Spatial and temporal modelling for perennial crop variety selection trials

Joanne De Faveri

Doctor of Philosophy

School of Agriculture, Food and Wine
University of Adelaide
URRBRAE, SA 5064

September 2013

Contents

Abstract

This thesis involves the investigation and development of methods for analysing data from variety selection trials in perennial crops. This involves identifying best varieties from data collected at multiple times in field trials, often from multiple locations and involving multiple traits. For accurate variety predictions the methods for analysis of such data need to account for the spatial correlation typically present in field trials and the temporal correlation induced by the repeated measures nature of the data. The methods also need to model the variety effects over time. The methods presented are based on the linear mixed model and estimation is performed using residual maximum likelihood (REML).

Spatial analysis methods are applied to data from multiple harvest times for two perennial crop data sets. These analyses show that spatial correlation is evident and the spatial analysis methods improve model fit. Simulation studies also show the spatial analysis methods provide better predictions of variety effects (closer to the true effects).

As the data from perennial crop variety selection trials is measured over time there is also a need to account for the temporal correlation between measurements. Separable models are presented that model the spatial and temporal residual covariance structure. These methods are suitable for large numbers of harvests. Application to a multi-harvest lucerne breeding data set shows these models to be an improvement on historical analysis approaches.

At the genetic level the variety effects need to be modelled over time. Two approaches are presented. The first approach involves applying factor analytic models to variety by harvest effects and using clustering to aid in interpretation and selection. The second approach uses cubic smoothing spline random regression. These approaches are applied to data from two traits from a lucerne breeding trial and are shown to successfully model the variety by harvest effects and aid in selection. As data is usually obtained from multiple trials at different locations, the above approaches are extended to the multi-environment situation and applied to a multi-harvest, multi-environment lucerne data set.

While the separable spatio-temporal residual models show an improvement on analysing each harvest time separately, they are very restrictive in that they assume common spatial correlation parameters across harvests (or traits). The initial spatial analyses on the two multi-harvest perennial crop data sets reveal that spatial correlation often varies between harvests and between traits. A more suitable non-separable covariance model is investigated that allows for differing spatial correlation across time or traits. The approach is based on the Multivariate Autoregressive model, initially for spatial correlation in one direction. Subsequently the model is extended to the two directional row-column situation using the theory of Multivariate Conditional Autoregressive models. These models are applied to the lucerne multi-harvest and multi-trait data using code written in R, and are shown in most cases to be a significant improvement to the separable residual models previously investigated.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

SIGNED:

DATE:

Acknowledgements

I would like to sincerely thank my official supervisors Wayne Pitchford and Brian Cullis for their invaluable guidance and support throughout this PhD. I would also like to thank Ari Verbyla who started me on this journey and made sure I reached the end. All three have contributed greatly to the direction of this research and I have learnt so much from all of them. I am very appreciative of all of their efforts.

I would like to thank Robin Thompson for his ideas which instigated this PhD and for being so generous with his wisdom. I would also like to thank Shoba Venkatanagappa and the NSW DPI lucerne breeding team for providing me with the lucerne data set to work on. I also thank Guangdi Li for providing the chicory data. The data sets provided challenging motivation for the methods in this PhD and were a vital part of this research. I am also very grateful to David Butler and Arthur Gilmour for their support in implementing models in ASReml and ASReml-R.

I gratefully acknowledge the financial support from the NSW Department of Primary Industries, CSIRO and The University of Adelaide for providing a generous scholarship. I also acknowledge my employer Queensland Department of Agriculture, Fisheries and Forestry for providing me with extended leave without pay while completing the PhD.

I thank my colleagues at CMIS and DAFF for their support and patience. I especially wish to thank Mary, Joanna, Maree and Janet for their continued encouragement.

Finally I would like to thank my family, especially Ari and my two wonderful children Matthew and Lara, for their endless support, love and understanding throughout these years. Without you I would not have got this far.

Chapter 1

Introduction

The challenging objective of plant breeding programmes is to improve crop performance through the development of new varieties. This requires the selection of best performing varieties for traits of interest, from breeding trials usually conducted across multiple environments (locations and years). To achieve the optimal selections it is important that the statistical methods used for analysing data from these breeding trials, are as accurate and efficient as possible.

Perennial (forage and horticultural) crop breeding trials are usually conducted in the field at several locations and multiple measurements are taken over time. The aim of these trials is to get accurate predictions for variety performance over time and across a range of environments. Interest therefore lies in the overall performance of varieties across time, as well as the interactions between variety and time, variety by environment and variety by time by environment.

While the analysis of data from variety selection trials in annual crops (such as wheat and barley) is well established and sound statistical methods (such as the spatial modelling techniques of Gilmour et al., 1997 and the multi-environment analysis methods of Smith et al., 2005) are in place and routinely used, the methods for analysis of data from perennial crops are not well developed. The repeated measures nature of the data from perennial crop breeding trials introduces additional modelling challenges above those of annual crops.

The aim of this thesis is to develop methods for analysing data from perennial crop variety selection trials to provide more accurate selections than those obtained from current methods, such as performing separate analyses of data from individual trials and harvests.

Throughout this thesis the terms variety effects and genetic effects are used interchangeably.

1.1 Motivating data

1.1.1 Experimental details

The motivating data sets used in this thesis are from variety selection trials in the perennial pasture crops lucerne and chicory.

Lucerne

Lucerne (*Medicago sativa L.*) is an important perennial forage legume grown for hay or grazing. Breeding lucerne involves maximizing agronomic performance, resistance to pests and diseases, and tolerance to abiotic stresses such as drought, acidic soils, salt and high temperatures (Irwin et al., 2001, Humphries & Auricht, 2001). The aim is therefore to identify lucerne varieties suited to varied climatic and environmental conditions.

Lucerne varieties are usually created by polycrossing a select number of male and female parental plants to produce synthetic populations (Busbice, 1969). This results in within variety genetic variation (Julier et al., 2000). In addition, cultivated lucerne is an outcrossing autotetraploid (Irwin et al., 2001), which also results in genetic heterogeneous varieties (Musial et al., 2002). This within variety genetic variability may mask the between variety variation that is primarily of interest in breeding trials.

There are two characteristics of lucerne varieties that impact on their yield, namely winter activity or dormancy, and persistence. Lucerne varieties vary in their winter activity or dormancy (Morley et al., 1957), ranging from dormant to highly active. Lodge (1985) and Lowe et al. (1985) show winter active varieties produce higher yields than dormant varieties but are less persistent. Lucerne varieties also differ in their crown structure or morphology which impacts on varietal persistence (Irwin et al., 2001).

Lucerne advanced breeding lines and commercial varieties are assessed for persistence and yield by conducting field trials which are sown in multiple environments, over multiple years (3-5), with multiple harvests varying in number and timing across the environments. Harvest times vary due to influence of rain, local temperature and seasonal conditions as these can influence the growth phase of lucerne significantly. The data arising from such breeding trials is therefore unbalanced across environments, and multivariate in nature.

In this thesis, the data from the 2003–2006 series of New South Wales Department of Primary Industries (NSW DPI) lucerne variety assessment trials is considered. These trials were planted at five sites across New South Wales and Queensland in Australia (Euloma, Leadville, Sandigo, Tamworth(TCCI) and Terry Hie Hie). The five trials involved testing 60 varieties (all varieties were grown in all trials, with eleven being commercial varieties) with 3 replicates at each trial. Each trial was designed as a Randomized Complete Block (RCB) in a rectangular array of 180 plots, with four of the trials laid out as 30 rows by 6 columns (with each block consisting of 30 rows by 2 columns) and one trial (Tamworth) laid out as 15 rows by 12 columns (with each block consisting of 5 rows by 12 columns). The experimental details are given in Table 1.1.

The variables analysed in this thesis are the yield and persistence of each variety.

Yield

Multiple harvests were conducted in each trial. Harvest time was determined by the influence of seasonal conditions on the growth phase of all varieties in the presence of varying winter activity and varying basic vegetative phase at each trial. Hence trials differed in their times between harvests and the total number of harvests. The number of harvests ranged from 3 to 10 as detailed in Table 1.1.

Yield was measured by cutting all trial plots at a consistent defined height at each harvest time and drying the samples to obtain dry matter weights expressed as kg DM/ha. The raw trial mean yields for each harvest at each trial are also given in Table 1.1. Two harvests, namely Sandigo harvest 2 and Tamworth harvest 1 were not measured in the same manner as the other harvests and were removed from the data and hence from the analysis presented in this thesis.

A plot of the yield data across harvest times for each of the 11 commercial varieties at each site is presented in Figure 1.1. This plot shows the extent of the differing harvest times and the differences in yield patterns across the sites. Within a site the varieties follow a similar pattern across the harvest times. This can be seen in further detail in Figure 1.2 where the yield at Terry Hie Hie is given for each variety over the harvest times. These profiles follow a similar trend over time but the trend is not of an obvious known functional form. There are obvious clear environmental effects (time and site) and the important issue will be to separate out the genetic effects from these environmental effects.

Persistence

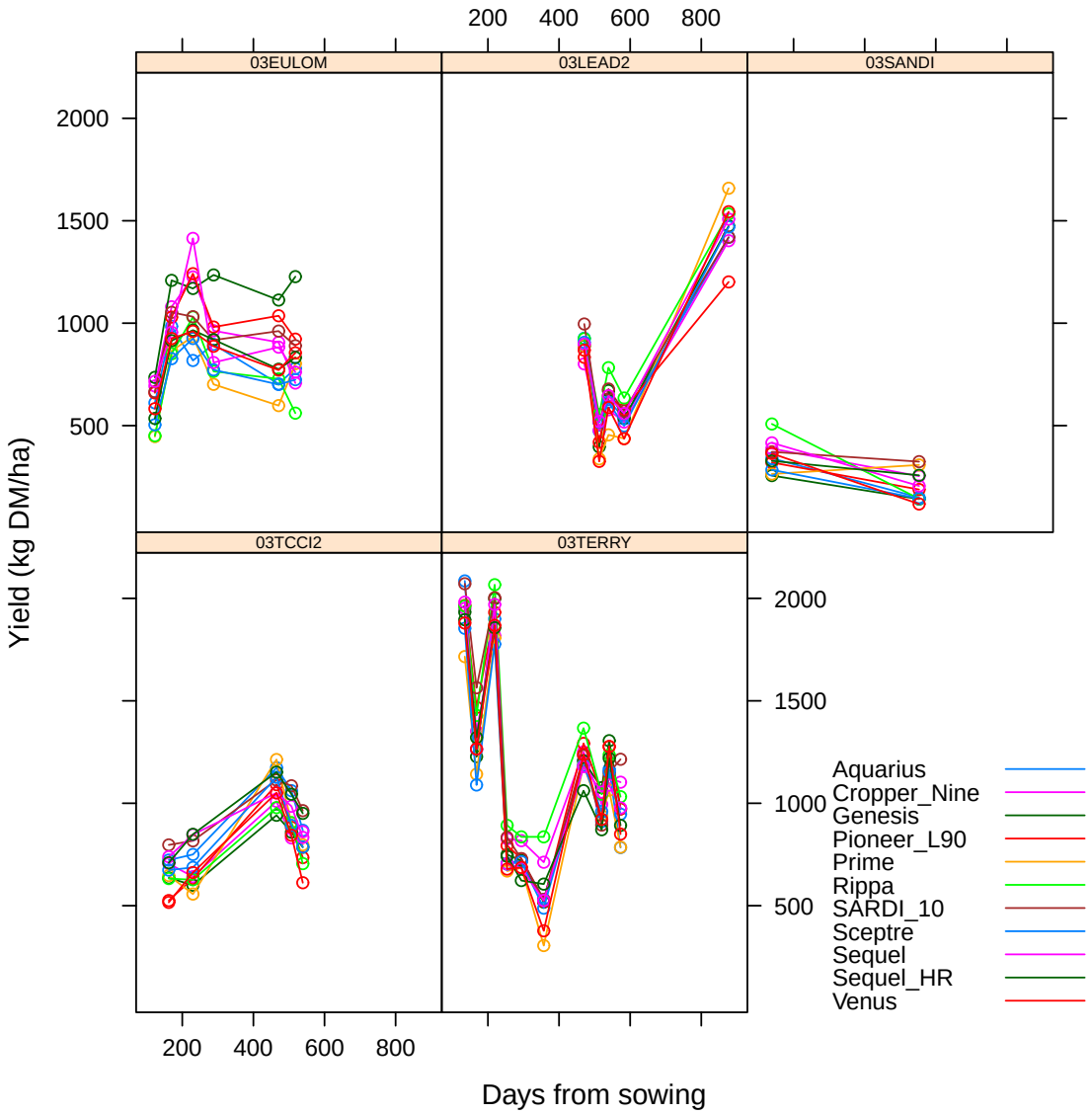
The persistence of each lucerne variety was recorded as the percentage of unit squares in a grid of 10 cm x 10 cm squares that had a lucerne plant/s present at each of the assessment dates at each site. The details of this method are given in Lodge & Gleeson (1984) where it is shown that this method reliably reflects changes in plant populations, for all but very high plant densities. The number of assessment times ranged from 3 to 7 as shown in Table 1.2.

Persistence data across assessment dates for the 11 commercial varieties at each site is presented in Figure 1.3. This plot shows the different assessment times at the different sites and the general trends of the commercial varieties at these sites. Figure 1.4 presents data from a single site (Terry Hie Hie) and displays the persistence data for all varieties (all three plots (replicates) of each variety) over the assessment dates at this site. The varieties at Terry Hie Hie follow a similar trend which resembles a quadratic curve over time, but differences are apparent between varieties in the level of persistence. There are clear plot effects with some plots high performers and others low.

Table 1.1: Experimental details of the 28 lucerne trial by harvest combinations together with the corresponding raw mean yield (kg DM/ha)

Trial	Harvest No.	Harvest Date	Plot size	Cut size	No. of Rows	No. of Cols	Mean yield (kg DM/ha)
Euloma	Sown	1/07/03					
Euloma	1	1/11/03	5.1×1.38	5.1×0.53	30	6	551
Euloma	2	18/12/03	5.1×1.38	5.1×0.53	30	6	910
Euloma	3	16/02/04	5.1×1.38	5.1×0.53	30	6	1084
Euloma	4	14/04/04	5.1×1.38	5.1×0.53	30	6	884
Euloma	5	14/10/04	5.1×1.38	5.1×0.53	30	6	834
Euloma	6	30/11/04	5.1×1.38	5.1×0.53	30	6	771
Leadville	Sown	27/06/03					
Leadville	1	14/10/04	5×2	5×0.5	30	6	909
Leadville	2	25/11/04	5×2	5×0.5	30	6	447
Leadville	3	21/12/04	5×2	5×0.5	30	6	637
Leadville	4	3/02/05	5×2	5×0.5	30	6	538
Leadville	5	24/11/05	5×2	5×0.5	30	6	1472
Sandigo	Sown	25/06/03					
Sandigo	1	17/11/03	5×2	5×0.5	30	6	381
Sandigo	2	20/10/04	5×2	5×0.5	30	6	-
Sandigo	3	04/01/05	5×2	5×0.5	30	6	237
Tamworth(TCCI)	Sown	11/07/03					
Tamworth(TCCI)	1	6/11/03	5×2	4×1.5	15	12	-
Tamworth(TCCI)	2	10/12/03	5×2	4×0.5	15	12	664
Tamworth(TCCI)	3	16/02/04	5×2	4×0.5	15	12	701
Tamworth(TCCI)	4	8/10/04	5×2	5×0.5	15	12	1056
Tamworth(TCCI)	5	19/11/04	5×2	5×0.5	15	12	954
Tamworth(TCCI)	6	21/12/04	5×2	5×0.5	15	12	843
Terry Hie Hie	Sown	22/07/03					
Terry Hie Hie	1	13/11/03	5×2	4×0.5	30	6	1934
Terry Hie Hie	2	16/12/03	5×2	4×0.5	30	6	1314
Terry Hie Hie	3	5/02/04	5×2	4×0.5	30	6	1948
Terry Hie Hie	4	11/03/04	5×2	5×0.5	30	6	808
Terry Hie Hie	5	20/04/04	5×2	5×0.5	30	6	779
Terry Hie Hie	6	22/06/04	5×2	5×0.5	30	6	624
Terry Hie Hie	7	12/10/04	5×2	5×0.5	30	6	1257
Terry Hie Hie	8	2/12/04	5×2	5×0.5	30	6	1012
Terry Hie Hie	9	23/12/04	5×2	5×0.5	30	6	1229
Terry Hie Hie	10	24/01/05	5×2	5×0.5	30	6	977

Figure 1.1: Plot of mean lucerne yield for each of the 11 commercial varieties across harvests at each site

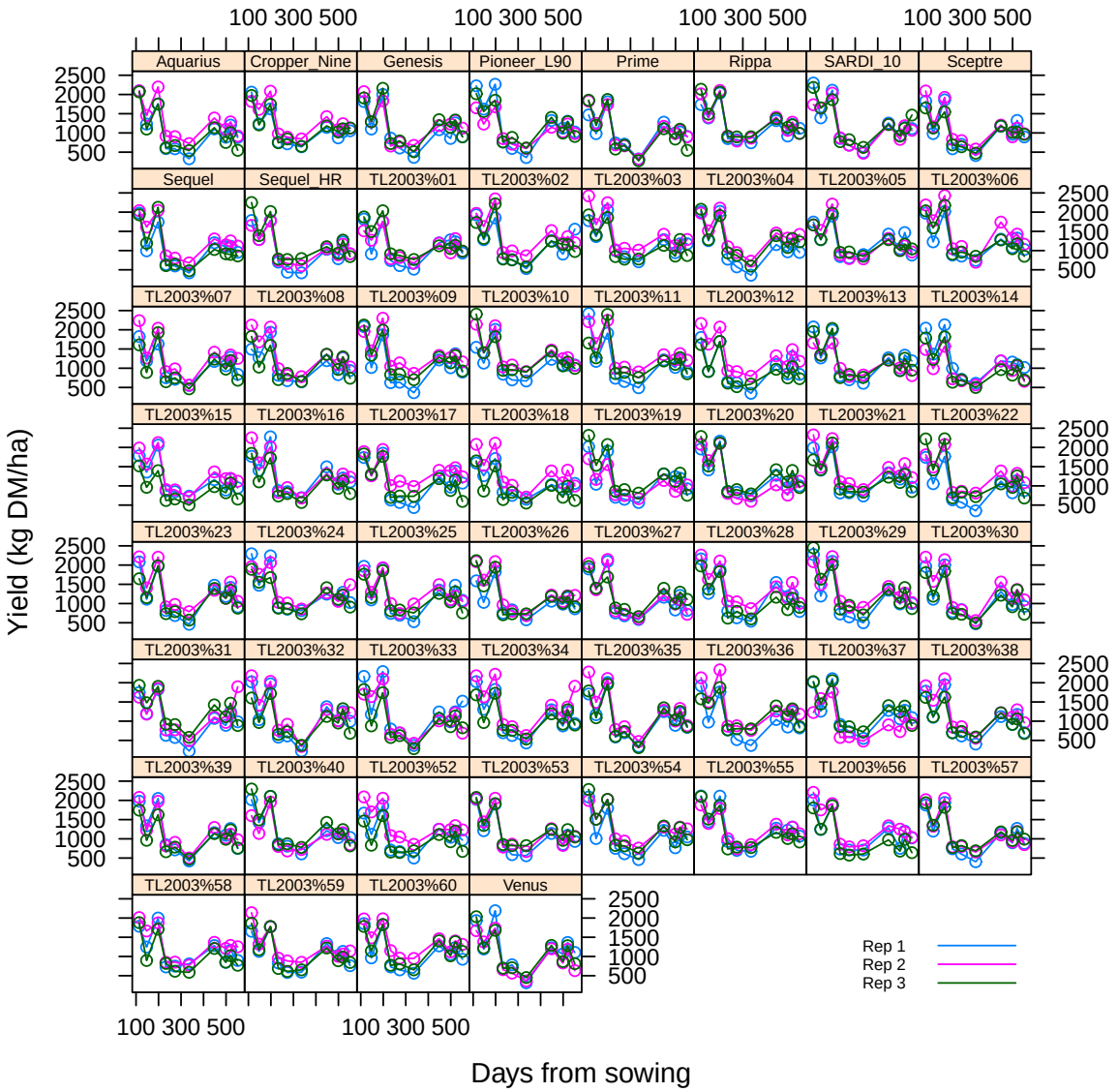


Chicory

Forage chicory (*Cichorium intybus L.*) is a summer active perennial herb that provides feed for livestock. The first cultivar released specifically as a forage crop (in New Zealand in 1985) was Grasslands Puna, and many commercial cultivars in Australia are based on this cultivar. Grasslands Puna and most varieties bred from it (e.g. Puna11) are winter dormant, while some other varieties e.g. La Lacerta (originally from Uruguay), Commander and Grouse have increased levels of winter activity.

The chicory data presented in this thesis arises from a variety selection trial planted at Keith, South Australia. The trial involved testing 22 varieties (20 of which were chicory varieties and 2 being English Plantain (*Plantago lanceolata L.*) with 4 replicates in a Randomised Complete Block (RCB) design, laid out in a rectangular array of 22 rows

Figure 1.2: Plot of raw lucerne yield data across harvests for Terry Hie Hie for all varieties and replicates



and 4 columns (with each column forming a replicate). There were 11 harvests conducted on the trial. Table 1.3 gives the harvest dates and mean yields for this trial.

Estimates of yield (herbage production) of each plot were obtained from visual assessments made on three sections of each plot, using a scoring scale of $1 - 10$. Quadrat cuts (10-15 over each trial) were then used to calibrate the visual scores and convert to total herbage production (kg dry matter (DM)/ha).

A plot of the raw yield (kg DM/ha) data across harvests (Figure 1.5) shows large differences between varieties across the harvests and clear harvest time effects. The first harvest contains substantially higher yields than subsequent harvests. As interest lies in the genetic variation of the chicory varieties (not English Plantain) it will be necessary to separate the chicory varieties from the two non chicory varieties in the trial.

Note that even though this data comes from a single site it has been included in this

Table 1.2: Persistence data from the lucerne trials

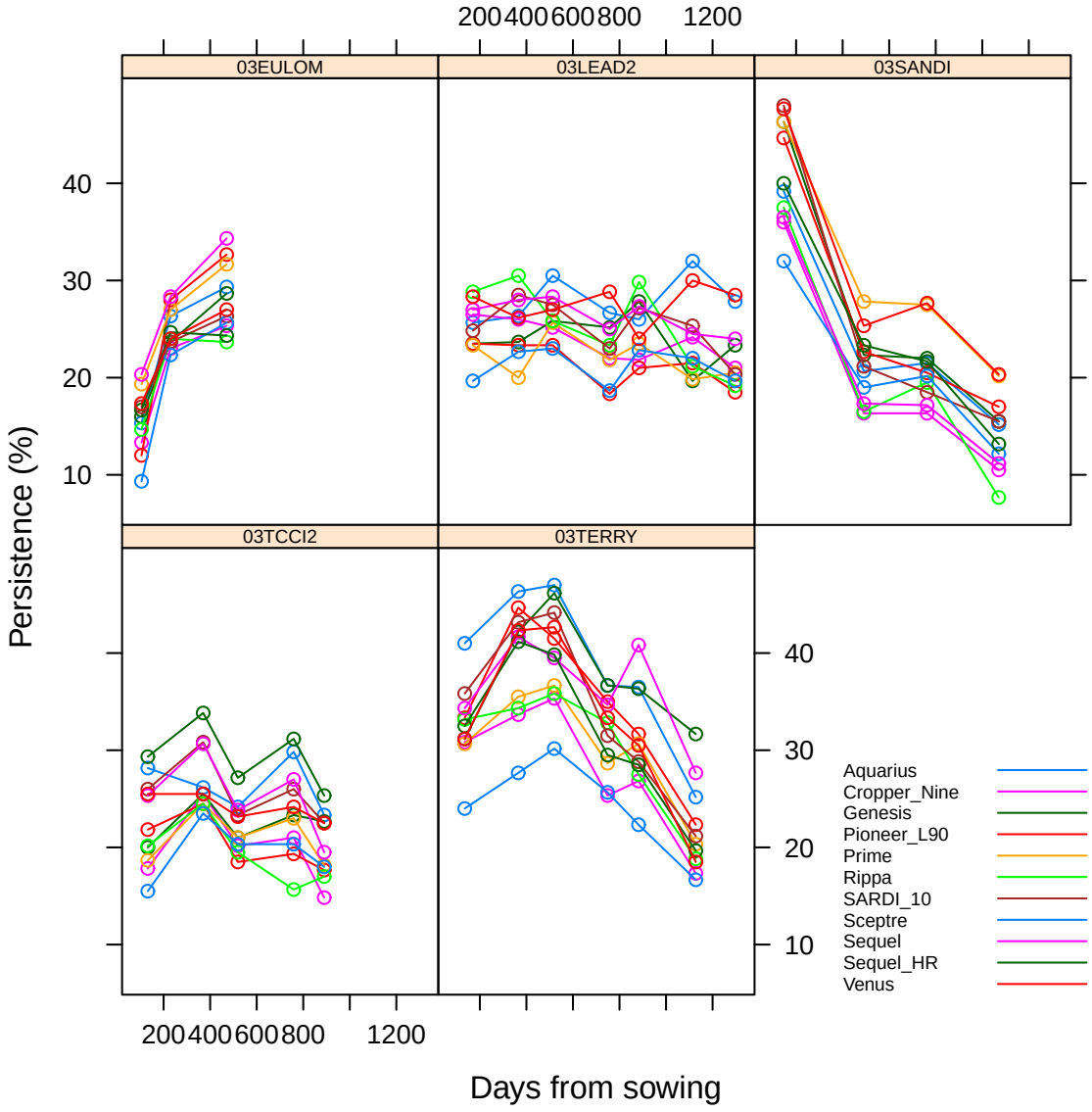
Trial	Assessment No.	Assessment Date	Mean Persistence (%)
Euloma	1	14/10/03	17.8
Euloma	2	16/02/04	26.1
Euloma	3	14/10/04	29.4
Leadville	1	18/12/03	26.9
Leadville	2	01/07/04	26.8
Leadville	3	25/11/04	28.3
Leadville	4	28/07/05	24.8
Leadville	5	01/12/05	26.2
Leadville	6	19/07/06	24.0
Leadville	7	18/01/07	22.3
Sandigo	1	25/11/03	43.8
Sandigo	2	03/11/04	21.0
Sandigo	3	02/08/05	20.3
Sandigo	4	06/06/06	13.6
Tamworth(TCCI)	1	10/11/03	25.0
Tamworth(TCCI)	2	05/07/04	29.4
Tamworth(TCCI)	3	01/12/04	24.6
Tamworth(TCCI)	4	29/07/05	25.8
Tamworth(TCCI)	5	07/12/05	21.5
Terry Hie Hie	1	13/11/03	34.1
Terry Hie Hie	2	30/06/04	40.8
Terry Hie Hie	3	02/12/04	41.7
Terry Hie Hie	4	19/07/05	33.9
Terry Hie Hie	5	29/11/05	32.8
Terry Hie Hie	6	02/08/06	22.9

thesis to allow models developed for multi-harvest data to be tested on another species (in addition to lucerne) to examine the utility of the methods for perennial crops in general.

Table 1.3: Harvest dates and mean yields for the Keith chicory trial harvests

Harvest No.	Harvest Date	Mean yield (kg DM/ha)
1	19/12/05	3007
2	16/3/06	362
3	22/6/06	858
4	5/9/06	180
5	14/11/06	140
6	21/2/07	142
7	12/7/07	694
8	3/9/07	260
9	1/11/07	532
10	26/3/08	754
11	28/5/08	6.8

Figure 1.3: Plot of mean lucerne persistence for each of the 11 commercial varieties over harvests, for each site

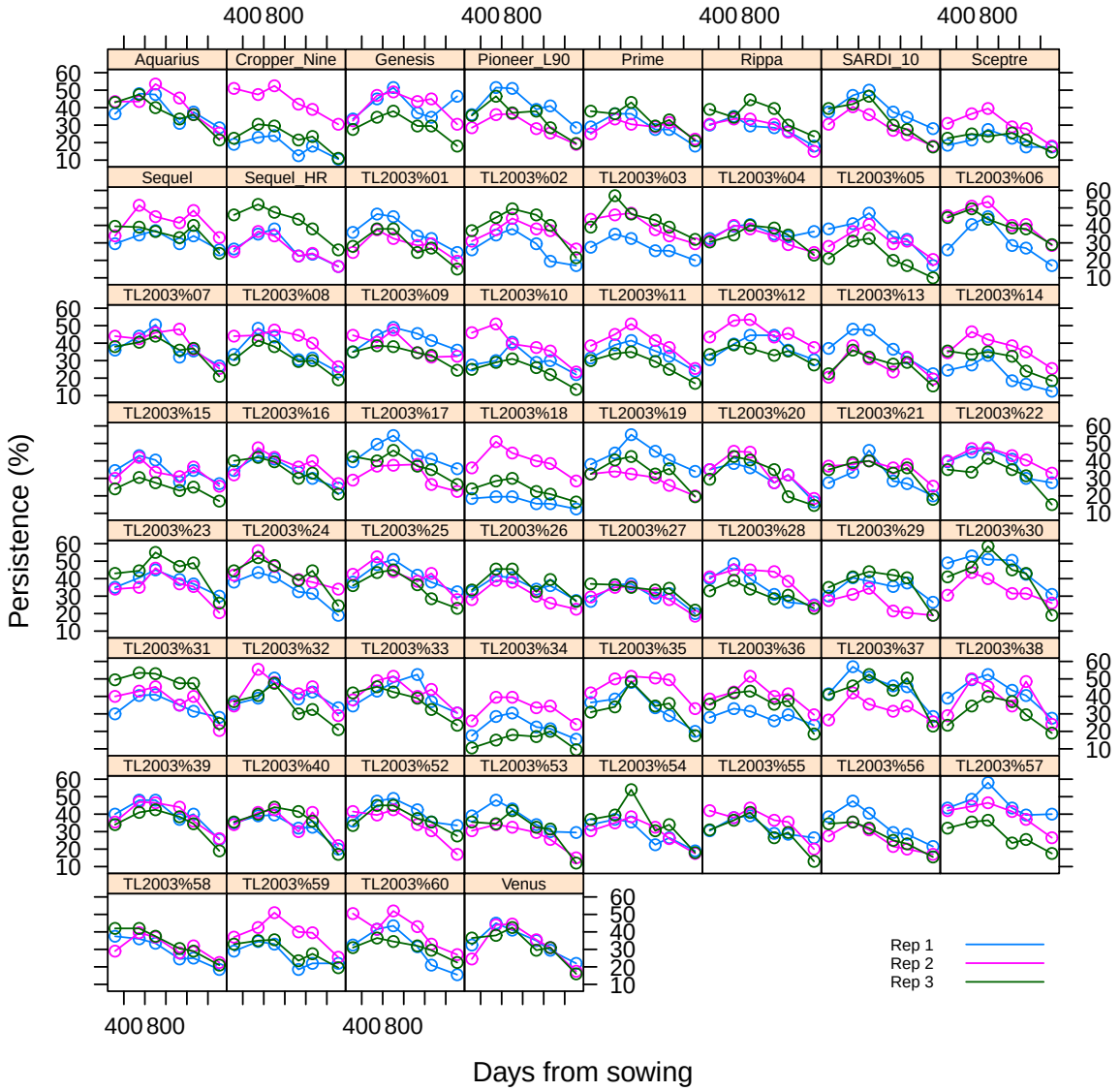


1.2 Overview of issues

The motivating data sets outlined above are typical of perennial forage and horticultural crop breeding trials that are usually conducted in the field at several locations with multiple measurements taken at multiple times. For forage crops these repeated measurements may be taken within a season or a number of seasons, while in horticultural tree-based crops these may be annual measurements taken over many years. The aim of these trials is to obtain reliable and accurate predictions for variety means and variety differences over time across a range of environments, and an understanding of variety by time by environment interactions. The ultimate aim is to select superior varieties for genetic improvement.

Field trials in general, usually show spatial variation (Gilmour et al., 1997, Stefanova

Figure 1.4: Plot of raw lucerne persistence data at Terry Hie Hie for all varieties and replicates

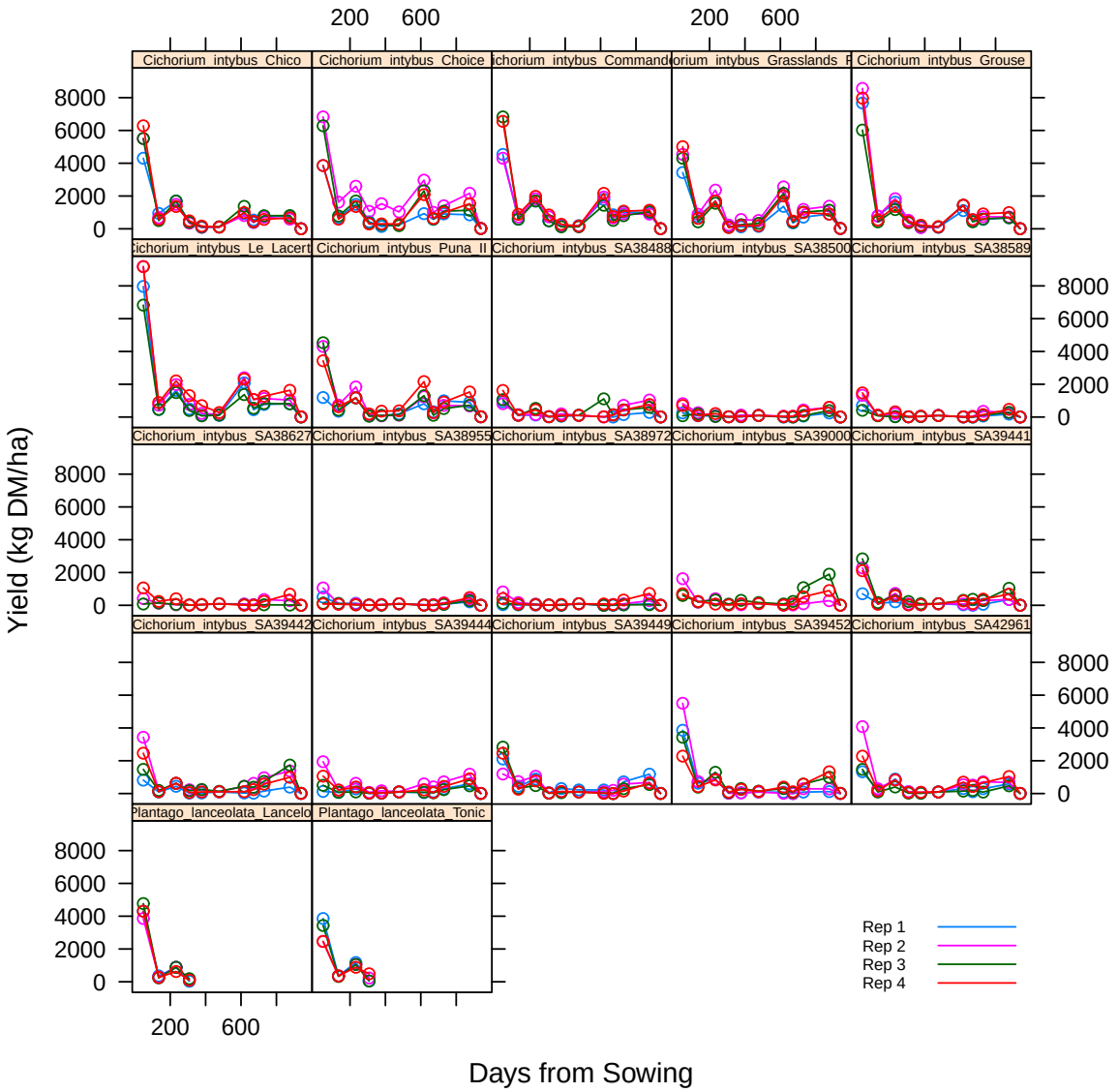


et al., 2009, Dutkowski et al., 2002), and in some cases interplot competition (Stringer & Cullis, 2002). This spatial variation and interplot competition needs to be taken into account otherwise variety estimates may be affected.

As repeated measurements are made on each trial (for example multiple harvests), there is a need to account for the temporal correlation between measurements made on each plot (or tree). The variety (genetic) effects also need to be modelled over time and account made for possibly heterogenous genetic variances and correlations between genetic effects at different times.

As variety selections are usually based on data from a number of field trials in different locations, methods of multi-environment (MET) analysis need to be employed. These methods are not yet well developed for perennial crops but there are numerous methods for MET analysis in annual crops. While the MET analysis methods for annual crops

Figure 1.5: Plot of raw chicory yield data across harvests for Keith for all varieties and replicates



address some of the spatial and multi-environment issues of perennial crop breeding trials, they are for single harvest trials and so are not directly applicable when multiple harvests are involved. Methods for perennial crop trials need to accommodate instances where harvest times are not regular across trials (and hence data is very unbalanced) and/or when there are long sequences of harvest times. An example of this is with perennial pasture variety selection trials where multiple harvests may be conducted each year over a number of years and where harvest times may vary across locations due to influence of rain, local temperature and other seasonal conditions.

In summary, the statistical methods for the analysis of data from perennial crop breeding trials need to account for the spatial variation and correlation within a trial, the temporal correlation between repeated measurements and model the genetic effects over time and across sites in a meaningful way.

1.3 Literature review

The following presents an overview of the literature covering recent spatial analysis methods, multi-environment trial (MET) analysis methods, spatio-temporal, and response profile modelling methods that are currently applied in the analysis of annual crop breeding data and other applications, which may be applicable to perennial crop breeding data.

1.3.1 Spatial analysis of field trials

Varietal selection field trials usually involve evaluation of many varieties over a potentially large and variable trial area. In analysing data from these field trials spatial variation needs to be modelled in order for accurate variety effects to be estimated.

Spatial variation has long been recognized as an important issue in field trials. The classic Fisherian approach (Brien & Bailey, 2006 and references therein) attempts to minimize the effect of spatial variation in the field by careful selection of trial sites, application of randomization of plots to varieties and inclusion of sources of variation due to the trial design in the analysis. This randomization approach is often not sufficient to model all sources of spatial variation and it is often necessary to build on the randomization model to include terms over and above the design components.

There have been many methods proposed to account for spatial variation including the earliest methods in which neighbouring plot yields were used as covariates in the analysis (see Bartlett, 1978 for an account). Wilkinson et al. (1983) proposed a significant improvement with a smooth trend plus independent error model with the trend being removed by differencing the data. Removing trend by differencing neighbouring plots was also used by Green et al. (1985) and Besag & Kempton (1986). There have been many other approaches to spatial analysis including the one dimensional models of Gleeson & Cullis (1987), where trend was modelled using time series models, and their extension to two dimensions by Cullis & Gleeson (1991), using a separable correlation structure.

Gilmour et al. (1997) extended the method of Cullis & Gleeson (1991) by identifying three major components of spatial variation to be modelled, namely local and global smooth spatial trend and extraneous variation. Local smooth trend arises because data from plots close together are more similar to those apart and may reflect for example, changes in soil moisture, fertility or depth. Global trend reflects non-stationary trend across the field and is usually aligned with the rows or columns of the field. Extraneous variation often arises due to management practices or experimental procedures, for example serpentine harvesting. To model global trend and extraneous variation Gilmour et al. (1997) fitted polynomial or spline functions (Verbyla et al., 1999) to the row or column co-ordinates. They modelled local trend using a covariance structure for the residuals. It is this separation of global trend and local trend which is an important extension from the previous spatial analysis methods.

Gilmour et al. (1997) present a mixed model for the data $\boldsymbol{y}$ (a $n \times 1$ vector) from an individual field trial consisting of n plots in an array of r rows by c columns (data ordered as rows within columns) as follows

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e} \tag{1.3.1}$$

where $\boldsymbol{\tau}$ is a vector of fixed effects with design matrix $\boldsymbol{X}$, $\boldsymbol{u}$ is the vector of random genotype effects with associated design matrix $\boldsymbol{Z}$ and $\boldsymbol{e}$ is the vector of residuals ordered as for the data vector (rows within columns).

The random effects in the mixed model above, (1.3.1) are assumed to follow a Normal distribution with mean $\mathbf{0}$ and variance

$$\text{var}\left(\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{G} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{R} \end{bmatrix}$$

The residual term $\boldsymbol{e}$ can be partitioned into two components, a vector $\boldsymbol{\xi}$ of local smooth spatial trend and a vector $\boldsymbol{\eta}$ of independent measurement errors. The local smooth spatial trend can be modelled using a two dimensional separable correlation structure (Gilmour et al., 1997). For example the covariance matrix for $\boldsymbol{\xi}$ is given by,

$$\text{var}\,(\boldsymbol{\xi}) = \sigma^2 \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r \tag{1.3.2}$$

where $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\Sigma}_r$ are the $c \times c$ and $r \times r$ spatial correlation matrices corresponding to the column and row dimensions respectively as in Gilmour et al. (1997). The symbol $\otimes$ in (1.3.2) represents the Kronecker product (see Result A.6 in Appendix).

These two matrices are typically assumed to arise from autoregressive processes of order 1 (labelled ar1) so that they are each functions of a single (autocorrelation) parameter (ϕ_r in the row direction and ϕ_c in the column direction). In this case, (1.3.2) is given symbolically by ar1(Column).ar1(Row).

The spatial correlation matrix for the row dimension $\boldsymbol{\Sigma}_r$ can be written as

$$\boldsymbol{\Sigma}_r = \begin{bmatrix} 1 & \phi_r & \phi_r^2 & \dots & \phi_r^{r-1} \\ \phi_r & 1 & \phi_r & \dots & \phi_r^{r-2} \\ \phi_r^2 & \phi_r & 1 & \dots & \phi_r^{r-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_r^{r-1} & \phi_r^{r-2} & \dots & \phi_r & 1 \end{bmatrix}$$

The measurement error process $\boldsymbol{\eta}$, has variance σ_η^2. Hence the variance of the error process $\boldsymbol{e} = \boldsymbol{\xi} + \boldsymbol{\eta}$ is given by

$$\boldsymbol{R} = \sigma^2 \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + \sigma_\eta^2 \boldsymbol{I}_n$$

Sources of global and extraneous variation can be included in $\boldsymbol{u}$ and/or $\boldsymbol{\tau}$ in (1.3.1). To determine the appropriate spatial correlation structure and to identify global and extraneous variation, Gilmour et al. (1997) used diagnostic tools such as the sample variogram and residual plots (of residuals against row (or column) number). The variogram dis-

plays the semi-variance of the residuals for pairs of plots a given distance apart (it will be discussed in further detail in subsequent chapters).

While Gilmour et al. (1997) included model-based components of variation that occur because of management process, extraneous and local variation or correlation, they omitted the variables inherent in the design of field trials. The combination of randomization and model-based components is now considered preferable and appropriate for analysis (Beeck et al., 2010).

While the spatial analysis methods of Gilmour et al. (1997) are now routinely used in the analysis of annual field crops, their use in analysing data for perennial crops is not so widespread. In general, more simplistic approaches have been attempted to account for spatial variation in perennial crops, for example in Smith & Casler (2004) and Smith & Kearney (2002), where the methods used have been based on Nearest Neighbour analyses first introduced by Papadakis in 1937 (see Bartlett, 1978 for a review). Exceptions to this include Stringer & Cullis (2002) and Smith et al. (2007) where the spatial analysis techniques of Gilmour et al. (1997) are applied to sugarcane selection trials, Costa e Silva et al. (2001), Dutkowski et al. (2002) and Hardner et al. (2010), where the above spatial analysis methods have been used in forest genetic trials.

1.3.2 Analysis of multi-environment trials (METs)

Usually plant variety improvement programs involve the evaluation of new varieties in field trials at a number of trial locations. These trials are referred to as METs. There have been many methods put forward for analysing METs in annual crops. Early methods were based on ANOVA methods which did not give an understanding of genotype (or variety) by environment (GxE) interaction. Other early methods which did allow investigation of GxE include using a regression based approach on environmental variables (for example Knight, 1970) and using regression of varietal yield on the mean yield of all varieties in each environment (for example Finlay & Wilkinson, 1963).

Mixed model approaches have become popular as they can easily handle incomplete data, (where not all varieties are grown at all environments) and they can model within trial variation appropriately. Smith et al. (2005) provided a review of such methods for analysing multi-environment trials (METs).

A two stage approach has often been used for the analysis of MET data. This involves first estimating variety means separately for each trial and then combining them for an overall analysis. The overall analysis may use mixed model methods for example (Cullis et al., 1996a, Cullis et al., 1996b, Frensham et al., 1997) or fixed effects models, for example the AMMI (Additive Main effects and Multiplicative Interaction) model of Gauch (1992). These two stage methods are an approximation to the overall analysis of individual plot data from all trials and may not provide a good approximation if there is spatial variation within trials or heterogeneity of error variance between trials. Cullis et al. (1998) presented a more comprehensive combined spatial MET analysis in which individual plot data is analysed and allowance is made for separate spatial modelling and separate error variance for each trial.

Several authors have suggested the use of factor analytic (fa) models for modelling GxE effects (e.g. Smith et al., 2001, Piepho, 1997, Thompson et al., 2003). Smith et al. (2001) extended the spatial MET analysis method of Cullis et al. (1998) to include factor analytic models to model the variety effects in different environments as a series of multiplicative terms and provide a parsimonious approximation to the fully unstructured variance covariance matrix. Smith et al. (2001) presented a mixed model for modelling the observations from a series of MET trials using the following notation. They consider data from t trials in which m varieties are grown. The j^{th} trial consists of n_j plots in a rectangular array of r_j rows by c_j columns ($n_j = r_j c_j$). The data vector for trial j, $\boldsymbol{y_j}$ is a $n_j \times 1$ vector of observations ordered as rows within columns. The data combined across trials is denoted by $\boldsymbol{y} = (\boldsymbol{y}_1^T \ldots \boldsymbol{y}_t^T)^T$; this is a $N \times 1$ vector with $N = \sum_{j=1}^{t} n_j$.
The mixed model for $\boldsymbol{y}$ is given by

$$\boldsymbol{y} = \boldsymbol{X\tau} + \boldsymbol{Z}_g\boldsymbol{u}_g + \boldsymbol{Z}_o\boldsymbol{u}_o + \boldsymbol{e} \tag{1.3.3}$$

where $\boldsymbol{\tau}$ and $\boldsymbol{u}_g$ are the vectors of fixed effects and random genetic effects respectively and $\boldsymbol{u}_o$ contains other non-genetic random effects.

The random effects from the linear mixed model (1.3.3) are assumed to follow a Normal distribution with zero mean vector and variance-covariance matrix

$$\text{var}\left(\begin{bmatrix} \boldsymbol{u}_g \\ \boldsymbol{u}_o \\ \boldsymbol{e} \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{G_g} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G_o} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R} \end{bmatrix}$$

The MET approach of Smith et al. (2001) assumes a separate spatial model for each trial and the residuals from different trials are assumed independent. Hence the full residual variance matrix $\boldsymbol{R}$ is given by a block diagonal matrix; $\boldsymbol{R} = \text{diag}\,(\boldsymbol{R}_j)$ where $\boldsymbol{R}_j$ is the residual variance matrix for the j^{th} trial, $j = 1, \ldots, t$. The spatial modelling approach of Gilmour et al. (1997), as described in the previous section is applied to each trial. Hence $\boldsymbol{R}_j = \sigma_j^2 \boldsymbol{\Sigma}_{c_j} \otimes \boldsymbol{\Sigma}_{r_j}$ for each trial.

The random effects term $\boldsymbol{u}_g$ is a $tm \times 1$ vector of genetic effects for m varieties in each of the t environments (ordered as varieties within trials). Smith et al. (2001) considered the trials to be like different traits measured on each variety, and assumed the variance structure for the genetic effects to be of the separable form

$$\text{var}\,(\boldsymbol{u}_g) = \boldsymbol{G_g} = \boldsymbol{G}_t \otimes \boldsymbol{G}_v$$

where $\boldsymbol{G}_t$ and $\boldsymbol{G}_v$ are the symmetric $t \times t$ and $m \times m$ component matrices for trials and varieties respectively. Smith et al. (2001) assume that the genetic effects for the same variety in different environments are correlated but that the varieties are independent. Hence $\boldsymbol{G}_v = \boldsymbol{I}_m$. However if there is a known family structure in the varieties alternative forms for the variety component matrix are possible. For example if a pedigree matrix

is known which specifies the relationship between varieties then $\boldsymbol{G}_v$ may take different forms, (Oakey et al., 2006).

If

$$\boldsymbol{G}_g \;=\; \boldsymbol{G}_t \otimes \boldsymbol{I}_m$$

then $\boldsymbol{G}_t$ is the so called genetic variance matrix of dimension $t \times t$ with diagonal elements being the genetic variances of individual trials and the off diagonal elements the genetic covariances between pairs of trials. The genetic variance matrix can take a number of possible forms including common variance/common covariance, heterogeneous variance/common covariance, heterogeneous variance/common correlation, and unstructured.

The unstructured matrix is the most general form for the genetic variance matrix. It involves estimating all of the genetic variances for the trials and covariances between pairs of trials, resulting in $t(t+1)/2$ parameters being required to be estimated. If the number of trials is large this matrix is often difficult to fit.

While the common covariance/correlation models require significantly less parameters to be estimated than the unstructured matrix they are not often flexible enough to model the genetic variance matrix. It is unlikely that variances will be constant across the different environments and that all pairs of environments will have the same genetic correlation. A more parsimonious structure which is a good approximation to the unstructured model is the factor analytic model proposed by Smith et al. (2001).

Factor analytic models

The factor analytic model (fa) is based on the multivariate technique of factor analysis (Mardia, 1988). In the general sense factor analysis is used to model the covariance structure between a set of t variates, $X_1, \ldots, X_t$ with the aim being to account for the covariances in terms of a smaller number of hypothetical factors. Smith et al. (2001) used the factor analysis approach to provide a variance structure for the genetic variance matrix $\boldsymbol{G}_t$.

The model is defined in terms of the unobserved variety effects in each environment. That is

$$u_{g_{ij}} = \sum_{r=1}^{k} \lambda_{jr} f_{ir} + \delta_{ij} \tag{1.3.4}$$

where $u_{g_{ij}}$ is the random effect for variety $i, (i = 1, \ldots, m)$ in environment $j, (j = 1, \ldots, t)$, f_{ir} is the value (or score) for variety i in the r^{th} hypothetical factor $(r = 1, \ldots, k)$, λ_{jr} is the coefficient (referred to as a factor loading) for environment j, and δ_{ij} is a residual. The residuals, δ_{ij} are independent with variance, Ψ_j (known as the specific variance of the j^{th} environment).

The model can be represented in vector notation as

$$\begin{aligned} \boldsymbol{u}_g &= (\boldsymbol{\lambda}_1 \otimes \boldsymbol{I}_m)\boldsymbol{f}_1 + \ldots + (\boldsymbol{\lambda}_k \otimes \boldsymbol{I}_m)\boldsymbol{f}_k + \boldsymbol{\delta} \\ &= (\boldsymbol{\Lambda} \otimes \boldsymbol{I}_m)\boldsymbol{f} + \boldsymbol{\delta} \end{aligned}$$

where $\boldsymbol{\lambda}_r$ is a $t \times 1$ vector of loadings, $\{\lambda_{jr}\}$, $\boldsymbol{f}_r$ is a $m \times 1$ vector of factor scores, $\{f_{ir}\}$, $\boldsymbol{\delta}$ is a $mt \times 1$ vector of residuals, $\{\delta_{ij}\}$, $\boldsymbol{\Lambda}$ is a $t \times k$ matrix of loadings, $\{\boldsymbol{\lambda}_1 \ldots \boldsymbol{\lambda}_k\}$, and $\boldsymbol{f}$ is a $mk \times 1$ vector of factor scores, $[\boldsymbol{f}_1^T \boldsymbol{f}_2^T \ldots \boldsymbol{f}_k^T]^T$.

The joint distribution of $\boldsymbol{f}$ and $\boldsymbol{\delta}$ is given by

$$\begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{\delta} \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{bmatrix} \boldsymbol{G_f} \otimes \boldsymbol{I}_m & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Psi} \otimes \boldsymbol{I}_m \end{bmatrix} \right]$$

where $\boldsymbol{\Psi}$ is a diagonal matrix of specific variances, one for each trial, that is $\boldsymbol{\Psi} = \text{diag}\,(\Psi_1 \ldots \Psi_t)$. The factor scores are commonly assumed to be independent and scaled to have unit variance, so that $\boldsymbol{G}_f = \boldsymbol{I}_k$. Hence the variance matrix for the variety effects in each environment is given by

$$\begin{aligned} \text{var}\,(\boldsymbol{u}_g) &= (\boldsymbol{\Lambda} \otimes \boldsymbol{I}_m)\text{var}\,(\boldsymbol{f})\,(\boldsymbol{\Lambda}^T \otimes \boldsymbol{I}_m) + \text{var}\,(\boldsymbol{\delta}) \\ &= (\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}) \otimes \boldsymbol{I}_m \end{aligned}$$

Therefore the factor analytic model results in the following form for $\boldsymbol{G}_t$,

$$\boldsymbol{G}_t = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}$$

so the genetic variances and covariances are given by

$$\begin{aligned} \sigma_{g_{jj}} &= \sum_{r=1}^{k} \lambda_{jr}^2 + \Psi_j \\ \sigma_{g_{ij}} &= \sum_{r=1}^{k} \lambda_{ir}\lambda_{jr} \end{aligned}$$

where $\sigma_{g_{jj}}$ is the genetic variance for trial j and $\sigma_{g_{ij}}$ is the genetic covariance between trials i and j, i.e $\boldsymbol{G}_t = \{\sigma_{g_{ij}}\}$.

Smith et al. (2001) showed that when $k > 1$, $\boldsymbol{\Lambda}$ is not necessarily unique and that $k(k-1)/2$ independent constraints need to be imposed on the elements of $\boldsymbol{\Lambda}$ to ensure a unique solution. Smith et al. (2001) set the $k(k-1)/2$ elements in the upper triangle of $\boldsymbol{\Lambda}$ to be zero. That is $\lambda_{jr} = 0$ for $j < r = 2, \ldots, k$. Hence for $k = 2$, the parameter λ_{12} is set to zero. The above constraint for $\boldsymbol{\Lambda}$ enables a unique solution to be computed but it has no biological meaning. Smith et al. (2001) also suggest using a rotation (based on principal components) of the loadings, once estimated as above, in order to obtain meaningful interpretation of the environmental loadings and variety scores. This rotation ensures that the first factor accounts for the maximum genetic covariance between environments, the second factor accounts for the next largest amount and is orthogonal to the first factor,

and so on.

The factor analytic model for variety effects, (1.3.4) is similar to a standard random regression model but with the difference that both the covariates and the regression coefficients are unknown and must be estimated from the data. Kelly et al. (2007) shows the factor analytic model to be generally the model of best fit across a range of variety selection trials in annual cereal crops.

1.3.3 Analysis of multi-harvest data

With perennial crops, when there are multiple harvests at each site there is a need to account for temporal correlation in the residuals (due to the repeated measurements on each plot) as well as spatial correlation. One way of doing this is to use a mixed model, assuming a 3 way (harvest $\times$ column $\times$ row) separable process for the residual variance structure as in Smith et al. (2007) in their analysis of multi-harvest sugarcane breeding data. This approach will be discussed further in Chapter 4.

Other, more simplistic examples of analysis of multi-harvest data include Smith & Casler (2004), where harvest yields of forage grass trials are summed over years and spatially adjusted using nearest neighbour covariates, and Resende et al. (2006) where tea data is analysed spatially at separate times, longitudinally ignoring spatial, and using a bivariate spatial model for 2 harvests.

While the separability assumption of spatio temporal processes fits well mathematically, it may not always be suitable in practice. It is likely that spatial parameters will vary over time and more flexible models for modelling the residual variance structure of multi-harvest data need to be investigated.

1.3.4 Temporal models

Repeated measures data arises when multiple measurements are made over time on the same trait on the same experimental units. The analysis of repeated measures data needs to take into account the correlation between measurements over time. There are many books that focus on the analysis of temporal, repeated measures, or longitudinal data, for example Crowder & Hand (1990), Jones (1993), Diggle et al. (2002). Papers of particular relevance to the analysis of repeated measures data in this thesis include Bjornsson (1978), Diggle (1988), Cullis & McGilchrist (1990), Verbyla & Cullis (1990).

Bjornsson (1978) recognized the need to account for serial correlation between repeated measurements in perennial crop trials and that the correlation often decreases with increasing time intervals. They incorporated serial correlation into their repeated measures model by using a general linear model with autoregressive errors. Diggle (1988) proposed a more comprehensive approach to the analysis of repeated measures data in which the serial correlation is modelled as well as additional components such as measurement error and variation between units.

Diggle model

Diggle (1988) recognized that in repeated measures data, as well as serial correlation between repeated measurements, a number of other features contribute to the covariance structure of the process. The model of Diggle (1988) takes into account the concept that measurements made closely in time on the same unit will vary, no matter how close the observations are. This reflects measurement error. The model also recognizes that on average some units perform well and others are low performers. This reflects a consistent 'unit' effect (a plot or animal effect) and results in a positive correlation between any two measurements on the same unit. Lastly the model incorporates the serial correlation between measurements over time on the same unit, recognizing that this correlation depends on the separation in time between the measurements and that this correlation typically decreases as the separation increases.

Assume there are m units, $i = 1, \ldots, m$, on which n_i measurements have been made on the i^{th} unit, with the total number of measurements being $n = \sum_{i=1}^{m} n_i$. Let Y_{ij} denote the j^{th} measurement on the i^{th} unit, and $\boldsymbol{Y}_i = (Y_{i1} \ldots Y_{in_i})^T$ give the vector of measurements on the i^{th} unit, and $\boldsymbol{Y} = (\boldsymbol{Y}_1^T \ldots \boldsymbol{Y}_m^T)^T$ give the complete vector of measurements. Let $\mu_{ij} = \mathrm{E}(Y_{ij})$ and similarly $\boldsymbol{\mu}_i = \mathrm{E}(\boldsymbol{Y}_i)$ and $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{Y})$. $\boldsymbol{Y}$ is assumed to have a multivariate normal distribution with mean $\boldsymbol{\mu} = \boldsymbol{X\beta}$ and $n \times n$ variance matrix $\boldsymbol{V}$

The model for Y_{ij} as given by Diggle (1988) can be written as

$$Y_{ij} = \mu_{ij} + Z_{ij} + U_i + W_i(t_{ij}) \tag{1.3.5}$$

where Z_{ij} are independent and identically distributed (i.i.d) $N(0, \tau^2)$ random variates representing measurement error, the U_i are iid $N(0, \nu^2)$ variates representing the variation in average response between units, and $W_i(t)$ is an independent stationary Gaussian process representing the serial correlation between measurements on the same unit, with $E[W_i(t)] = 0$ and $\mathrm{cov}\,(W_i(t), W_i(s),\ =) \sigma^2\rho(|t-s|)$, with t_{ij} representing the time at which measurement Y_{ij} is taken.

This results in the variance covariance matrix $\mathrm{var}\,(\boldsymbol{Y}) = \boldsymbol{V}$ where $\boldsymbol{V}$ is a block diagonal $n \times n$ matrix with non zero entries being the $n_i \times n_i$ submatrices $\mathrm{var}\,(\boldsymbol{Y}_i) = \boldsymbol{V}_i$ where

$$\boldsymbol{V}_i = \tau^2\boldsymbol{I} + \nu^2\boldsymbol{J} + \sigma^2\boldsymbol{R}(\boldsymbol{t}_i)$$

where $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{J}$ is a square matrix with all elements equal to 1, and $\boldsymbol{R}(\boldsymbol{t}_i)$ is a symmetric covariance matrix with $(k,l)^{th}$ element given by $\rho(|t_k - t_l|)$ and $\boldsymbol{t}_i = (t_{i1} \ldots t_{in_i})^T$. Diggle (1988) suggest using a continuous time form of a first order autoregressive process for $\rho(u)$. That is

$$\rho(u) = \phi^u \quad \text{where} \quad \phi = \exp(-\alpha)$$

If $\sigma^2 = 0$, (1.3.5) reduces to the uniform correlation (or split plot in time) model and if additionally $\nu^2 = 0$, the model reduces further to the classical linear model with

independent errors.

The approach of Diggle (1988) for the analysis of repeated measurements provides a flexible but economical specification of the covariance structure, enabling the covariance to be modelled using only 4 parameters. It enables a more accurate specification of the covariance structure than the uniform correlation model while not requiring the large number of parameters of the unstructured covariance model. The method allows irregularly spaced measurement times and differing sets of measurement times for the different experimental units. The model also allows for a flexible specification of the mean response profile to reflect treatment trends over time and differences between treatments in these trends. Diggle (1988) proposed using the semi-variogram as a diagnostic tool to determine the appropriate covariance structure in the modelling process.

Random regression models

Another approach which implicitly models the covariance structure between repeated measurements is that of random regression (or random coefficient regression) introduced by Laird & Ware (1982).

Random regression analysis is the standard approach for the genetic analysis of repeated animal measurements over time. This form of analysis is commonly used to model lactation curves and cattle growth data (Meyer, 1998, Meyer, 1999, Jensen, 2001, Strabel et al., 2005). Schaeffer (2004) provides a review of applications of random regression models in animal breeding.

Random regression with random coefficients is a method that models the covariance structure of repeated measures data, allowing the unit random effects to vary over time. Random regression models assume an underlying average profile over time and allow subject effects to be random deviations from this mean trend. Laird & Ware (1982) introduced random regression models as mixed models with fixed parameters at the overall mean level and random parameters at the individual level. By treating the regression coefficients as random effects a matrix of covariances between random regression coefficients is implied. These covariances implicitly model the covariance structure between all measurements along the profile (Jamrozik & Shaeffer, 1997).

If $\boldsymbol{y}_i$ represents the measurements on individual i at times $\boldsymbol{t}_i$ then a standard linear random coefficient regression model can be written as

$$\begin{aligned} \boldsymbol{y}_i &= (\tau_1 + u_{i1})\mathbf{1} + (\tau_2 + u_{i2})\boldsymbol{t}_i + \boldsymbol{e}_i \\ &= \tau_1\mathbf{1} + \tau_2\boldsymbol{t}_i + u_{i1}\mathbf{1} + u_{i2}\boldsymbol{t}_i + \boldsymbol{e}_i \end{aligned}$$

which specifies an overall mean linear profile with fixed population intercept and slope given by τ_1 and τ_2 and the individual subjects having random intercepts (u_{i1}) and slopes (u_{i2}) that vary about the mean parameters. In addition allowance is made for these intercepts and slopes to be correlated within subjects. The random intercept and slope effects are assumed to be normally distributed with zero mean and variance matrix given

by

$$\boldsymbol{u}_i = \begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{G}_{rr} = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix} \right]$$

The error vector $\boldsymbol{e}_i$ is assumed independent of the random effects $\boldsymbol{u}_i$ and to be Normally distributed with mean $\mathbf{0}$ and variance matrix $\sigma^2 \boldsymbol{I}$. We can write the above model in matrix notation as

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\tau} + \boldsymbol{Z}_i \boldsymbol{u}_i + \boldsymbol{e}_i$$

where $\boldsymbol{u}_i$ is the vector of random effects with $\text{var}(\boldsymbol{u}_i) = \boldsymbol{G}_{rr}$ and $\boldsymbol{e}_i \sim N(0, \sigma^2 \boldsymbol{I}_{n_i})$. The within subject variance matrix which gives the covariance between any two measurements on subject i is given by

$$\text{var}(\boldsymbol{y}_i) = \boldsymbol{\Sigma}_i = \boldsymbol{Z}_i \boldsymbol{G}_{rr} \boldsymbol{Z}_i^T + \sigma^2 \boldsymbol{I}_{n_i}$$

The simple linear random regression model can be extended to handle non-linear trends in the data by using orthogonal polynomials, such as Legendre polynomials (Meyer & Kirkpatrick, 2005). These polynomials are able to model a variety of curves but often have problems with estimation when high order polynomials are involved. An alternative, more flexible specification is to use splines, for example cubic smoothing splines (Verbyla et al., 1999 and White et al., 1999). Meyer (2005) provides an account of recent advances in random regression analysis methods, in particular the incorporation of B-splines as an alternative to orthogonal polynomials.

Meyer & Kirkpatrick (2005) present a review of quantitative genetics concepts and how they relate when the trait of interest is a curve or trajectory, rather than individual values. They conclude that random regression models provide a powerful approach to quantitative genetic analyses of such data and that standard quantitative genetic concepts extend readily to traits that are represented by curves. They also show that the models may be readily incorporated in the linear mixed model.

Cubic smoothing splines

Cubic smoothing splines provide a flexible approach to modelling longitudinal data. The ability to formulate the cubic smoothing spline as a linear mixed model (Verbyla et al., 1999), means that they are an ideal approach to modelling variety selection data from multiple harvests over time.

Green & Silverman (1994) discuss the properties of cubic smoothing splines. In summary, suppose there are real numbers $t_1, \ldots, t_n$ on the interval $[a, b]$ with $a < t_1 < t_2 \ldots < t_n < b$. A function f is defined as a cubic spline on the interval $[a, b]$ if f satisfies the following conditions. Firstly f is a cubic polynomial on each of the intervals $[a, t_1), [t_1, t_2), \ldots, (t_n, b]$ and secondly, the cubic polynomial sections join together at the points t_i so that the function f, it's first and second derivatives are continuous at each t_i

and on the whole interval $[a, b]$. A natural cubic spline on $[a, b]$ is a cubic spline with the added property that it's second and third derivatives are zero at the end points a and b, implying that the spline is linear when it passes through the end points.

If a natural cubic spline f is defined at design points $t_1, \ldots, t_n$ (called knots), so that $f_i = f(t_i)$ and the second derivatives of the spline at the knot points are given by $\gamma_i = f''(t_i)$, Green & Silverman (1994) show that, given the vectors $\boldsymbol{f} = (f_1 \ldots f_n)^T$ and $\boldsymbol{\gamma} = (\gamma_2 \ldots \gamma_{n-1})$, the spline can be calculated at points other than t_i.

To see this, define $h_j = t_{j+1} - t_j$, for $j = 1, 2, \ldots, n-1$ and define banded matrices, $\boldsymbol{Q}$ and $\boldsymbol{G}_s$ of dimension $n \times (n-2)$ and $(n-2) \times (n-2)$ where $\boldsymbol{Q}$ and $\boldsymbol{G}_s$ have non zero elements given by

$$\begin{aligned} \boldsymbol{Q}_{ii} &= \frac{1}{h_i} \\ \boldsymbol{Q}_{i+1,i} &= -(\frac{1}{h_i} + \frac{1}{h_{i+1}}) \\ \boldsymbol{Q}_{i+2,i} &= \frac{1}{h_{i+1}} \end{aligned} \tag{1.3.6}$$

and

$$\begin{aligned} \boldsymbol{G}_{s;i,i+1} &= G_{s;i+1,i} = \frac{h_{i+1}}{6} \\ \boldsymbol{G}_{s;ii} &= \frac{h_i + h_{i+1}}{3} \end{aligned} \tag{1.3.7}$$

for $i = 1, 2, \ldots, n-2$, with all other elements equal to zero.

Green & Silverman (1994) show that the vector of second derivatives of f at design points $t_2, \ldots, t_{n-2}$ is given by

$$\boldsymbol{\gamma} = \boldsymbol{G}_s^{-1} \boldsymbol{Q}^T \boldsymbol{f}$$

and by definition $\gamma(t_1) = \gamma(t_n) = 0$.

They further show, that for a cubic spline defined on an interval $[t_L, t_R]$, with $h = t_R - t_L$, $f_L = f(t_L)$, $f_R = f(t_R)$, $\gamma_L = \gamma(t_L)$ and $\gamma(t_R)$, an expression for $f(t)$ is given by

$$\begin{aligned} f(t) &= \frac{(t - t_L) f_R + (t_R - t) f_L}{h} \\ &- \frac{1}{6}(t - t_L)(t_R - t)[(1 + \frac{t - t_L}{h})\gamma_R + (1 + \frac{t_R - t}{h})\gamma_L] \end{aligned}$$

Using this expression for each interval defined by the knot points of the spline, values for the spline may be obtained over the complete interval $[t_1, t_n]$.

Verbyla et al. (1999) show, that given observations $y_i, i = 1, 2, \ldots, n$ at points $t_i, t_1 < t_2 < \ldots < t_n$ where y_i can be expressed as $y_i = f(t_i) + e_i$, where $f(t)$ is a smooth unknown function and $e_i \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{R})$, in vector form $\boldsymbol{y} = \boldsymbol{f} + \boldsymbol{e}$, then the vector $\boldsymbol{f}$ that maximizes

the penalized log-likelihood

$$l = -\frac{1}{2}\log\det\sigma^2\boldsymbol{R} - \frac{1}{2\sigma^2}\left[(\boldsymbol{y}-\boldsymbol{f})^T\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{f}) + \lambda_s\int f''(t)^2 dt\right]$$

is a cubic smoothing spline, with smoothing parameter λ_s which specifies the amount of smoothing. They further show that the cubic smoothing spline estimate of $\boldsymbol{f}$ at the design points t_i, can be written as

$$\tilde{\boldsymbol{f}} = \boldsymbol{X}_s\hat{\boldsymbol{\beta}}_s + \boldsymbol{Z}_s\tilde{\boldsymbol{u}}_s$$

where

$$\begin{aligned}\hat{\boldsymbol{\beta}}_s &= (\boldsymbol{X}_s^T\boldsymbol{H}^{-1}\boldsymbol{X}_s)^{-1}\boldsymbol{X}_s^T\boldsymbol{H}^{-1}\boldsymbol{y}\\ \tilde{\boldsymbol{u}}_s &= (\boldsymbol{Z}_s^T\boldsymbol{R}^{-1}\boldsymbol{Z}_s + \lambda_s\boldsymbol{G}_s^{-1})^{-1}\boldsymbol{Z}_s^T\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{X}_s\hat{\boldsymbol{\beta}}_s)\end{aligned}$$

where $\boldsymbol{X}_s$ is a $n\times 2$ matrix whose columns are a vector of 'ones' and the vector $(t_1, t_2, \ldots, t_n)^T$, $\boldsymbol{Z}_s = \boldsymbol{Q}(\boldsymbol{Q}^T\boldsymbol{Q})^{-1}$ where $\boldsymbol{Q}$ is defined in (1.3.6) and $\boldsymbol{G}_s$ is defined in (1.3.7) above.

Therefore the cubic smoothing spline can be written as the sum of an estimated straight line ($\boldsymbol{X}_s\boldsymbol{\beta}_s$), plus a random component ($\boldsymbol{Z}_s\boldsymbol{u}_s$), that is

$$\boldsymbol{f} = \boldsymbol{X}_s\boldsymbol{\beta}_s + \boldsymbol{Z}_s\boldsymbol{u}_s \tag{1.3.8}$$

where $\boldsymbol{u}_{s_i} \sim N(0, \sigma_s^2\boldsymbol{G}_s)$, and hence can be incorporated into the mixed model framework.

1.3.5 Spatio-temporal models

The combination of the two aspects of spatial and temporal correlation in spatio-temporal models has received a large amount of interest over recent years and is also well documented in the literature. Reviews by Sahu & Mardia (2005) and Kyriakidis & Journel (1999) cover a variety of spatio-temporal methods and applications. The book by Finkenstadt et al. (2006) presents spatio-temporal methods in point processes, biological growth modelling, geostatistics and Gaussian Markov random fields. A book of abstracts edited by Sahu (2005) provides an insight into recent spatio-temporal applications from a range of areas such as real estate, environmental monitoring and disease epidemics. In all of these applications the main focus is on how to jointly model the spatial and temporal correlation between measurements.

Many of the methods for spatial analysis can be grouped depending on how the data are observed, that is whether the spatial location is continuously indexed over a region (geostatistical data) or discretely indexed (lattice data, where the lattice may be regular in spacing or irregular). For geostatistical data spatial dependence is specified through a covariance function, which is then used with kriging to predict values at spatial locations where no measurements are taken. For data collected on a lattice, for example yields of trees in an orchard, it is often not necessary or sensible to predict at locations other

than those in the lattice. While geostatistical methods may be used for some lattice data, there are methods that are specifically designed for data observed on a lattice which may be more suitable, especially for irregular lattices (Sain & Cressie, 2007). Hrafnkelsson & Cressie (2003) present a comparison of geostatistical methods and lattice based methods for the analysis of environmental count data measured on a fine grid and concluded that the lattice based methods were suitable for analysis of such data, giving similar results to the geostatistical methods but being faster to implement.

One approach for analysing spatial lattice data is to use spatial autoregressive models which represent the data at a spatial location as a function of neighbouring locations. Two forms of spatial autoregressive models that are commonly used are conditional autoregressive models (CAR) and simultaneous autoregressive models (SAR), (see Cressie, 1993).

Multivariate extensions of the CAR model have been used for analysing spatio-temporal data (Sahu & Mardia, 2005). Waller et al. (1997) implemented a spatio-temporal CAR model for the analysis of disease rates in space and time.

SAR models have also been extended to handle spatio-temporal data (STAR) models (see Banerjee et al., 2004). Pace et al. (2000) implemented these spatio-temporal autoregressive models in a real estate application and demonstrated the benefits of modelling the spatial as well as temporal correlation.

1.3.6 Conditional and simultaneous autoregressive models (CAR and SAR)

Conditional autoregressive models (CAR)

CAR models are defined by conditional distributions as given in Besag (1974).

$$\begin{aligned} \mathrm{E}\left(y_i|y_{-i}\right) &= \mu_i + \sum_{j=1}^{n} c_{ij}(y_j - \mu_j) \\ \mathrm{var}\left(y_i|y_{-i}\right) &= \sigma_i^2 \end{aligned}$$

for $i, j = 1, \ldots, n$ where y_{-i} indicates all y_j such that $j \neq i$.

From the conditional distributions the joint distribution can be specified using the Hammersley-Clifford Theorem and Brook's lemma (Banerjee et al., 2004). It can be shown in the Gaussian case that

$$\boldsymbol{y} \sim N(\boldsymbol{\mu}, (\boldsymbol{I} - \boldsymbol{C})^{-1}\boldsymbol{M})$$

provided $(\boldsymbol{I}-\boldsymbol{C})$ is invertible and $(\boldsymbol{I}-\boldsymbol{C})^{-1}\boldsymbol{M}$ is symmetric and positive definite (Cressie, 1993), where $\boldsymbol{M} = \mathrm{diag}\,(\sigma_1^2, \ldots, \sigma_n^2)$ and $\boldsymbol{C}$ is a $n \times n$ matrix specifying spatial dependencies between locations with $c_{ii} = 0$ and $c_{ij} \neq 0$ if j is a neighbour of i, $j \in N_i$. The requirement that $(\boldsymbol{I} - \boldsymbol{C})^{-1}\boldsymbol{M}$ is symmetric gives the condition $c_{ij}\sigma_j^2 = c_{ji}\sigma_i^2$.

Simultaneous autoregressive models (SAR)

SAR models can be written as

$$y_i = \mu_i + \sum_{j=1}^{n} c_{ij}(y_j - \mu_j) + \epsilon_i$$

for $i = 1, \ldots, n$. Cressie (1993) shows that any SAR model can be expressed as a CAR model but not vice versa. Hence the CAR model is a more general form and therefore preferable.

Link between CAR models and AR1

The CAR model can be expressed in a form that gives the same variance covariance structure as an autoregressive (ar1) process. This link between the two models, plus the fact that the CAR model can be applied to lattice data on a 2d grid in a single non separable model (as an alternative to the separable ar1(Column).ar1(Row) model) makes the CAR model a potentially useful model for modelling spatial data from perennial crop breeding trials. The added attraction is that the CAR model can be extended to the multivariate case and so may provide a method for modelling the multiple measurements on each plot in the lattice.

Stringer (2006) shows the connection between CAR models and autoregressive ar1 models, as used in the analysis of spatial field trials. This link is presented in the following sections.

The covariance matrix for an autoregressive process of order 1 (ar1) is given by

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \phi & \phi^2 & \ldots & \phi^{r-1} \\ \phi & 1 & \phi & \ldots & \phi^{r-2} \\ \phi^2 & \phi & 1 & \ldots & \phi^{r-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{r-1} & \phi^{r-2} & \ldots & \phi & 1 \end{bmatrix}$$

The inverse of this matrix is given by

$$\mathbf{\Sigma}^{-1} = \frac{1}{1-\phi^2} \begin{bmatrix} 1 & -\phi & 0 & \ldots & 0 \\ -\phi & 1+\phi^2 & -\phi & \ldots & 0 \\ 0 & -\phi & 1+\phi^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & -\phi & 1 \end{bmatrix}$$

Cressie (1993) shows that any Gaussian distribution on a finite set of sites ($Y \sim$

$N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be expressed as a CAR model. To see this we let

$$\boldsymbol{M} = \text{diag}\,(\boldsymbol{\Sigma})$$

and

$$\boldsymbol{C} = \boldsymbol{I} - \boldsymbol{M}\boldsymbol{\Sigma}^{-1}$$

therefore

$$\begin{aligned} \boldsymbol{M}\boldsymbol{\Sigma}^{-1} &= \boldsymbol{I} - \boldsymbol{C} \\ \boldsymbol{\Sigma} &= (\boldsymbol{I} - \boldsymbol{C})^{-1}\boldsymbol{M} \end{aligned}$$

If we use this result and note that $\boldsymbol{M} = \text{diag}\,(\boldsymbol{\Sigma}) = \text{diag}\left(\boldsymbol{\Sigma}^{-1}\right)^{-1}$ we can write $\boldsymbol{M}$ and $\boldsymbol{C}$ in terms of the inverse of the ar1 covariance matrix, above. That is

$$\begin{aligned} \boldsymbol{M} &= \text{diag}\left(\frac{1}{1-\phi^2}, \frac{1+\phi^2}{1-\phi^2}, \frac{1+\phi^2}{1-\phi^2}, \ldots, \frac{1}{1-\phi^2}\right)^{-1} \\ &= \text{diag}\left(1-\phi^2, \frac{1-\phi^2}{1+\phi^2}, \frac{1-\phi^2}{1+\phi^2}, \ldots, 1-\phi^2\right) \end{aligned}$$

and $\boldsymbol{C} = \boldsymbol{I} - \boldsymbol{M}\boldsymbol{\Sigma}^{-1}$ where

$$\boldsymbol{M}\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 1 & -\phi & 0 & \ldots & 0 \\ \frac{-\phi}{1+\phi^2} & 1 & \frac{-\phi}{1+\phi^2} & \ldots & 0 \\ 0 & \frac{-\phi}{1+\phi^2} & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & -\phi & 1 \end{bmatrix}$$

hence

$$\boldsymbol{C} = \begin{bmatrix} 0 & \phi & 0 & \ldots & 0 \\ \frac{\phi}{1+\phi^2} & 0 & \frac{\phi}{1+\phi^2} & \ldots & 0 \\ 0 & \frac{\phi}{1+\phi^2} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \phi & 0 \end{bmatrix}$$

Now, $C_{i,i-1} = C_{i,i=1} = \frac{\phi}{1+\phi^2}$ for all $(i, i-1)$ and $(i, i+1)$ terms, except if $i = 1$ or n, where $C_{12} = C_{n,n-1} = \phi = \frac{\phi}{1+\phi^2} + \frac{\phi^3}{1+\phi^2}$. Hence we can write $\boldsymbol{C}$ as

$$\boldsymbol{C} = \eta\boldsymbol{N} + \nu\boldsymbol{J}$$

where $\eta = \frac{\phi}{1+\phi^2}$, $\boldsymbol{N}$ is a neighbour matrix with elements indexed by $(i, i+1)$ and $(i, i-1)$

equal to 1, and all other terms equal to 0, $\nu = \frac{\phi^3}{1+\phi^2}$ and $\boldsymbol{J}$ is a matrix with elements $(1,2)$ and $(n, n-1)$ equal to 1 and 0 otherwise.

Zimmerman & Harville (1991) note that if $\frac{\phi^2}{1-\phi^2}$ is added to the first and last elements of $\boldsymbol{\Sigma}^{-1}$, then it can be shown that $\boldsymbol{M} = \text{diag}\left(\frac{1-\phi^2}{1+\phi^2}\right)$ and $\boldsymbol{C} = \eta \boldsymbol{N}$ where $\eta = \frac{\phi}{1+\phi^2}$.

1.3.7 Multivariate CAR models

As mentioned in the previous section, the CAR model may be extended to the multivariate setting. Multivariate Conditional Autoregressive (MCAR) models can be used to model multivariate spatial data where a p dimensional random variable $\boldsymbol{Y}_i$ is associated with each spatial location on a lattice. This may provide a suitable approach for modelling the spatio-temporal residual correlation structure which arises from the multiple harvest measurements made on spatially referenced field plots in perennial crop breeding trials.

Mardia (1988) introduced MCAR models by generalizing CAR models to the multivariate case. Kim (2001) presented a 'two fold' CAR model which modelled counts for two different types of diseases over each lattice point, but was restricted to the bivariate case. Gelfand & Vounatsou (2003) and Banerjee et al. (2004) built on the work of Mardia (1988) to develop multivariate CAR models for hierarchical modelling, using them in applications such as spatial modelling of child growth, spatial patterns of gene frequency and spatio-temporal survival data. Jin et al. (2005) proposed a generalised hierarchical multivariate CAR model that reduces the computational effort in modeling. Pettit et al. (2002) presented a CAR model for irregularly spaced multivariate data, which models the covariance function in terms of the Euclidean distance between points. Sain & Cressie (2007) proposed a multivariate CAR model which allows for more general forms for the spatial correlation and cross correlations between variables at different sites. This method differs to the previous methods in that it does not force the dependence between different variables at different locations to be symmetric, allowing for more flexible modelling of the spatial dependence structure.

A multivariate CAR (MCAR) model can be written in terms of conditional distributions, as

$$\begin{aligned} \text{E}\left(\boldsymbol{Y}_i | \boldsymbol{Y}_{-i}\right) &= \boldsymbol{\mu}_i + \sum_{j=1}^{n} \boldsymbol{\Lambda}_{ij}(\boldsymbol{Y}_j - \boldsymbol{\mu}_j) \\ \text{var}\left(\boldsymbol{Y}_i | \boldsymbol{Y}_{-i}\right) &= \boldsymbol{\Gamma}_i \end{aligned}$$

where each $\boldsymbol{Y}_i$ is a p dimensional random variable and $\boldsymbol{Y}_{-i}$ indicates all $\boldsymbol{Y}_j$ such that $j \neq i$ for $i, j = 1, \ldots, n$.

Each $\boldsymbol{\Lambda}_{ij}$ are $p \times p$ matrices with $\boldsymbol{\Lambda}_{ii} = -\boldsymbol{I}$ for $i = 1, \ldots, n$ and $\boldsymbol{\Lambda}_{ij} = \boldsymbol{0}$ for j not a neighbour of i. Each $\boldsymbol{\Gamma}_i$ are also $p \times p$ matrices. Two assumptions need to be made on the MCAR model to ensure the existence of a joint distribution for $\boldsymbol{Y}$ (Sain & Cressie, 2007).

The first assumption

$$\boldsymbol{\Lambda}_{ij}\boldsymbol{\Gamma}_j = \boldsymbol{\Gamma}_i{\boldsymbol{\Lambda}_{ji}}^T$$

ensures symmetry of var $(\boldsymbol{Y})$.
The second assumption

$$\text{Block}(-{\boldsymbol{\Gamma}_i}^{-1}\boldsymbol{\Lambda}_{ij}) \text{ or Block}(-\boldsymbol{\Lambda}_{ij}) \text{ is positive definite,}$$

where Block$(\boldsymbol{\Lambda}_{ij})$ is a block matrix with the $(i,j)^{th}$ block $= \boldsymbol{\Lambda}_{ij}$, ensures positive definiteness of var $(\boldsymbol{Y})$.
Then from Mardia (1988)

$$\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma} = [\boldsymbol{\Gamma}(\boldsymbol{I} - \boldsymbol{\Lambda})]^{-1}$, $\boldsymbol{\Lambda}$ is a $np \times np$ matrix with $(i,j)^{th}$ block $= \boldsymbol{\Lambda}_{ij}$, $\boldsymbol{\Lambda}_{ii} = \boldsymbol{0}$ for $i, j = 1, \ldots, n$ and $\boldsymbol{\Gamma}$ is a $np \times np$ block diagonal matrix with $p \times p$ diagonal entries $\boldsymbol{\Gamma}_i$.

1.4 Outline of thesis

The aim of this thesis is to investigate methods for analysing data from multi-harvest, multi-environment variety selection trials in perennial crops.

In Chapter 2 a review of mixed models applicable to multi-harvest, multi-environment trials and their estimation using REML is presented. These models and estimation methods form the basis for the approaches used in the rest of the thesis.

In Chapter 3 the spatial analysis methods of Gilmour et al. (1997) are applied to the analysis of data from individual harvests of the motivating lucerne and chicory data sets. These analyses show that in many trials spatial variation is evident. In some trials the spatial correlation parameters are shown to vary across harvests while in others they are similar. This highlights the need to develop flexible spatial and temporal models for the residual correlation structure in the models for perennial crop data.

In Chapter 4 simulation studies are conducted to investigate the effect of the spatial analysis methods on the estimation of genetic effects and how the estimates from spatial models and RCB compare to the true genetic effects. Simulations are also performed to investigate the impact of fitting measurement error in the spatial model.

In Chapter 5 methods for analysing data from multiple harvests at a single site using separable spatial and temporal residual covariance structures are presented. Models for analysing variety effects over time are developed. These methods are applied to the multi-harvest analysis of the lucerne yield and persistence data from a single site, in Chapter 6.

In Chapter 7 the separable residual models and models for genetic effects for multi-harvest data are extended to the multi-environment situation for the analysis of the combined multi-harvest data across sites. The methods are applied in the multi-environment analysis of the lucerne yield and persistence data.

The assumption of separability of spatial and temporal residual correlation processes made in Chapters 5 to 7 is a big assumption which may not always hold in practice. It is likely that the spatial correlation parameters will vary over harvests in many situations. It is therefore desirable to investigate non-separable residual models which allow for varying spatial correlation across the harvest times. In Chapter 8 the Multivariate Autoregressive model (MVAR1) is investigated as a suitable non-separable spatio-temporal model for the modelling the residual correlation structure in one direction (e.g. Row). Code is presented to fit these models and they are applied to both multi-harvest and multi-trait examples using the lucerne yield and persistence data. The models are found to provide a significantly better fit than the separable residual models of Chapters 5 and 6.

In Chapter 9 the non-separable MVAR1 models are extended to the two directional lattice situation (with spatial correlation in both Row and Column directions) using the theory of MCAR models. Code has been written to implement the resulting two dimensional multivariate autoregressive models (2dMVAR1) and they are applied to the lucerne data. These models are compared to the separable models of Chapters 5 and 6 and shown to be a significant improvement in most cases.

The final chapter provides an overall summary of the thesis and discusses further research that may provide further insight into this area of work.

Chapter 2

Review: Statistical models

2.1 Introduction

The methods developed in this thesis are based on the linear mixed model. A linear mixed model has both fixed and random effects and can be regarded as an extension of the classical linear model. In this chapter a linear mixed model for data obtained from multiple harvests made on field trials grown in multiple environments (hereafter referred to as multi-harvest METs) is presented, together with details of estimation using best linear unbiased prediction (BLUP) and residual maximum likelihood (REML).

2.2 Linear mixed model

Consider a series of multi-environment trials (MET) consisting of t trials in which m varieties are grown (not all varieties need to be grown in all trials; i will index the variety ($i = 1, \ldots, m$)), in which multiple harvests are made on each trial. The j^{th} trial consists of n_j plots in a rectangular array consisting of c_j columns by r_j rows ($n_j = c_j r_j$). Denote the number of harvests for the j^{th} trial by h_j and let k refer to the harvests within each trial ($k = 1, \ldots, h_j$). The total number of trial by harvest combinations is $h_+ = \sum_{j=1}^{t} h_j$.

Let $\boldsymbol{y}_{jk}$ be the $n_j \times 1$ vector of observations for harvest k in trial j ($j = 1, \ldots, t$, $k = 1, \ldots, h_j$), ordered as rows within columns and let $\boldsymbol{y}_j$ be the $h_j n_j \times 1$ vector of observations (for example yield) for trial j, ordered as rows within columns within harvests. The data combined across trials is denoted by $\boldsymbol{y} = (\boldsymbol{y}_1^T \boldsymbol{y}_2^T \ldots \boldsymbol{y}_t^T)^T$; this is an $N \times 1$ vector with $N = \sum_{j=1}^{t} h_j n_j$.

A linear mixed model for the vector of observations, $\boldsymbol{y}$ can be written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{Z}_g \boldsymbol{u}_g + \boldsymbol{Z}_o \boldsymbol{u}_o + \boldsymbol{e} \tag{2.2.1}$$

where $\boldsymbol{\tau}$ is a vector of fixed effects with design matrix $\boldsymbol{X}$ (assumed to be of full column rank), $\boldsymbol{u}_g$ is the $h_+ m \times 1$ vector of random variety (or genetic) effects with associated design matrix $\boldsymbol{Z}_g$, $\boldsymbol{u}_o$ is a vector of other random effects with associated design matrix $\boldsymbol{Z}_o$ and $\boldsymbol{e}$ is the $N \times 1$ vector of residuals.

The random effects from the linear mixed model (2.2.1) are assumed to follow a Normal

distribution with zero mean vector and variance-covariance matrix

$$\text{var}\left(\begin{bmatrix} \boldsymbol{u}_g \\ \boldsymbol{u}_o \\ \boldsymbol{e} \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{G}_g(\boldsymbol{\gamma_g}) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_o(\boldsymbol{\gamma_o}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}(\boldsymbol{\phi}) \end{bmatrix}$$

where $\boldsymbol{G}_g$, $\boldsymbol{G}_o$ and $\boldsymbol{R}$ are variance matrices which are functions of the vectors of variance parameters $\boldsymbol{\gamma_g}$, $\boldsymbol{\gamma_o}$ and $\boldsymbol{\phi}$ respectively.

The distribution of the data $\boldsymbol{y}$ is therefore Normal with mean $\boldsymbol{X\tau}$ and variance matrix

$$\text{var}\,(\boldsymbol{y}) = \boldsymbol{H} = \boldsymbol{Z_g G_g Z_g}^T + \boldsymbol{Z_o G_o Z_o}^T + \boldsymbol{R}$$

Variance models for residual effects ($\boldsymbol{R}$)

In some situations the vector of residuals $\boldsymbol{e}$ will be indexed by a factor or factors defining sections in the data. Residuals for different sections may be denoted by $\boldsymbol{e}_i$ and hence $\boldsymbol{e}$ may be written as a series of vectors, $\boldsymbol{e} = (\boldsymbol{e}_1^T \boldsymbol{e}_2^T \ldots \boldsymbol{e}_s^T)^T$. For example, in data from multi-harvest METs these sections may represent different trials, or in some situations harvests within trials. It is assumed that each section has it's own variance matrix but the residuals from the different sections are independent. Hence the variance matrix for the residuals, $\boldsymbol{R}$ can be written as

$$\boldsymbol{R} = \text{diag}\,(\boldsymbol{R_i}) = \begin{bmatrix} \boldsymbol{R}_1 & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_2 & \ldots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \ldots & \boldsymbol{0} & \boldsymbol{R}_s \end{bmatrix}$$

Variance models for genetic effects ($\boldsymbol{G}_g$)

The $h_+m \times 1$ vector of genetic effects, $\boldsymbol{u}_g$ for m varieties in each of the h_+ environments (where environment comprises of harvests and trials) can be considered as a two-dimensional (variety by environment) array of effects, $\boldsymbol{U}_g$ (of m rows and h_+ columns), where $\boldsymbol{u}_g = \text{vec}[\boldsymbol{U}_g]$.

The $h_+m \times h_+m$ variance matrix of the variety by environment effects is assumed to have the separable form

$$\boldsymbol{G}_g = \boldsymbol{G}_e \otimes \boldsymbol{G}_v$$

where $\boldsymbol{G}_e$ is a $h_+ \times h_+$ genetic variance matrix representing the variances and covariances between the harvests and trials, and $\boldsymbol{G}_v$ is a $m \times m$ symmetric positive definite matrix representing the structure for the varieties.

It is commonly assumed that $\boldsymbol{G}_v = \boldsymbol{I}_m$ (that is the varieties are independent but if a pedigree structure is known for the varieties other forms of $\boldsymbol{G}_v$ may be applicable (see Oakey et al., 2006).

Variance models for non genetic random effects ($\boldsymbol{G}_o$)

The vector of non-genetic random effects $\boldsymbol{u}_o$ may include many random effect terms. Hence the vector $\boldsymbol{u}_o$ may be composed of b subvectors, such that, $\boldsymbol{u}_o = (\boldsymbol{u}_{o1}^T \boldsymbol{u}_{o2}^T \dots \boldsymbol{u}_{ob}^T)^T$, where the subvector $\boldsymbol{u}_{oi}$ is the vector of effects for the i^{th} random term and is of size $q_i \times 1$. These subvectors are assumed to be independently normally distributed with variance matrices $\boldsymbol{G}_{oi}$. Therefore the variance matrix for the non genetic random effects, $\boldsymbol{G}_o$, can be written as

$$\boldsymbol{G}_o = \text{diag}\,(\boldsymbol{G}_{oi}) = \begin{bmatrix} \boldsymbol{G}_{o1} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_{o2} & \dots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{G}_{ob} \end{bmatrix}$$

In this thesis a key issue will be the choice of models for $\boldsymbol{G}_g$, $\boldsymbol{G}_o$ and $\boldsymbol{R}$. Different forms for these variance matrices will be discussed in subsequent chapters, depending on the situation involved.

2.3 Estimation and prediction

Estimation in the linear mixed model involves estimating the fixed and random effects, τ and $\boldsymbol{u}_g$ and $\boldsymbol{u}_o$, and the variance parameters, $\boldsymbol{\kappa} = (\boldsymbol{\gamma_g}^T \boldsymbol{\gamma_o}^T \boldsymbol{\phi}^T)^T$. This involves two linked processes, namely the estimation of the fixed and random effects for given $\boldsymbol{\kappa}$, which is done using Best Linear Unbiased Prediction (BLUP) (first introduced by Henderson, 1950), and the estimation of the variance parameters $\boldsymbol{\kappa}$ using Residual Maximum Likelihood (REML) (Patterson & Thompson, 1971). Thompson (2008) presents a review of estimation in mixed models and how the methods have evolved. In this chapter some of the main concepts of BLUP and REML and their derivations will be presented.

In order to simplify the account of the estimation of parameters in the linear mixed model, the mixed model of (2.2.1) is condensed to be written in the general form as

$$\boldsymbol{y} = \boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{e}$$

where $\boldsymbol{\tau}$ is the vector of p fixed effects with design matrix $\boldsymbol{X}$ (assumed to be of full column rank), $\boldsymbol{u} = (\boldsymbol{u}_g^T \boldsymbol{u}_o^T)^T$ is the vector of b random variety and other non genetic effects, with associated design matrix $\boldsymbol{Z} = [\boldsymbol{Z}_g \boldsymbol{Z}_o]$, and $\boldsymbol{e}$ is the vector of residuals.

The random effects are assumed to follow a Normal distribution with zero mean vector and variance-covariance matrix

$$\text{var}\left(\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{G}(\gamma) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}(\boldsymbol{\phi}) \end{bmatrix}$$

where $\boldsymbol{G} = \text{diag}\,(\boldsymbol{G}_g, \boldsymbol{G}_o)$ and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_g^T \boldsymbol{\gamma}_o^T)^T$ is the vector of variance parameters associated with $\boldsymbol{u}$ (partitioned into genetic and other non genetic variance parameters) and the vector

of variance parameters associated with $\boldsymbol{e}$ is given by $\boldsymbol{\phi}$.

The distribution of $\boldsymbol{y}$ is therefore Normal with mean $\boldsymbol{X\tau}$ and variance matrix

$$\text{var}\,(\boldsymbol{y}) = \boldsymbol{H} = \boldsymbol{ZGZ}^T + \boldsymbol{R}$$

2.3.1 Best linear unbiased prediction

Best linear unbiased prediction of the fixed and random effects in the linear mixed model (2.2.1) results in the best linear unbiased estimates (BLUEs) of the fixed effects and the best linear unbiased predictors (BLUPs) of the random effects. Note there is a distinction made between the *estimation* of fixed effects and *prediction* of random effects. This is denoted by using $\hat{\boldsymbol{\tau}}$ to represent the BLUE of $\boldsymbol{\tau}$ and $\tilde{\boldsymbol{u}}$ to represent the BLUP of $\boldsymbol{u}$.

Properties of the BLUPs of the random effects are discussed by Robinson (1991) and Thompson (2008). They are *linear* functions of the data $\boldsymbol{y}$, they are *unbiased* in that $E(\tilde{\boldsymbol{u}} - \boldsymbol{u}) = 0$, and they are *best* in that of all unbiased predictors they minimize the mean squared error (MSE) between the true and predicted effects.

A number of derivations of Best Linear Unbiased Prediction (BLUP) are discussed in Robinson (1991), including those given by Henderson (1963) and Henderson (1950).

Henderson (1963) show that the predictor of the of the linear combination $\boldsymbol{k}_1^T\boldsymbol{\tau} + \boldsymbol{k}_2^T\boldsymbol{u}$ of fixed and random effects (where $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$ are $p \times 1$ and $b \times 1$ vectors respectively), which has minimum MSE among the class of unbiased predictors, is given by $\boldsymbol{k}_1^T\hat{\boldsymbol{\tau}} + \boldsymbol{k}_2^T\tilde{\boldsymbol{u}}$, where $\hat{\boldsymbol{\tau}}$ is a solution to the generalised least squares (GLS) equations

$$\hat{\boldsymbol{\tau}} = (\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{y} \tag{2.3.2}$$

and

$$\tilde{\boldsymbol{u}} \;=\; \boldsymbol{GZ}^T\boldsymbol{H}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\tau}}) \tag{2.3.3}$$

where $\tilde{\boldsymbol{u}}$ can be written as

$$\tilde{\boldsymbol{u}} \;=\; \boldsymbol{GZ}^T\boldsymbol{P}\boldsymbol{y}$$

where

$$\boldsymbol{P} = \boldsymbol{H}^{-1} - \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1} \tag{2.3.4}$$

To see this, let $\boldsymbol{a}^T\boldsymbol{y}$ be a linear unbiased predictor of $\boldsymbol{k}_1^T\boldsymbol{\tau} + \boldsymbol{k}_2^T\boldsymbol{u}$ so that

$$\begin{aligned} E(\boldsymbol{a}^T\boldsymbol{y}) &= E(\boldsymbol{k}_1^T\boldsymbol{\tau} + \boldsymbol{k}_2^T\boldsymbol{u}) \\ \boldsymbol{a}^T E(\boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{e}) &= E(\boldsymbol{k}_1^T\boldsymbol{\tau}) + 0 \\ \boldsymbol{a}^T\boldsymbol{X\tau} &= \boldsymbol{k}_1^T\boldsymbol{\tau} \end{aligned}$$

which implies

$$\boldsymbol{a}^T\boldsymbol{X} = \boldsymbol{k}_1^T$$

and

$$\boldsymbol{k}_1 = \boldsymbol{X}^T\boldsymbol{a} \tag{2.3.5}$$

Therefore, clearly not all values of $\boldsymbol{k}_1$ will result in the correct expected value and there is a constraint that needs to be imposed.

To obtain the *best* linear unbiased predictor, the mean square error (MSE) needs to be minimized. The MSE is given by

$$\begin{aligned}
MSE &= \mathrm{E}\left((\boldsymbol{a}^T\boldsymbol{y} - \boldsymbol{k}_1^T\boldsymbol{\tau} - \boldsymbol{k}_2^T\boldsymbol{u})^2\right) \\
&= \mathrm{E}\left((\boldsymbol{a}^T\boldsymbol{y} - \boldsymbol{k}_2^T\boldsymbol{u})^2\right) - 2\mathrm{E}\left((\boldsymbol{a}^T\boldsymbol{y} - \boldsymbol{k}_2^T\boldsymbol{u})(\boldsymbol{k}_1^T\boldsymbol{\tau})\right) + \mathrm{E}\left((\boldsymbol{k}_1^T\boldsymbol{\tau})^2\right) \\
&= \mathrm{var}\left((\boldsymbol{a}^T\boldsymbol{y} - \boldsymbol{k}_2^T\boldsymbol{u})\right) + (\mathrm{E}\left(\boldsymbol{a}^T\boldsymbol{y} - \boldsymbol{k}_2^T\boldsymbol{u}\right))^2 - 2\mathrm{E}\left(\boldsymbol{a}^T\boldsymbol{y} - \boldsymbol{k}_2^T\boldsymbol{u}\right)(\boldsymbol{k}_1^T\tau) + (\boldsymbol{k}_1^T\boldsymbol{\tau})^2 \\
&= \mathrm{var}\left((\boldsymbol{a}^T\boldsymbol{y} - \boldsymbol{k}_2^T\boldsymbol{u})\right) + (\boldsymbol{k}_1^T\boldsymbol{\tau})^2 - 2(\boldsymbol{k}_1^T\boldsymbol{\tau})(\boldsymbol{k}_1^T\boldsymbol{\tau}) + (\boldsymbol{k}_1^T\boldsymbol{\tau})^2 \\
&= \mathrm{var}\left((\boldsymbol{a}^T\boldsymbol{y} - \boldsymbol{k}_2^T\boldsymbol{u})\right) \\
&= \boldsymbol{a}^T\boldsymbol{H}\boldsymbol{a} + \boldsymbol{k}_2^T\boldsymbol{G}\boldsymbol{k}_2 - \boldsymbol{a}^T\mathrm{cov}\left(\boldsymbol{y},\ \boldsymbol{u}\right)\boldsymbol{k}_2 - \boldsymbol{k}_2^T\mathrm{cov}\left(\boldsymbol{u},\ \boldsymbol{y}\right)\boldsymbol{a} \\
&= \boldsymbol{a}^T\boldsymbol{H}\boldsymbol{a} + \boldsymbol{k}_2^T\boldsymbol{G}\boldsymbol{k}_2 - 2\boldsymbol{a}^T\boldsymbol{Z}\boldsymbol{G}\boldsymbol{k}_2
\end{aligned}$$

The MSE needs to be minimized subject to the constraint in (2.3.5). This can be achieved using Lagrangian multipliers. The function to be minimized is

$$B = MSE + 2\boldsymbol{\lambda}^T(\boldsymbol{k}_1 - \boldsymbol{X}^T\boldsymbol{a})$$

where $\boldsymbol{\lambda}$ is the vector of Lagrange multipliers. To minimize B the partial derivatives of B with respect to $\boldsymbol{a}$ and $\boldsymbol{\lambda}$ are taken and equated to zero. Using Result A.3 the partial derivative of B with respect to $\boldsymbol{a}$ is

$$\frac{\partial B}{\partial \boldsymbol{a}} = 2\boldsymbol{H}\boldsymbol{a} - 2\boldsymbol{Z}\boldsymbol{G}\boldsymbol{k}_2 - 2\boldsymbol{X}\boldsymbol{\lambda}$$

and equating to zero and solving for $\boldsymbol{a}$ results in

$$\boldsymbol{a} = \boldsymbol{H}^{-1}\boldsymbol{Z}\boldsymbol{G}\boldsymbol{k}_2 + \boldsymbol{H}^{-1}\boldsymbol{X}\boldsymbol{\lambda} \tag{2.3.6}$$

The partial derivative of B with respect to $\boldsymbol{\lambda}$ is

$$\frac{\partial B}{\partial \boldsymbol{\lambda}} = 2(\boldsymbol{k}_1 - \boldsymbol{X}^T\boldsymbol{a})$$

and equating to zero, using (2.3.6) and solving for $\boldsymbol{\lambda}$ results in

$$\boldsymbol{\lambda} = (\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{k}_1 - \boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{Z}\boldsymbol{G}\boldsymbol{k}_2)$$

Now substituting this into (2.3.6) results in

$$\begin{aligned}
\boldsymbol{a} &= \boldsymbol{H}^{-1}\boldsymbol{Z}\boldsymbol{G}\boldsymbol{k}_2 + \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{k}_1 - \boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{Z}\boldsymbol{G}\boldsymbol{k}_2) \\
&= \boldsymbol{H}^{-1}\boldsymbol{Z}\boldsymbol{G}\boldsymbol{k}_2 + \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{k}_1 - \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{Z}\boldsymbol{G}\boldsymbol{k}_2) \\
&= (\boldsymbol{H}^{-1} - \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1})\boldsymbol{Z}\boldsymbol{G}\boldsymbol{k}_2 + \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{k}_1 \\
&= \boldsymbol{P}\boldsymbol{Z}\boldsymbol{G}\boldsymbol{k}_2 + \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{k}_1
\end{aligned}$$

where $\boldsymbol{P}$ is given by (2.3.4).
Hence

$$\begin{aligned}
\boldsymbol{a}^T\boldsymbol{y} &= \boldsymbol{k}_1^T(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{y} + \boldsymbol{k}_2^T\boldsymbol{G}\boldsymbol{Z}^T\boldsymbol{P}\boldsymbol{y} \\
&= \boldsymbol{k}_1^T\hat{\boldsymbol{\tau}} + \boldsymbol{k}_2^T\tilde{\boldsymbol{u}}
\end{aligned}$$

where the BLUE of $\boldsymbol{\tau}$ is

$$\hat{\boldsymbol{\tau}} = (\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{y}$$

and the BLUP of $\boldsymbol{u}$ is given by

$$\tilde{\boldsymbol{u}} = \boldsymbol{G}\boldsymbol{Z}^T\boldsymbol{P}\boldsymbol{y}$$

The above form for the BLUPs requires the inversion of $\boldsymbol{H}$, the $n\times$ n variance matrix, which may be computationally demanding. An alternative approach that does not require the inverse of $\boldsymbol{H}$ was suggested by Henderson (1950) based on the so-called mixed model equations.

$$\boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X}\hat{\boldsymbol{\tau}} + \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z}\tilde{\boldsymbol{u}} = \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{y} \tag{2.3.7}$$

$$\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X}\hat{\boldsymbol{\tau}} + (\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})\tilde{\boldsymbol{u}} = \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y} \tag{2.3.8}$$

The mixed model equations (2.3.7) and (2.3.8) are most often expressed in matrix form as

$$\begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\tau}} \\ \tilde{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix}$$

To obtain estimates for $\hat{\boldsymbol{\tau}}$ and $\tilde{\boldsymbol{u}}$, the mixed model equations need to be solved. Firstly (2.3.8) is written in terms of $\tilde{\boldsymbol{u}}$ as

$$\tilde{\boldsymbol{u}} = (\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}(\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X}\hat{\boldsymbol{\tau}})$$

and substituting this into (2.3.7) gives

$$\boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X}\hat{\boldsymbol{\tau}} + \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}(\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X}\hat{\boldsymbol{\tau}}) = \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{y}$$

Collecting terms results in

$$\begin{aligned} &\boldsymbol{X}^T(\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1})\boldsymbol{X}\hat{\boldsymbol{\tau}} \\ &= \boldsymbol{X}^T(\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1})\boldsymbol{y} \end{aligned} \tag{2.3.9}$$

and using Result A.1 in the Appendix it can be seen that

$$\begin{aligned} \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1} &= (\boldsymbol{R} + \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T)^{-1} \\ &= \boldsymbol{H}^{-1} \end{aligned}$$

Substituting this result in (2.3.9) gives

$$\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X}\hat{\boldsymbol{\tau}} = \boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{y}$$

resulting in

$$\hat{\boldsymbol{\tau}} = (\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{y}$$

which is the BLUE given in (2.3.2).

Substituting $\hat{\boldsymbol{\tau}}$ into (2.3.8) and rearranging leads to

$$\begin{aligned} \tilde{\boldsymbol{u}} &= (\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}(\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y} - \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{y}) \\ &= (\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1})\boldsymbol{y} \\ &= \boldsymbol{G}\boldsymbol{Z}^T\boldsymbol{H}^{-1}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1})\boldsymbol{y} \qquad \text{(using Result A.2)} \\ &= \boldsymbol{G}\boldsymbol{Z}^T\boldsymbol{P}\boldsymbol{y} \\ &= \boldsymbol{G}\boldsymbol{Z}^T\boldsymbol{H}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\tau}}) \end{aligned}$$

which is the BLUP given in (2.3.3).

Note that the mixed model equations may be written for the expanded linear mixed model (2.2.1) and the resulting BLUEs and BLUPs as

$$\begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z}_o & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z}_g \\ \boldsymbol{Z}_o^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}_o^T\boldsymbol{R}^{-1}\boldsymbol{Z}_o + \boldsymbol{G}_o^{-1} & \boldsymbol{Z}_o^T\boldsymbol{R}^{-1}\boldsymbol{Z}_g \\ \boldsymbol{Z}_g^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}_g^T\boldsymbol{R}^{-1}\boldsymbol{Z}_o & \boldsymbol{Z}_g^T\boldsymbol{R}^{-1}\boldsymbol{Z}_g + \boldsymbol{G}_g^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\tau}} \\ \tilde{\boldsymbol{u}}_o \\ \tilde{\boldsymbol{u}}_g \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}_o^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}_g^T\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix}$$

The best linear unbiased estimates (BLUEs) of the fixed effects are given by

$$\hat{\boldsymbol{\tau}} = (\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{y}$$

and the best linear unbiased predictors (BLUPs) of the random effects are

$$\tilde{\boldsymbol{u}}_o = \boldsymbol{G}_o \boldsymbol{Z}_o^T \boldsymbol{P} \boldsymbol{y}$$
$$\tilde{\boldsymbol{u}}_g = \boldsymbol{G}_g \boldsymbol{Z}_g^T \boldsymbol{P} \boldsymbol{y}$$

These estimates of fixed and random effects require estimates of the variance matrices $\boldsymbol{G_o}$, $\boldsymbol{G_g}$ and $\boldsymbol{R}$ and their associated variance parameters. In practice these parameters are estimated using Residual Maximum Likelihood (REML). On substituting these REML estimates, empirical best linear unbiased estimators and predictors (EBLUEs and EBLUPs) are obtained.

2.3.2 Residual maximum likelihood (REML) estimation

One possible approach for estimating the variance parameters in the linear mixed model is to use the method of maximum likelihood (ML). The log likelihood for $\boldsymbol{y}$ may be maximized with respect to the fixed effects $\boldsymbol{\tau}$ and the variance parameters $\boldsymbol{\kappa}$, to obtain maximum likelihood estimates. However the resulting maximum likelihood estimates of the variance parameters are biased due to the fact that the estimation of the fixed effects is not taken into account when estimating the variance parameters. In some cases ML can lead to inconsistent estimators (Neyman & Scott, 1948).

An alternative method which does account for the estimation of fixed effects is residual maximum likelihood (REML), introduced by Patterson & Thompson (1971). Unlike maximum likelihood, REML estimation does not involve maximizing the whole likelihood of $\boldsymbol{y}$ but the likelihood function based on the residuals after fitting the fixed effects. This approach is preferred to ML due to it resulting in unbiased estimates of the variance parameters.

The following derivation of the REML likelihood function given by Verbyla (1990), involves partitioning the full likelihood of $\boldsymbol{y}$ into a conditional likelihood and a marginal likelihood. The marginal log likelihood is the REML log likelihood. Maximizing this log likelihood provides estimates of the variance parameters $\boldsymbol{\kappa}$.

Verbyla (1990) considers a matrix $\boldsymbol{L} = [\boldsymbol{L_1} \boldsymbol{L_2}]$ where the matrices $\boldsymbol{L_1}$ and $\boldsymbol{L_2}$ are of dimensions $n \times p$ and $n \times (n-p)$ respectively, and satisfy the conditions ${\boldsymbol{L_1}}^T \boldsymbol{X} = \boldsymbol{I_p}$ and ${\boldsymbol{L_2}}^T \boldsymbol{X} = \boldsymbol{0}$. If the data $\boldsymbol{y}$ is transformed to $\boldsymbol{L}^T \boldsymbol{y}$ where

$$\boldsymbol{L}^T \boldsymbol{y} = \begin{bmatrix} \boldsymbol{L}_1^T \boldsymbol{y} \\ \boldsymbol{L}_2^T \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}$$

then the distribution of the transformed data is given by

$$\begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\tau} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{L_1^T H L_1} & \boldsymbol{L_1^T H L_2} \\ \boldsymbol{L_2^T H L_1} & \boldsymbol{L_2^T H L_2} \end{bmatrix} \right)$$

Noting that

$$f(\boldsymbol{y}_1, \boldsymbol{y}_2) = f(\boldsymbol{y}_1|\boldsymbol{y}_2) f(\boldsymbol{y}_2)$$

the distribution of $\boldsymbol{L^T y}$ can be written as the product of the conditional distribution of $\boldsymbol{y_1}$ given $\boldsymbol{y_2}$ (involving $\boldsymbol{\tau}$) and the marginal distribution based on $\boldsymbol{y_2}$ (free of the fixed effects). This marginal distribution is the basis for REML.

For the conditional distribution, the expected value of $\boldsymbol{y}_1$ given $\boldsymbol{y}_2$ is

$$\mathrm{E}\,(\boldsymbol{y_1}|\boldsymbol{y_2}) = \boldsymbol{\tau} + \boldsymbol{L}_1^T \boldsymbol{H} \boldsymbol{L}_2 (\boldsymbol{L}_2^T \boldsymbol{H} \boldsymbol{L}_2)^{-1} \boldsymbol{y}_2$$

The variance of $\boldsymbol{y}_1$ given $\boldsymbol{y}_2$ is

$$\begin{aligned}
\mathrm{var}\,(\boldsymbol{y_1}|\boldsymbol{y_2}) &= [\boldsymbol{L_1^T H L_1} - \boldsymbol{L_1^T H L_2} (\boldsymbol{L_2^T H L_2})^{-1} \boldsymbol{L_2^T H L_1}] \\
&= (\boldsymbol{L}_1^T [\boldsymbol{H} - \boldsymbol{H} \boldsymbol{L}_2 (\boldsymbol{L}_2^T \boldsymbol{H} \boldsymbol{L}_2)^{-1} \boldsymbol{L}_2^T \boldsymbol{H}] \boldsymbol{L}_1) \\
&= (\boldsymbol{L}_1^T \boldsymbol{X} (\boldsymbol{X}^T \boldsymbol{H}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{L}_1) \\
&= (\boldsymbol{X}^T \boldsymbol{H}^{-1} \boldsymbol{X})^{-1}
\end{aligned}$$

using Results A.4 and A.5 in Appendix A.

Hence the conditional distribution of $\boldsymbol{y}_1$ given $\boldsymbol{y}_2$ is

$$\boldsymbol{y}_1|\boldsymbol{y}_2 \sim N(\boldsymbol{\tau} + \boldsymbol{L}_1^T \boldsymbol{H} \boldsymbol{L}_2 (\boldsymbol{L}_2^T \boldsymbol{H} \boldsymbol{L}_2)^{-1} \boldsymbol{y}_2, (\boldsymbol{X}^T \boldsymbol{H}^{-1} \boldsymbol{X})^{-1})$$

The marginal distribution of $\boldsymbol{y}_2$ is given by

$$\boldsymbol{y}_2 \sim N(\boldsymbol{0}, \boldsymbol{L}_2^T \boldsymbol{H} \boldsymbol{L}_2).$$

The log likelihood of $\boldsymbol{y}_1$ given $\boldsymbol{y}_2$ is given by (ignoring terms not involving $\boldsymbol{\tau}$)

$$l = -\frac{1}{2} (\boldsymbol{y}_1 - (\boldsymbol{\tau} + \boldsymbol{L}_1^T \boldsymbol{H} \boldsymbol{L}_2 (\boldsymbol{L}_2^T \boldsymbol{H} \boldsymbol{L}_2)^{-1} \boldsymbol{y}_2)^T (\boldsymbol{X}^T \boldsymbol{H}^{-1} \boldsymbol{X}) (\boldsymbol{y}_1 - (\boldsymbol{\tau} + \boldsymbol{L}_1^T \boldsymbol{H} \boldsymbol{L}_2 (\boldsymbol{L}_2^T \boldsymbol{H} \boldsymbol{L}_2)^{-1} \boldsymbol{y}_2)$$

If we differentiate this log likelihood with respect to $\boldsymbol{\tau}$ and equate to zero, we obtain the following estimate of $\boldsymbol{\tau}$,

$$\begin{aligned}
\hat{\boldsymbol{\tau}} &= \boldsymbol{y}_1 - \boldsymbol{L}_1^T \boldsymbol{H} \boldsymbol{L}_2 (\boldsymbol{L}_2^T \boldsymbol{H} \boldsymbol{L}_2)^{-1} \boldsymbol{y}_2 \\
&= \boldsymbol{L}_1^T (\boldsymbol{y} - \boldsymbol{H} \boldsymbol{L}_2 (\boldsymbol{L}_2^T \boldsymbol{H} \boldsymbol{L}_2)^{-1} \boldsymbol{y}_2) \\
&= \boldsymbol{L}_1^T (\boldsymbol{H} \boldsymbol{H}^{-1} \boldsymbol{y} - \boldsymbol{H} \boldsymbol{L}_2 (\boldsymbol{L}_2^T \boldsymbol{H} \boldsymbol{L}_2)^{-1} \boldsymbol{L}_2^T \boldsymbol{H} \boldsymbol{H}^{-1} \boldsymbol{y}) \\
&= \boldsymbol{L}_1^T (\boldsymbol{H} - \boldsymbol{H} \boldsymbol{L}_2 (\boldsymbol{L}_2^T \boldsymbol{H} \boldsymbol{L}_2)^{-1} \boldsymbol{L}_2^T \boldsymbol{H}) \boldsymbol{H}^{-1} \boldsymbol{y} \\
&= \boldsymbol{L}_1^T (\boldsymbol{X} (\boldsymbol{X}^T \boldsymbol{H}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T) \boldsymbol{H}^{-1} \boldsymbol{y} \quad \text{(using Result A.5 in Appendix A)} \\
&= (\boldsymbol{X}^T \boldsymbol{H}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{H}^{-1} \boldsymbol{y}
\end{aligned}$$

which is the estimate obtained from the mixed model equations (2.3.7) and (2.3.8).

The REML log-likelihood is the marginal log likelihood based on $\boldsymbol{y}_2$. The likelihood is given by

$$L_r = \frac{1}{(2\pi)^{\frac{n-p}{2}}|\boldsymbol{L}_2^T\boldsymbol{H}\boldsymbol{L}_2|^{\frac{1}{2}}}\exp(-\frac{1}{2}\boldsymbol{y}_2^T(\boldsymbol{L}_2^T\boldsymbol{H}\boldsymbol{L}_2)^{-1}\boldsymbol{y}_2)$$

The log-likelihood (ignoring constant terms) is

$$l_r = \log L_r = -\frac{1}{2}\log\left(\left|\boldsymbol{L}_2^T\boldsymbol{H}\boldsymbol{L}_2\right|\right) - \frac{1}{2}\boldsymbol{y}_2^T(\boldsymbol{L}_2^T\boldsymbol{H}\boldsymbol{L}_2)^{-1}\boldsymbol{y}_2$$

which can be written in terms free of $\boldsymbol{L}_2$ (using results in Verbyla, 1990) as

$$l_r = -\frac{1}{2}\log\left(|\boldsymbol{H}|\right) - \frac{1}{2}\log\left(\left|\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X}\right|\right) - \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{y} \tag{2.3.10}$$

where $\boldsymbol{P}$ is given by (2.3.4). The REML log-likelihood (2.3.10) is used to estimate the variance parameters by solving the set of score equations below. With $\boldsymbol{\kappa}$ denoting the vector of variance parameters associated with $\boldsymbol{G}_o$, $\boldsymbol{G}_g$ and $\boldsymbol{R}$, the score for κ_i is given by

$$\begin{aligned}
\boldsymbol{U}(\kappa_i) = \frac{\partial l_r}{\partial \kappa_i} &= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{H}^{-1}\dot{\boldsymbol{H}}_i\right) - \frac{1}{2}\mathrm{tr}\left((\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\frac{\partial(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})}{\partial\kappa_i}\right) - \frac{1}{2}\boldsymbol{y}^T\frac{\partial\boldsymbol{P}}{\partial\kappa_i}\boldsymbol{y} \\
&= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{H}^{-1}\dot{\boldsymbol{H}}_i\right) + \frac{1}{2}\mathrm{tr}\left((\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\dot{\boldsymbol{H}}_i\boldsymbol{H}^{-1}\boldsymbol{X}\right) - \frac{1}{2}\boldsymbol{y}^T\frac{\partial\boldsymbol{P}}{\partial\kappa_i}\boldsymbol{y} \\
&= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{H}^{-1}\dot{\boldsymbol{H}}_i\right) + \frac{1}{2}\mathrm{tr}\left(\boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\dot{\boldsymbol{H}}_i\right) - \frac{1}{2}\boldsymbol{y}^T\frac{\partial\boldsymbol{P}}{\partial\kappa_i}\boldsymbol{y} \\
&= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{H}^{-1}\dot{\boldsymbol{H}}_i - \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\dot{\boldsymbol{H}}_i\right) - \frac{1}{2}\boldsymbol{y}^T\frac{\partial\boldsymbol{P}}{\partial\kappa_i}\boldsymbol{y} \\
&= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{P}\dot{\boldsymbol{H}}_i\right) - \frac{1}{2}\boldsymbol{y}^T\frac{\partial\boldsymbol{P}}{\partial\kappa_i}\boldsymbol{y} \\
&= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{P}\dot{\boldsymbol{H}}_i\right) + \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\dot{\boldsymbol{H}}_i\boldsymbol{P}\boldsymbol{y} \quad \text{(using Result A.3 in Appendix A)} \\
&= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{P}\dot{\boldsymbol{H}}_i\right) + \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{q}_i
\end{aligned} \tag{2.3.11}$$

where $\dot{\boldsymbol{H}}_i = \partial\boldsymbol{H}/\partial\kappa_i$, (the first derivative of $\boldsymbol{H}$ with respect to κ_i), and $\boldsymbol{q}_i = \boldsymbol{H}_i\boldsymbol{P}\boldsymbol{y}$ is called the working variate for κ_i.

The REML estimate of $\boldsymbol{\kappa}$ is obtained by solving the set of score equations $\boldsymbol{U}(\boldsymbol{\kappa}) = 0$. In general numerical methods must be used, for example using Fisher-scoring or Newton-Raphson iterative procedures. These iterative procedures work by taking an initial estimate for the vector of variance parameters $\boldsymbol{\kappa}$, and updating this estimate by an amount which depends on the score of the estimate and a descent direction, which in the Fisher-scoring case is based on the inverse of the Expected Information matrix of $\boldsymbol{\kappa}$ and in the Newton-Raphson case is based on the inverse of the Observed Information matrix. That is, given an initial estimate $\boldsymbol{\kappa}^{(0)}$, both algorithms update the vector $\boldsymbol{\kappa}$ to $\boldsymbol{\kappa}^{(1)}$ using

$$\boldsymbol{\kappa}^{(1)} = \boldsymbol{\kappa}^{(0)} + \boldsymbol{I}(\boldsymbol{\kappa}^{(0)}, \boldsymbol{\kappa}^{(0)})^{-1}\boldsymbol{U}(\boldsymbol{\kappa}^{(0)}) \tag{2.3.12}$$

where $\boldsymbol{U}(\boldsymbol{\kappa}^{(0)})$ is the score vector (2.3.11) and $\boldsymbol{I}(\boldsymbol{\kappa}^{(0)}, \boldsymbol{\kappa}^{(0)})$ is the expected information matrix $\boldsymbol{I}_e$ of $\boldsymbol{\kappa}$ evaluated at $\boldsymbol{\kappa}^{(0)}$ for Fisher-scoring and is the observed information matrix $\boldsymbol{I}_o(\boldsymbol{\kappa}^{(0)}, \boldsymbol{\kappa}^{(0)})$ for the Newton-Raphson procedure. The elements of the observed information matrix, are

$$\begin{aligned}
\boldsymbol{I}_o(\kappa_i, \kappa_j) &= -\frac{\partial^2 l_r}{\partial\kappa_i \partial\kappa_j} \\
&= -\left[-\frac{1}{2}\frac{\partial tr(\boldsymbol{P}\dot{\boldsymbol{H}}_i)}{\partial\kappa_j} + \frac{1}{2}\frac{\partial \boldsymbol{y}^T\boldsymbol{P}\dot{\boldsymbol{H}}_i\boldsymbol{P}\boldsymbol{y}}{\partial\kappa_j}\right] \\
&= -\left[-\frac{1}{2}tr(-\boldsymbol{P}\dot{\boldsymbol{H}}_j\boldsymbol{P}\dot{\boldsymbol{H}}_i + \boldsymbol{P}\dot{\boldsymbol{H}}_{ij}) + \frac{1}{2}\boldsymbol{y}^T(\boldsymbol{P}\dot{\boldsymbol{H}}_{ij}\boldsymbol{P} - 2\boldsymbol{P}\dot{\boldsymbol{H}}_i\boldsymbol{P}\dot{\boldsymbol{H}}_j\boldsymbol{P})\boldsymbol{y}\right] \\
&= -\left[-\frac{1}{2}tr(\boldsymbol{P}\dot{\boldsymbol{H}}_{ij}) + \frac{1}{2}tr(\boldsymbol{P}\dot{\boldsymbol{H}}_j\boldsymbol{P}\dot{\boldsymbol{H}}_i) - \boldsymbol{y}^T\boldsymbol{P}\dot{\boldsymbol{H}}_i\boldsymbol{P}\dot{\boldsymbol{H}}_j\boldsymbol{P}\boldsymbol{y} + \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\dot{\boldsymbol{H}}_{ij}\boldsymbol{P}\boldsymbol{y}\right] \\
&= \frac{1}{2}tr(\boldsymbol{P}\dot{\boldsymbol{H}}_{ij}) - \frac{1}{2}tr(\boldsymbol{P}\dot{\boldsymbol{H}}_i\boldsymbol{P}\dot{\boldsymbol{H}}_j) + \boldsymbol{y}^T\boldsymbol{P}\dot{\boldsymbol{H}}_i\boldsymbol{P}\dot{\boldsymbol{H}}_j\boldsymbol{P}\boldsymbol{y} - \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\dot{\boldsymbol{H}}_{ij}\boldsymbol{P}\boldsymbol{y} \quad (2.3.13)
\end{aligned}$$

where $\dot{\boldsymbol{H}}_{ij} = \partial^2\boldsymbol{H}/\partial\kappa_i\partial\kappa_j$
The elements of the expected information matrix are

$$\begin{aligned}
\boldsymbol{I}_e(\kappa_i, \kappa_j) &= \mathrm{E}\left(-\frac{\partial^2 l_r}{\partial\kappa_i\partial\kappa_j}\right) \\
&= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{P}\dot{\boldsymbol{H}}_i\boldsymbol{P}\dot{\boldsymbol{H}}_j\right) \quad (2.3.14)
\end{aligned}$$

Calculating the trace terms in the above information matrices (2.3.13) and (2.3.14) can be very demanding computationally and sometimes not feasible when dealing with large data sets. To overcome this problem an alternative iterative method called the average information (AI) algorithm Gilmour et al. (1995) may be used. In this method the average information matrix $\boldsymbol{I}_{\boldsymbol{A}}$ is obtained by (approximately) averaging the observed and expected information matrices and approximating $\boldsymbol{y}^T\boldsymbol{P}\dot{\boldsymbol{H}}_{ij}\boldsymbol{P}\boldsymbol{y}$ by it's expectation, $\mathrm{tr}\left(\boldsymbol{P}\dot{\boldsymbol{H}}_{ij}\right)$. The elements of the average information matrix are

$$\begin{aligned}
\boldsymbol{I}_A(\kappa_i, \kappa_j) &= \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\dot{\boldsymbol{H}}_i\boldsymbol{P}\dot{\boldsymbol{H}}_j\boldsymbol{P}\boldsymbol{y} \\
&= \frac{1}{2}\boldsymbol{q}_i^T\boldsymbol{P}\boldsymbol{q}_j
\end{aligned}$$

Hence given an estimate for $\boldsymbol{\kappa} = \boldsymbol{\kappa}^{(m)}$, an update based on the average information matrix is calculated using (2.3.12) with $\boldsymbol{I} = \boldsymbol{I}_A$.

Software and testing

The software used in this thesis to estimate the variance parameters from the linear mixed model using REML, is ASReml in the R environment (Butler et al., 2009). ASReml implements the Average Information (AI) algorithm (Gilmour et al., 1995).

To test the significance of random effects in the linear mixed model the Residual Maximum Likelihood Ratio Test (REMLRT) can be used. The REMLRT may be used

to compare the fit of two models only if they are nested and contain the same fixed effects. For two nested models, M_0 and M_1 with M_1 having p_1 variance parameters and M_0 having p_0 variance parameters, with $p_1 > p_0$, the Residual Maximum Likelihood Ratio Test Statistic (REMLRS) is calculated as $-2(l_0 - l_1)$ where l_0 is the residual log-likelihood for model M_0 and l_1 is the residual log-likelihood for model M_1.

The standard REMLRS is asymptotically distributed as a chi-squared statistic with $p_1 - p_0$ degrees of freedom. If however the test involves a null hypothesis where the parameter is on the boundary of the parameter space the REMLRT needs to be adjusted. For a test of a single variance component the theoretical asymptotic distribution of the REMLRS is a mixture of chi-squared variates where the mixing probabilities are 0.5, one with 0 degrees of freedom (a spike at 0) and the other with 1 degree of freedom. The approximate P value for the REMLRS is $0.5(1 - Pr(\chi^2 \leq d))$ where d is the observed value of the REMLRS (see Stram & Lee, 1994).

To compare the goodness of fit of two models (that may be non-nested) the Akaike Information (AIC) criterion may be used. The AIC value for a model is calculated as $-2(l - p)$, where l is the residual log-likelihood for the model and p is the number of variance parameters in the model. Models with smaller AIC values provide a better fit to the data.

To test the significance of fixed effects in a linear mixed model the Wald test may be used. The traditional Wald statistic is asymptotically distributed as a chi-squared distribution. This test is known to be anti-conservative (Butler et al., 2009). Kenward & Roger (1997) presented an adjusted Wald statistic together with an F approximation which they showed performed well across a variety of situations. The Wald tests used in this thesis are based on these adjusted Wald tests, using "conditional" Wald statistics in Asreml-R (Butler et al., 2009), giving an approximate F test for testing fixed effects.

The approach for model selection used in this thesis has been to use REMLRT (adjusted where necessary) for individual model comparisons involving nested models (presented in the text) and to present AIC values for all models in the tables. This allows for an approach to compare across all models, nested and non-nested. The percent variance explained in factor analytic models has also been investigated to assess model fit. This approach is similar to that of Beeck et al. (2010) and Smith et al. (2007).

Chapter 3

Spatial analysis for perennial crops

3.1 Introduction

In this chapter, a spatial analysis approach based on that of Gilmour et al. (1997) (outlined in Chapter 1) combined with a randomization (or design) based approach will be illustrated in analyses of the motivating lucerne and chicory data sets. The spatial analysis methods will be applied to each individual harvest from each trial.

Spatial analyses in perennial crops using these techniques are not widespread. Some exceptions include Smith et al. (2007) in sugarcane and Jones et al. (2009) in grapes. A number of spatial analyses in perennial crops have been performed using more simplistic methods such as nearest neighbour adjustments (Smith & Casler (2004), Smith & Kearney, 2002). In all mentioned analyses spatial correlation has been evident.

This chapter aims to demonstrate the presence of spatial variation and it's different forms in perennial pasture trials, the improvement in analysis by accounting for this spatial variation and give an insight into the issues that will need to be addressed in modelling the residual correlation structure in an analysis across harvests within a multi-harvest trial. One of the main considerations is how spatial correlation may differ between harvest times within a trial for perennial crops. Similar studies investigating spatial correlation over time in perennial crops have not been found in the literature.

3.2 Mixed model for spatial analysis

The spatial analysis is based on the linear mixed model of (2.2.1).

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{Z}_g\boldsymbol{u}_g + \boldsymbol{Z}_o\boldsymbol{u}_o + \boldsymbol{e}$$

where $\boldsymbol{\tau}$ is a vector of fixed effects with design matrix $\boldsymbol{X}$ (assumed to be of full column rank), $\boldsymbol{u}_g$ is the $h_+m \times 1$ vector of random variety (or genetic) effects with associated design matrix $\boldsymbol{Z}_g$, $\boldsymbol{u}_o$ is a vector of other random effects with associated design matrix $\boldsymbol{Z}_o$ and $\boldsymbol{e}$ is the vector of residuals.

The random effects from the linear mixed model are assumed to follow a Normal

distribution with zero mean vector and variance-covariance matrix

$$\text{var}\left(\begin{bmatrix} \boldsymbol{u}_g \\ \boldsymbol{u}_o \\ \boldsymbol{e} \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{G}_g(\boldsymbol{\gamma_g}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{G}_o(\boldsymbol{\gamma_o}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{R}(\boldsymbol{\phi}) \end{bmatrix}$$

where $\boldsymbol{G}_g$, $\boldsymbol{G}_o$ and $\boldsymbol{R}$ are variance matrices which are functions of the vectors of variance parameters $\boldsymbol{\gamma_g}$, $\boldsymbol{\gamma_o}$ and $\boldsymbol{\phi}$ respectively.

The spatial analysis presented aims to investigate the spatial models for each harvest within each trial. This is analogous to treating each trial by harvest combination as a separate trait and performing a separate spatial analysis on each trait.

Within the framework of the whole multi-environment, multi-harvest data set, presented in Chapter 2, the residual and genetic variance matrices are given by

$$\begin{aligned} \boldsymbol{R} &= \text{diag}\left((\boldsymbol{R}_{jk})\right) \quad (j = 1, \ldots, t, \quad k = 1, \ldots, h_j) \\ &= \text{diag}\left((\boldsymbol{R}_s)\right) \quad (s = 1, \ldots, h_+) \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{G}_g &= \boldsymbol{G}_e \otimes \boldsymbol{G}_v \\ &= \text{diag}\left((\sigma^2_{g_{jk}}) \otimes \boldsymbol{I}_m\right) \quad (j = 1, \ldots, t, \quad k = 1, \ldots, h_j) \\ &= \text{diag}\left((\sigma^2_{g_s}) \otimes \boldsymbol{I}_m\right) \quad (s = 1, \ldots, h_+) \end{aligned}$$

The residual variance structure, $\boldsymbol{R}$, is defined to have h_+ sections relating to the complete set of trial by harvest combinations. Each harvest therefore has it's own residual variance structure and the residuals are assumed independent between harvests.

The genetic variance structure $\boldsymbol{G}_g$ assumes a different genetic variance for each harvest $\sigma_{g_{jk}}$, and also assumes the genetic effects are not correlated between harvests. It also assumes that the varieties are independent ($\boldsymbol{G}_v = \boldsymbol{I}_m$).

The spatial modelling method follows that of Gilmour et al. (1997), (with the spatial variation partitioned into global and local spatial trend and extraneous variation), but with the added inclusion of the randomization or design terms. The approach is a sequential one, commencing with the randomisation based model and then building on this model to include other identified sources of variation. After fitting the randomisation terms the next step in the model building process is to model the local spatial variation. Following Gilmour et al. (1997) this local spatial variation is modelled using a separable correlation structure for the residuals from each trial by harvest combination. Therefore the residual variance matrix for harvest k at trial j is given by

$$\boldsymbol{R}_{jk} = \sigma^2_{jk} \boldsymbol{\Sigma}_{c_{jk}} \otimes \boldsymbol{\Sigma}_{r_{jk}}$$

where $\boldsymbol{\Sigma}_{c_{jk}}$ and $\boldsymbol{\Sigma}_{r_{jk}}$ are the $c_j \times c_j$ and $r_j \times r_j$ spatial correlation matrices for harvest k at trial j corresponding to the column and row dimensions respectively.

Diagnostic tools such as the sample variogram and residual plots (of residuals against row (or column) number) are then used to assess the suitability of the model and to identify any global or extraneous variation. To model global trend and extraneous variation Gilmour et al. (1997) fitted design factors or polynomial or spline functions (Verbyla et al., 1999) to the row or column co-ordinates.

3.2.1 Variogram

The (theoretical) variogram for a two dimensional spatially correlated process $S(.)$ at locations $\boldsymbol{s}$ and $\boldsymbol{t}$ is defined in terms of the semi-variance, which is given by the expected squared difference of the process between locations $\boldsymbol{s}$ and $\boldsymbol{t}$. That is

$$\omega(\boldsymbol{s},\boldsymbol{t}) = \frac{1}{2}E[(S(\boldsymbol{s}) - S(\boldsymbol{t}))^2]$$

If $S(.)$ has zero mean then this gives half the variance of the difference between the two locations. Therefore

$$\omega(\boldsymbol{s},\boldsymbol{t}) = \frac{1}{2}\text{var}\,(S(\boldsymbol{s}) - S(\boldsymbol{t})) = \frac{1}{2}[V(\boldsymbol{s},\boldsymbol{s}) + V(\boldsymbol{t},\boldsymbol{t}) - 2V(\boldsymbol{s},\boldsymbol{t})]$$

where $V(.,.)$ is the covariance function of $S(.)$. It is assumed that the spatial process has a spatially constant mean and variance.

In the case of the spatially correlated ar1(Col).ar1(Row) error process (discussed in Chapter 1) the semi-variance is given by

$$\omega(\boldsymbol{s},\boldsymbol{t}) = \omega(\boldsymbol{l}) = \sigma^2(1 - \phi_r^{|l_r|}\phi_c^{|l_c|})$$

where $\boldsymbol{l} = (l_r, l_c) = |\boldsymbol{s} - \boldsymbol{t}|$ with l_r being the distance between $\boldsymbol{s}$ and $\boldsymbol{t}$ in the row direction and similarly l_c, the separation in the column direction. This semi-variance is a smooth exponentially increasing function in the row and column directions as the distance between plots increases, reaching a plateau at the process variance σ^2. The greater the autoregressive correlation coefficients ϕ_r and ϕ_c, the slower the function rises to the plateau.

In practice the sample variogram is calculated from observed half squared differences of residuals for pairs of plots a given distance apart and is viewed as a 3 dimensional plot. Therefore for the data $\boldsymbol{y}$ given in (2.2.1) the sample variogram for two locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ (where $\{\boldsymbol{s}_i\}$ is the vector of the plot co-ordinates), is defined as

$$\tilde{v}_{ij} = \frac{1}{2}[\tilde{e}_i(s_i) - \tilde{e}_j(s_j)]^2$$

where

$$\tilde{\boldsymbol{e}} = \{\tilde{e}_i(s_i)\} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\tau}} - \boldsymbol{Z}_g\tilde{\boldsymbol{u}}_g - \boldsymbol{Z}_o\tilde{\boldsymbol{u}}_o$$

As data from field trials are usually laid out in a regular array there will be many

values of $\tilde{v}_{ij}$ that have been calculated for the same absolute separation between plots. These values are averaged for each displacement (with mean given by $\bar{v}_{ij}$) and the sample variogram is given as a 3 dimensional plot of the values $(l_{ijr}, l_{ijc}, \bar{v}_{ij})$, where $l_{ijr} = |s_{ir} - s_{jr}|$ and $l_{ijc} = |s_{ic} - s_{jc}|$ are the displacements in the row and column directions respectively.

The sample variogram can be used in the model building process to identify global and extraneous variation. Variograms that depart from a smooth function indicate the presence of extraneous variation while failure of the variogram to reach a plateau in either the row or column direction provides evidence of global trend. The need to account for measurement error can be seen in the variogram when there is a jump discontinuity at zero separation.

Stefanova et al. (2009) suggest using the faces of the sample variogram (that is the slices of the variogram corresponding to zero column or zero row displacement) together with approximate 95% coverage intervals as an extra, diagnostic tool to aid in model selection.

The approach to calculating the sample variogram faces and coverage intervals is detailed in Stefanova et al. (2009). In summary, the approach involves firstly fitting the linear mixed model to the observed data and obtaining estimates of the variance components and fixed effects. Then the sample variogram values are computed for the row and column faces. Using the initial estimated variance parameters, simulated data sets are generated by simulating values for the random effects. The linear mixed model is then fitted to each of these simulated data sets and sample variogram values calculated for each simulation. From these sample variogram values, mean values and 2.5% and 97.5% percentiles can be calculated for each displacement for both the row face and column face. A plot is then constructed of the observed variogram faces together with the means and coverage intervals from the simulations.

Stefanova et al. (2009) propose that using the variogram slices together with 95% coverage intervals provides a more formal approach to model selection than using the standard sample variogram, thus providing a method that results in easier and more informed model selection.

3.3 Spatial analysis of individual harvest data at each site

The spatial analysis methods are applied in the following analyses of the yield and persistence data from individual harvests from the lucerne variety trials and the yield analyses from the chicory variety trial harvests. Each individual harvest from each trial is analysed separately. To present the spatial modelli ng approach in full, the yield data from a single harvest from the Terry Hie Hie site will be analysed in detail. The remaining analyses of the 28 lucerne yield, 25 lucerne persistence and 11 chicory, trial by harvest combinations will be presented in summary.

3.3.1 Lucerne yield

The lucerne yield data was transformed prior to analysis using a cube root transformation ($y_{transf} = (y+1)^{1/3}$). This transformation was required to stabilize the variance and better approximate the assumed Gaussian distribution. The cube root transformation was chosen (over the more commonly used log transformation) after careful consideration of residual plots, in particular Normal Quantile-Quantile plots of residuals from analyses of each harvest. The cube root transformation provided a less severe transformation, that better approximated the Gaussian distribution, than the log transformation. The cube root transformation is often used in transforming volume data and given the lucerne yield data arose from cutting lucerne at a certain height from a plot of set length and width, it was considered sensible for this application.

The first analysis presented here considers data from a single harvest (harvest 5) from the Terry Hie Hie trial. The data from this harvest may be referred to as $\boldsymbol{y}_{jk}$ where this denotes the k^{th} harvest of the j^{th} trial (here $k = 5$ and $j = 5$ if the trials are numbered $1, \ldots, 5$ in alphabetical order).

The sequence of models fitted to the yield (kg DM/ha) data from harvest 5 (24/11/2005) at the Terry Hie Hie site are given in Table 3.1. The first model fitted was a Randomised Complete Block (RCB) model with random variety effects. Hence the mixed model for the data $\boldsymbol{y}_{jk}$ can be given by equation 3.2.1 with $\boldsymbol{u}_o$ containing Rep effects plus random Variety effects, $\boldsymbol{\tau}$ containing a mean harvest term and the errors assumed to be independent with

$$\boldsymbol{R}_{jk} = \sigma^2_{jk}\boldsymbol{I}_{n_j} = \sigma^2_{jk}\boldsymbol{I}_{c_j} \otimes \boldsymbol{I}_{r_j}$$

This residual correlation structure is denoted by id(Col).id(Row), where id refers to the identity matrix. This model reflects the experimental design or randomisation analysis.

The first attempt at improving the fit over and above the randomisation model was to fit a spatial model to the residuals using a separable autoregressive process in the row and column dimensions (ar1(Col).ar1(Row)). This is model 2 in Table 3.1. Hence the mixed model contains the same terms as model 1 (above) but now

$$\boldsymbol{R}_{jk} = \sigma^2_{jk}(\boldsymbol{I}_{c_j} + \sum_{p=1}^{c_j-1} \phi^p_{c_{jk}} \boldsymbol{F}_p) \otimes (\boldsymbol{I}_{r_j} + \sum_{p=1}^{r_j-1} \phi^p_{r_{jk}} \boldsymbol{F}_p)$$

where $\phi_{r_{jk}}$ and $\phi_{c_{jk}}$ are the autoregressive spatial correlation parameters in the row and column directions respectively, $\boldsymbol{F}_p$ is a matrix with ones on the p^{th} sub and super diagonals and zeroes elsewhere.

This model was a significant improvement on the randomisation model as seen by the Residual Maximum Likelihood Ratio Test (REMLRT) of 70.43 on 2 df which is highly significant ($P < 0.001$).

The residual plot from this model (Figure 3.1) shows the residuals being quite different in their mean levels for each column. That is, there is a clear effect associated with the columns, with some columns having very low residuals (e.g. column 1) and other columns

Table 3.1: Summary of models fitted to Terry Hie Hie Harvest 5 (20/04/2004) lucerne yield data.

Model	Local spatial Correlation	Rand terms	Global/extraneous spatial terms[a]	REML log-lik	REMLRT
1.	RCB: id(Col).id(Row)	Rep	-	49.62	
2.	ar1(Col).ar1(Row)	Rep	-	84.83	P<0.001
3.	ar1(Col).ar1(Row)	Rep	ran(Col)	92.40	P<0.001
4.	ar1(Col).ar1(Row)	Rep	lin(Row) + ran(Col)	93.50	P=0.010[b]
5.	id(Col).ar1(Row)	Rep	lin(Row) + ran(Col)	93.49	P=0.980
6.	ar1(Col).id(Row)	Rep	lin(Row) + ran(Col)	81.50	P<0.001

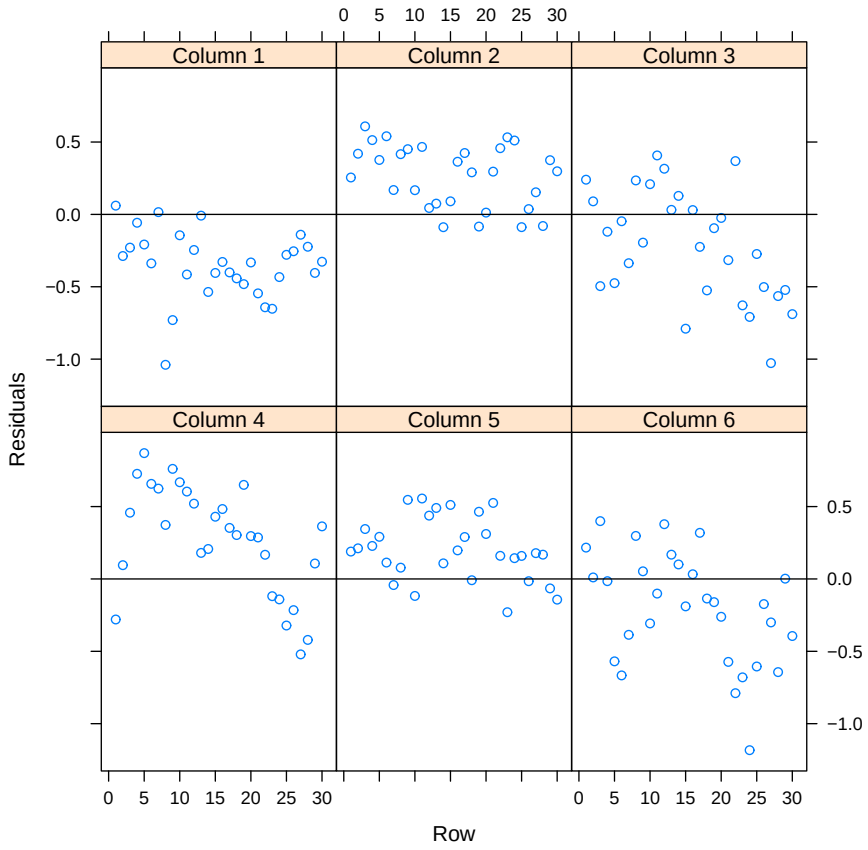
[a] lin(Row) = fixed linear regression over row number; ran(Col) = random effects for columns
[b] tested using Wald test

having high residuals (e.g. column 2). This effect is also clearly seen in the sample variogram (Figure 3.2), where there is a saw-toothed up-down pattern in the plot in the column direction. This pattern is likely to be the result of extraneous variation within the trial. These effects are also seen in the plot of the row face of the sample variogram (corresponding to zero column displacement) in Figure 3.5 (a). The sill of the variogram is much lower than the mean of the simulations and is near the lower boundary of the coverage interval. As shown in Stefanova et al. (2009), this suggests random column effects.

This effect can be accommodated in the model by including a random column effect in $\boldsymbol{u}_o$, in the mixed model. This is fitted as model 3 in Table 3.1. This effect is shown to be highly significant based on the non standard REMLRT statistic (Stram & Lee, 1994) of 15.12 ($P < 0.001$). The resulting sample variogram (Figure 3.3) no longer shows the saw-toothed trend. Note that at large separations in both the row and column dimensions the variogram is based on only a few points and should not be given much weight. The variogram in Figure 3.3 no longer shows the trend due to the column effect but there is evidence of global trend in the row direction with the variogram steadily increasing and failing to reach a plateau in the row direction. This is also evident in the plot of the row face of the sample variogram (Figure 3.5 (b)), where there is a steady increase in semi-variance with increasing row displacement. This is indicative of a linear (global) trend present in the residuals.

This global trend can be accommodated in the model in $\boldsymbol{\tau}$ using a linear regression over row number, denoted lin(Row). Model 4 builds on model 3 by including this fixed linear row regression. Note that models 3 and 4 cannot be compared using the REMLRT test as they do not contain the same fixed effects. However the linear row term can be tested using the Wald test. It is shown to be significant ($P = 0.01$). The variogram from model 4 is given in Figure 3.4. The variogram no longer shows the trends associated with the column effects or the global trend in the row direction. The row and column faces of the variogram (Figure 3.5 (c) and (f)) now lie within the 95% coverage intervals and follow the means of the simulations. The residual spatial autoregressive correlation

Figure 3.1: Plot of residuals (from model 2 in Table 3.1 fitted to lucerne yield data from harvest 5 at Terry Hie Hie) versus row number for each column

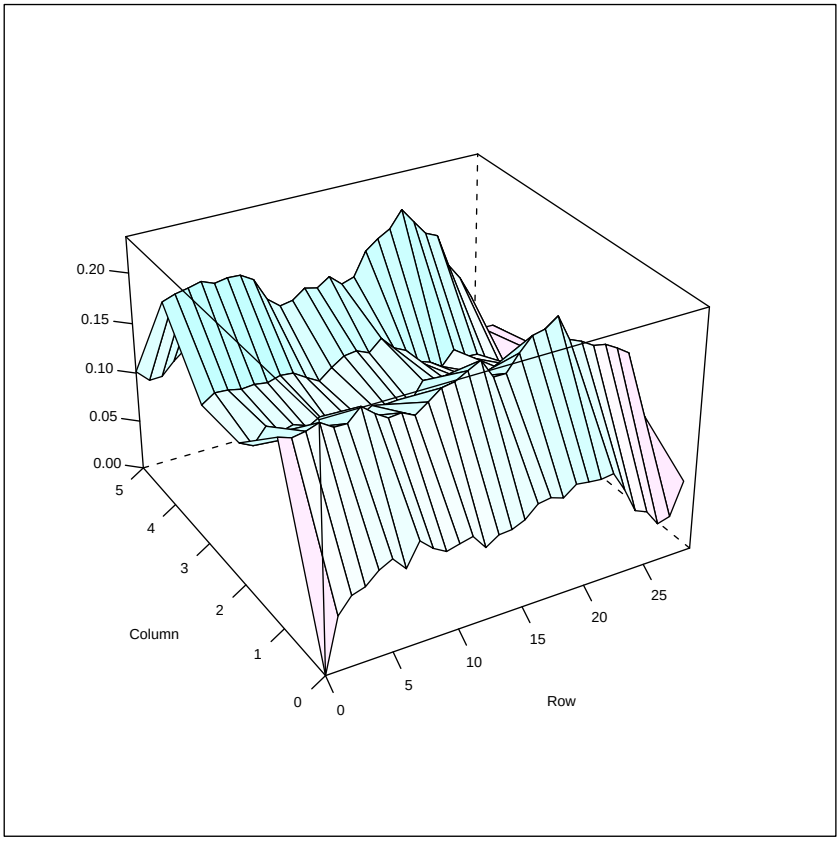


parameters estimated in model 4 are 0.003 and 0.26 for the row and column directions respectively. Tests to assess whether these parameters were significantly different to zero were performed by fitting models 5 and 6. The tests showed the column spatial correlation parameter was not significant but the row correlation parameter was highly significant. Hence the final model for yield at Terry Hie Hie, harvest 5 (model 5) includes random effects for Variety, Rep, Column, a fixed linear Row effect and assumes an id(Col).ar1(Row) residual error process to model the local spatial correlation in the row direction. Therefore the residual variance model fitted in the final model is given by

$$\boldsymbol{R}_{jk} = \sigma^2_{jk}(\boldsymbol{I}_{c_j} \otimes (\boldsymbol{I}_{r_j} + \sum_{p=1}^{r_j-1} \phi^j_{r_{jk}} \boldsymbol{F}_p)$$

Figure 3.6 shows the BLUPs for each variety from the RCB model fitted to the yield data from harvest 5 at Terry Hie Hie, (model 1) plotted against the BLUPs from the spatial model (model 5). In this plot the cut-off lines for the top 20% of varieties (top 12 varieties) under each method are displayed. It is of interest to compare the percentage of varieties in common in the top ranking varieties from the two methods of analysis. There are a number of varieties that are ranked in the top 20% under one approach but not the other.

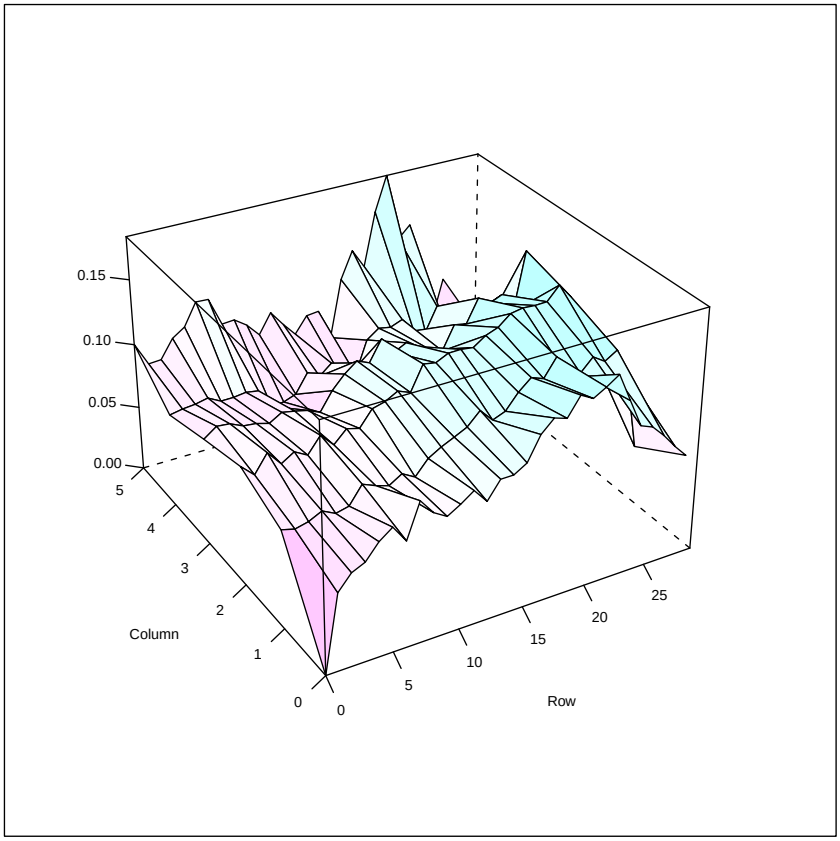
Figure 3.2: Sample variogram of residuals from model 2 (in Table 3.1) fitted to lucerne yield data from harvest 5 at Terry Hie Hie



The analysis of the remaining 27 trial by harvest combinations from the 5 lucerne trials, as detailed in Table 3.2, follows the same modelling approach. The first model fitted in each case was the RCB model. The next step was to build on this randomisation model to fit the ar1(Col).ar1(Row) model to the residuals. Using diagnostic tools such as the variogram and plots of residuals against row number for each column (and column number for each row), the adequacy of the model was assessed for each harvest. Based on these diagnostics further terms were added to the model, where required, to account for global and extraneous spatial variation. These terms are given in Table 3.2 (Linear and Random Row, Random Column, Row1) and are included in $\boldsymbol{u}_o$ or $\boldsymbol{\tau}$ in (3.2.1). The significance of the ar1 spatial autocorrelation parameters in both row and column directions was tested (using Residual Maximum Likelihood Ratio Tests (REMLRT)) and in some cases it was sufficient to model the correlation in one dimension only (hence a id(Col).ar1(Row) or ar1(Col).id(Row) model was fitted to the residuals.

To compare the final "best" spatial model against the RCB model for each harvest, AIC values were calculated for the two models. The model with the smaller AIC value is superior in terms of goodness of fit. In Table 3.2, the AIC values have been presented for each of the best spatial models, as differences from the AIC value for the RCB model. Hence negative values indicate the spatial to be a better fit than the RCB model. In cases

Figure 3.3: Sample variogram of residuals from model 3 (in Table 3.1) fitted to lucerne yield data from harvest 5 at Terry Hie Hie

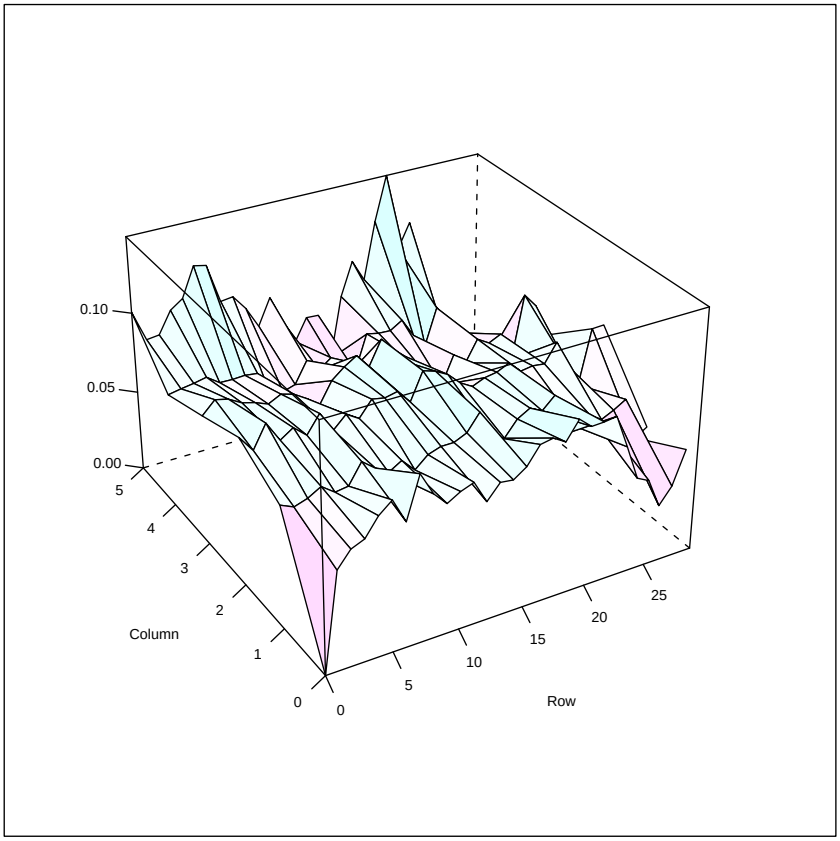


where extra fixed effect terms (e.g. lin(Row)) have been included in the spatial model the AIC value for the RCB model has been calculated with the RCB model also including these same fixed effects. In all but one harvest, the spatial models were a significant improvement on the RCB analysis. This illustrates spatial variation is an important component of lucerne breeding trials that needs to be taken into account.

All three forms of spatial variation, local, global and extraneous (as described by Gilmour et al., 1997) were evident in the lucerne trials. There was global trend in the form of linear row effects present in harvests 2 to 5 at the Leadville site. There was apparent extraneous variation in the data from harvest 10 at Terry Hie Hie. The first row of this trial had very high residuals compared to the other rows. To account for this variation, a factor with two levels (1 for plots in first row and 2 for others) was fitted as a fixed effect.

Table 3.2 also presents the modelled local spatial trends as shown in the column and row autocorrelation parameters. The correlation between adjacent plots was consistently high in the row direction at all harvests at Euloma and Leadville, with correlations ranging from 0.59 to 0.63 at Euloma and 0.31 to 0.67 at Leadville. The other trials had high correlations at some harvests and not others, for example the correlations in the column direction at Tamworth ranged from 0.18 to 0.59.

Figure 3.4: Sample variogram of residuals from model 4 (in Table 3.1) fitted to lucerne yield data from harvest 5 at Terry Hie Hie



Heritabilities have been calculated for the spatial method and the non spatial randomisation method using the following formula (Cullis et al. (2006)),

$$h^2 = 1 - \frac{A}{2\sigma_{gj}^2} \tag{3.3.1}$$

where A is the average pairwise prediction error variance of variety effects and σ_{gj}^2 is the genetic variance.

These heritabilities are also included in Table 3.2. In most cases the genetic variation was very low and for some harvests it was zero hence many of the heritabilities are also zero. In many cases the spatial method provided substantial increases in heritability, for example at harvests 2 to 5 at Leadville where the heritabilities were essentially 0 under the non spatial method and ranged from 0.22 to 0.46 under the spatial analysis method.

Table 3.2: Summary of terms fitted to the lucerne yield data in the best spatial model for each of the 28 lucerne trial by harvest combinations to account for non-genetic extraneous, global and spatial variation in the data. AIC values for each model are given as differences from the AIC for the RCB model. P values (based on the Wald test) are given for the fixed effect terms accounting for global and extraneous trend in the model.

Trial / Harvest	Residual Model (Col.Row)	Lin Row	Ran Row	Ran Col	Row1	Spatial correlation Column	Row	AIC Δ
Euloma 1	ar1.ar1					0.21	0.63	-82.5
Euloma 2	ar1.ar1					0.30	0.63	-85.8
Euloma 3	ar1.ar1	P<0.001				0.24	0.63	-96.8^f
Euloma 4	id.ar1					-	0.59	-61.3
Euloma 5	ar1.ar1					0.26	0.61	-80.5
Euloma 6	id.ar1					-	0.62	-69.3
Leadville 1	id.ar1					-	0.45	-28.8
Leadville 2	id.ar1	P<0.001				-	0.66	-67.3^f
Leadville 3	ar1.ar1	P<0.001				0.23	0.67	-76.0^f
Leadville 4	id.ar1	P<0.001				-	0.59	-55.2^f
Leadville 5	id.ar1	P<0.001		✓		-	0.31	-20.2^f
Sandigo 1	ar1.ar1	P=0.01				0.27	0.60	-66.4^f
Sandigo 2	ar1.ar1		✓	✓		0.32	0.32	-55.1
Tamworth 1	ar1.ar1					0.59	0.45	-88.2
Tamworth 2	ar1.ar1					0.30	0.41	-43.3
Tamworth 3	ar1.ar1					0.18	0.17	-5.1
Tamworth 4	ar1.id			✓		0.23	-	-31.1
Tamworth 5	ar1.ar1			✓		0.44	0.21	-44.9
Terry Hie Hie 1	id.ar1					-	0.27	-10.8
Terry Hie Hie 2	id.ar1			✓		-	0.33	-19.7
Terry Hie Hie 3	id.ar1			✓		-	0.24	-12.8
Terry Hie Hie 4	id.ar1			✓		-	0.11	-24.0
Terry Hie Hie 5	id.ar1	P<0.001		✓		-	0.26	-79.5^f
Terry Hie Hie 6	id.ar1			✓		-	0.37	-57.7
Terry Hie Hie 7	id.ar1					-	0.20	-3.2
Terry Hie Hie 8	id.ar1			✓		-	0.21	-14.9
Terry Hie Hie 9	id.id					-	-	0
Terry Hie Hie 10	id.ar1				P<0.001	-	0.27	-7.3

Lin Row = fixed linear regression over row number; Ran Row = random effects for rows; Ran Col = random effects for columns

^{f} denotes AIC calculated for RCB model with same fixed effects as best spatial model

Figure 3.5: Plot of faces of the sample variogram (solid line) and mean (dotted line) and approximate 95% coverage intervals (dashed lines) for models 2,3,4 (in Table 3.1) fitted to lucerne yield data from harvest 5 at Terry Hie Hie; (a) Model 2 row face, (b) Model 3 row face, (c) Model 4 row face, (d) Model 2 column face, (e) Model 3 column face, (f) Model 4 column face

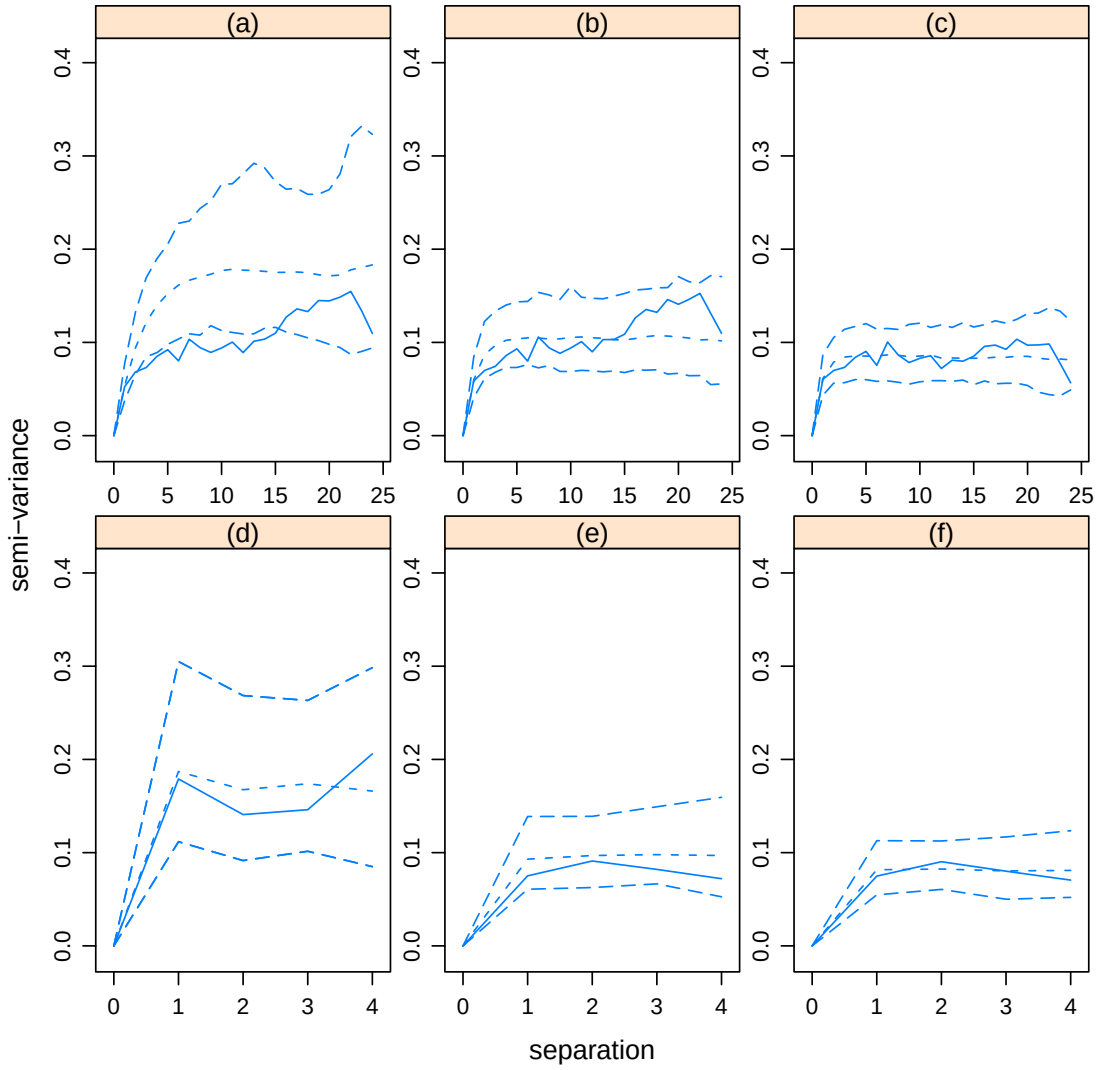


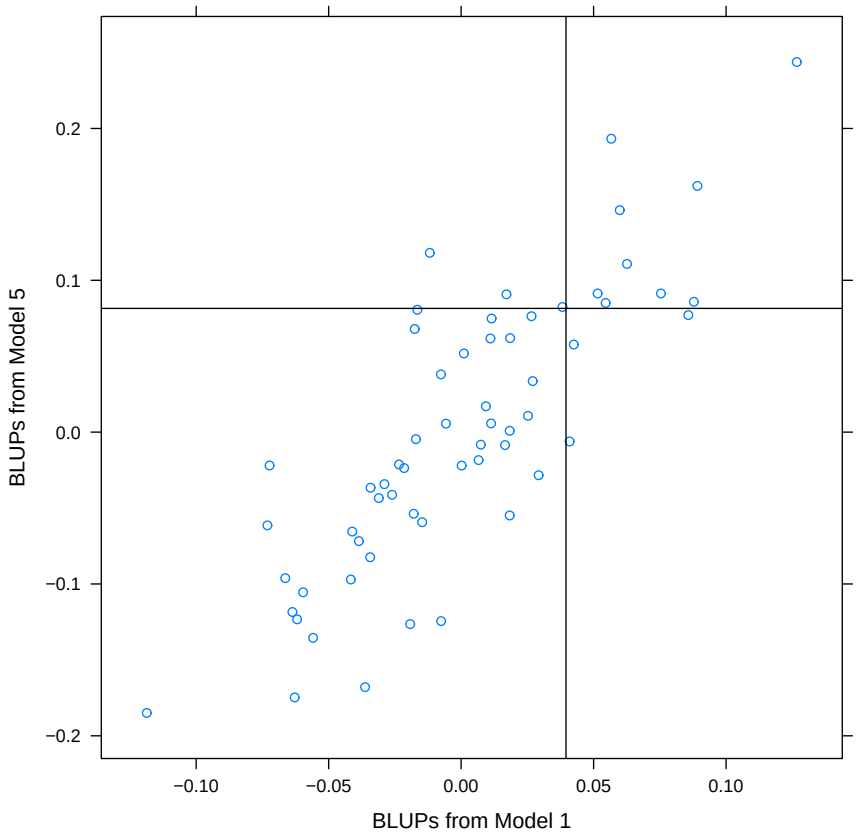
Figure 3.7 presents the residuals from the analysis of each harvest at Terry Hie Hie. It can be seen that the residuals exhibit temporal correlation between harvests. This correlation is generally highest between successive times.

3.3.2 Lucerne persistence

The lucerne persistence data (based on percentages) was transformed prior to analysis, using an empirical logit transformation ($\log((P + 0.5)/(n - P + 0.5))$), where P is the number of squares in the measurement grid (of $n = 100$ squares) that have a plant or plants present. This transformation maps the percentage data from the $(0, 1)$ range to the real line, thereby providing a better approximation to the Gaussian distribution.

The analysis of the lucerne persistence data proceeded as for the yield data (above),

Figure 3.6: BLUPs from model 1 (RCB) vs BLUPs from spatial model (model 5 in Table 3.1) for harvest 5 at Terry Hie Hie. Vertical and horizontal lines show the top 20% of varieties based on each method



commencing with fitting a Randomized Complete Block (RCB) model to data at each harvest and then building on this model using the sequential spatial modelling approach of Gilmour et al. (1997). The terms fitted to account for global and extraneous spatial variation are given in Table 3.4. There were random Column effects identified in 12 of the 25 harvests.

There was significant local spatial correlation between neighbouring plots in 15 of the harvests, mostly in the row direction. These spatial correlation parameters are given in Table 3.4. Once again, the difference between spatial correlations and trend terms between sites and harvests within a site is observed. The AIC values show the spatial model providing a better fit to the data than the RCB model in 21 out of 25 harvests.

3.3.3 Chicory

The chicory yield data was transformed prior to analysis using the cube root transformation ($y_{transf} = (y + 1)^{1/3}$). This transformation was required to stabilize the variance and better approximate the Gaussian distribution. The modelling process for the chicory data followed that of the lucerne data (above).

There was evidence of local spatial correlation with 9 out of 11 harvests having sig-

Table 3.3: Genetic and Residual variances and heritabilities calculated for the RCB and spatial models fitted to the Yield data from 28 lucerne trial by harvest combinations

Trial / Harvest	RCB		Spatial		Heritabilities	
	Genetic	Residual	Genetic	Residual	RCB	Spatial
Euloma 1	4020.33	23649	1091.94	27298	0.33	0.23
Euloma 2	0.01	38688	0.17	42047	0	0
Euloma 3	0.01	102329	0.02	98699	0	0
Euloma 4	0.00	47948	0.01	52571	0	0
Euloma 5	2693.76	44896	690.79	49342	0.14	0.09
Euloma 6	201.95	50488	0.00	29069	0.01	0
Leadville 1	1842.60	18426	3701.16	41124	0.22	0.28
Leadville 2	0.00	16660	1663.68	13864	0	0.46
Leadville 3	0.00	32529	2219.58	20178	0	0.41
Leadville 4	0.00	25195	836.10	16722	0	0.22
Leadville 5	0.01	53354	1855.85	37117	0	0.17
Sandigo 1	1420.25	10925	1035.72	11508	0.29	0.38
Sandigo 2	4148.80	20744	2521.68	12008	0.38	0.41
Tamworth 1	473.68	47368	1990.32	49758	0.03	0.25
Tamworth 2	0.00	25513	0.00	25746	0	0
Tamworth 3	0.00	37990	0.00	38649	0	0
Tamworth 4	0.00	31703	0.00	23796	0	0
Tamworth 5	0.01	41747	1686.70	33734	0	0.19
Terry Hie Hie 1	0.02	50580	0.01	51034	0	0
Terry Hie Hie 2	1635.00	32700	2633.04	32913	0.14	0.23
Terry Hie Hie 3	0.00	33497	0.00	33975	0	0
Terry Hie Hie 4	2103.40	10517	1346.85	8979	0.37	0.31
Terry Hie Hie 5	944.08	11801	1437.16	7564	0.19	0.44
Terry Hie Hie 6	13131.85	11419	13175.36	7486	0.77	0.85
Terry Hie Hie 7	1825.44	15212	655.72	16393	0.27	0.13
Terry Hie Hie 8	1084.02	15486	439.50	14650	0.17	0.09
Terry Hie Hie 9	2289.60	12720	-	-	0.34	-
Terry Hie Hie 10	1936.14	32269	2677.78	19127	0.14	0.33

nificant autocorrelation between plots in the row direction. In these harvests the spatial correlation parameters ranged from 0.24 to 0.64 and are given in Table 3.5. There were no terms identified to account for global trend or extraneous variation. The AIC values show the spatial model providing a better fit to the data than the RCB model in 9 out of 11 harvests.

3.4 Discussion

The analyses presented in this chapter demonstrate the presence of spatial variation in lucerne and chicory pasture trials. This variation occurs in the form of global and extraneous trends and local spatial correlation between neighbouring plots. These different types of spatial variation are unlikely to have been identified using classical randomisation methods or Nearest Neighbour methods of analysis.

Table 3.4: Summary of terms fitted to the Persistence data in the best model for the 26 lucerne trial by harvest combinations, to account for non-genetic extraneous, global and spatial variation in the data. AIC values are given as differences from the AIC for the RCB model. P values (based on the Wald test) are given for the fixed effect terms accounting for global and extraneous trend in the model.

Trial / Harvest	Residual Model (Col.Row)	Lin[a] Row	Ran[a] Row	Ran[a] Col	Spatial correlation Column	Spatial correlation Row	AIC Δ
Euloma 1	id.ar1				-	0.20	-3.2
Euloma 2	ar1.ar1				0.21	0.17	-6.6
Euloma 3	id.ar1	P<0.001			-	0.15	-1.2[f]
Leadville 1	id.id				-	-	0
Leadville 2	id.id			✓	-	-	-11.0
Leadville 3	id.ar1			✓	-	0.21	-10.3
Leadville 4	id.ar1				-	0.34	-9.9
Leadville 5	id.ar1				-	0.32	-10.2
Leadville 6	id.ar1			✓	-	0.20	-94.9
Leadville 7	id.id				-	-	0
Sandigo 1	ar1.ar1				0.23	0.25	-8.6
Sandigo 2	ar1.ar1				0.25	0.36	-15.5
Sandigo 3	id.ar1			✓	-	0.31	-29.1
Sandigo 4	ar1.ar1				0.47	0.45	-44.4
Tamworth 1	id.ar1			✓	-	0.30	-65.2
Tamworth 2	id.id			✓	-	-	-27.4
Tamworth 3	ar1.id			✓	0.33	-	-37.0
Tamworth 4	ar1.id			✓	0.17	-	-15.1
Tamworth 5	id.id			✓	-	-	-32.0
Terry Hie Hie 1	id.id			✓	-	-	-11.1
Terry Hie Hie 2	id.id				-	-	0
Terry Hie Hie 3	id.id				-	-	0
Terry Hie Hie 4	id.id			✓	-	-	-5.9
Terry Hie Hie 5	id.ar1				-	0.17	-1.9
Terry Hie Hie 6	id.id			✓	-	-	-31.2

[a] Lin Row = fixed linear regression over row number; Ran Col = random effects for columns
[f] denotes AIC calculated for RCB model with same fixed effects as best spatial model

Figure 3.7: Pairwise plots of residuals from individual analyses of yield data from harvests at Terry Hie Hie

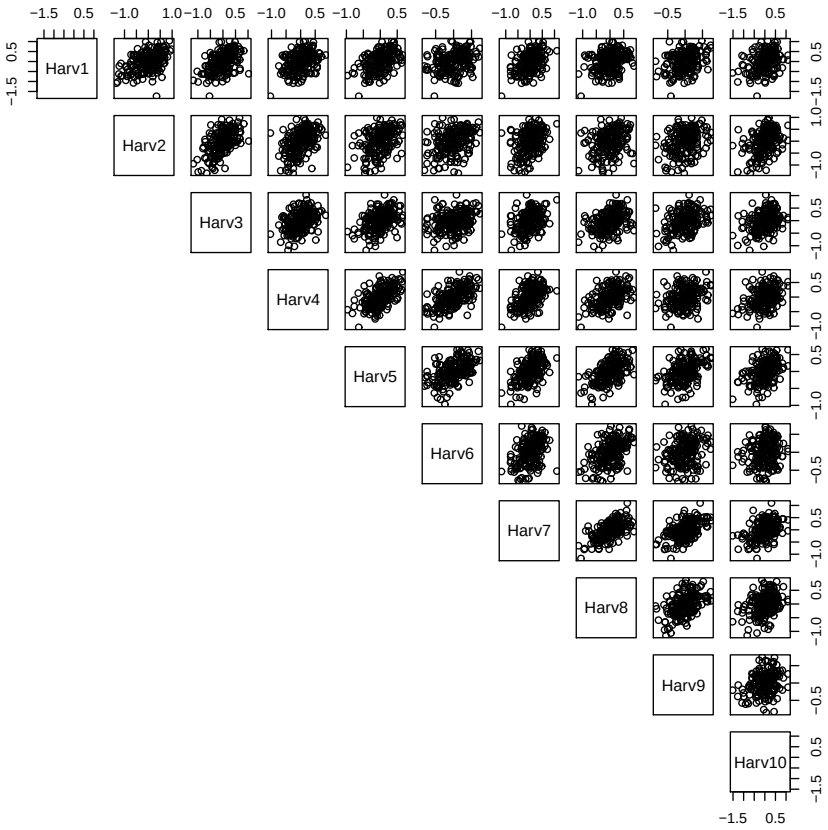


Table 3.5: Summary of terms fitted to the chicory data, to account for non-genetic extraneous, global and spatial variation, in the best model for each of the 11 harvests at Keith. AIC values are given as differences from the AIC for the RCB model. P values (based on the Wald test) are given for the fixed effect terms accounting for global and extraneous trend in the model.

Trial / Harvest	Residual	Spatial correlation		AIC
	Model(Col.Row)	Column	Row	Δ
Keith 1	id.ar1	-	0.31	-2.2
Keith 2	id.ar1	-	0.24	-1.0
Keith 3	id.id	-	-	0
Keith 4	id.ar1	-	0.64	-23.4
Keith 5	id.ar1	-	0.61	-15.0
Keith 6	id.ar1	-	0.32	-1.0
Keith 7	id.id	-	-	0
Keith 8	id.ar1	-	0.32	-2.3
Keith 9	id.ar1	-	0.54	-8.3
Keith 10	id.ar1	-	0.55	-13.9
Keith 11	id.ar1	-	0.61	-19.4

In all analyses the local spatial correlation parameters were positive, reflecting the ideas of Fisher (1935), in that plots that are in close proximity are likely to be more alike than those further apart. There was no evidence of interplot competition which would be indicated by negative spatial correlation parameters (Stringer & Cullis, 2002).

With the lucerne yield data all harvests except two showed local spatial correlation in the row direction. Many harvests also showed local spatial correlation in the column direction, including the harvests at Tamworth where the layout was different to the other trials (with 12 columns at Tamworth rather than 6 at the other sites). The spatial correlation parameters across the harvests ranged from $0 - 0.67$ in the row direction and $0 - 0.59$ in the column direction. Within a trial the spatial correlation parameters were similar across the harvests for some trials, (for example at Euloma where the row spatial correlation parameters ranged from $0.59 - 0.63$), while for other trials the spatial correlation varied across harvests (for example at Tamworth the column correlation parameters ranged from $0.18 - 0.59$). Many harvests showed extraneous variation associated with the columns, and at the Leadville site there was clear global trend in the row direction for all harvests besides the first.

There are many possible explanations for the extraneous and global variation evident in the lucerne trials, depending on the trial and harvest. For example in the analysis of the yield data at Terry Hie Hie the first three harvests show no significant extraneous variation associated with the columns but then four out of the next six harvests had such extraneous variation, with the external columns (1 and 6) having lower residuals than the internal columns. This may be due to the change in harvest area of the plots with harvests 1 to 3 only harvesting a subsection of the length of the plot while harvests 4-10 cut the whole length of the plot (see Table 1.1. Therefore in harvests 4-10 the outside of the trial area (outside of columns 1 and 6) are harvested whereas in harvests 1-3 this is not included. This may have resulted in lower values for the external plots and hence the column effect. The linear row effect for yield at Leadville may be related to the harvesting pattern of the trial as the plots were harvested across columns (all row1 then row2, row3 etc.). Hence the trend could be associated with the time of harvest. Alternatively it could be associated with soil trends or slope of the trial site.

The lucerne persistence data showed local spatial correlation between plots at a number of harvests, especially in the row direction. The spatial correlation parameters across the harvests ranged from $0 - 0.43$ in the row direction and $0 - 0.47$ in the column direction. In general the local spatial correlation observed was lower for the persistence data than the yield data (Tables 3.4 and 3.2). There was also evidence of extraneous variation associated with the columns at many harvests. With the lucerne persistence data, a similar effect of the external columns (to that of the Terry Hie Hie yield) is evident, but for some harvests, e.g. at Sandigo harvest 3, the residuals are lower for the edge columns, whereas for Terry Hie Hie the edge columns have higher residuals than the other columns.

The chicory yield data exhibited significant local spatial correlation in the row direction at many harvests. The spatial correlation parameters varied across the harvests, ranging from $0 - 0.64$. With only four columns in the trial it was decided not to fit local spatial

correlation models in the column direction. There was no obvious extraneous or global spatial variation identified.

It has been shown that one spatial model is not necessarily suitable for all trials and harvests within a trial. Each spatial model for each trial and harvest needs to be formulated using diagnostic aids and tests. There are many reasons why the spatial variation is likely to be different across trials and harvests within a trial. Trial sites will vary in their soil fertility and soil moisture, and management practices are likely to differ between trials, all impacting on the spatial variation. Within a trial the spatial variation may be expected to differ between harvest times due to factors such as seasonal changes, growth phase of the crop and changes in soil moisture levels. The results of analyses here have shown that while the spatial correlation parameters between plots are consistent across harvests at some trials, they vary considerably for others.

The importance of accounting for the different forms of spatial variation can be clearly seen in the improvement in AIC values between the RCB model and spatial model at most harvests. Accounting for the spatial variation will result in more accurate variety predictions. The difference between these predictions under the two models can be seen in Figure 3.6 where the BLUPs from the spatial model and RCB model for the yield at harvest 5 at Terry Hie Hie are plotted against one another. In the top 20% of varieties under the two methods there are 9 out of 12 (75%) of varieties in common between the two methods. Hence 3 of the varieties selected in the top 20% under the improved spatial method would not have been selected if the RCB analysis had been used.

3.5 Conclusions

Spatial variation is clearly evident in variety selection trials in perennial pasture crops. Variety selections will be affected if this spatial variation is not accounted for. Hence approaches for analysing variety selection data from perennial crops need to incorporate spatial analysis methods.

While in this chapter each harvest time has been analysed separately to better understand the spatial variation at each harvest, in order to obtain variety predictions for selection of new varieties, it would be more informative for the breeder to analyse the data across harvests and sites. Methods for analysing multi-site variety selection trials in annual crops are well developed (see Smith et al., 2005 for a review) and the methods may be extended to handle perennial crops. There are two extra issues that need consideration in such an analysis; one being how best to model the residual correlation structure within a trial with multiple harvests and secondly how best to model the genetic effects over time. The residual correlation structure will need to take into account both spatial variation between plots and temporal correlation between harvests. As has been shown in this chapter the spatial correlation is likely to vary between harvests and hence the residual correlation model will ideally provide for spatio-temporal interaction. This study has provided new insight into the spatial and temporal issues impacting multi-harvest data in perennial crops. In subsequent chapters methods addressing these issues will be

presented.

Chapter 4

Simulation study

4.1 Introduction

In Chapter 3, spatial analysis methods based on the approach of Gilmour et al. (1997) and Stefanova et al. (2009), were presented and analyses performed on data from the individual harvests of the lucerne and chicory trials. It was shown that the spatial models provided a better fit than the RCB model for most harvests (based on REMLRT and AIC values, see Tables (3.2, 3.4, 3.5). It was also shown (in the detailed analysis of harvest 5 at Terry Hie Hie), that the variety BLUPs obtained from the two methods differed (Figure 3.6).

While it is clear that the spatial analysis methods provide models that are a better fit to the data than the RCB model, it is important to assess the impact these models have on estimating the genetic effects and how the estimates obtained from the spatial analysis model and RCB model compare to the true genetic effects. This may be investigated using simulation methods, where data may be generated with known genetic effects.

A further consideration is the issue of fitting a measurement error component, as advocated by Stefanova et al. (2009). In many cases the fitting of measurement error in models to analyse data from field trials has been problematic and has not been routinely implemented. For these reasons the fitting of measurement error was omitted in the analyses presented in Chapter 3. It is important to assess the impact of fitting measurement error, or omitting it, on the estimation of genetic effects. This may also be investigated using simulation methods.

In this chapter, simulation studies based on the lucerne yield data, are conducted to investigate the above issues. To date there have been no known simulation studies published looking into these issues. Firstly, a simulation study is presented to compare the variety predictions obtained under the spatial model to those obtained from the RCB model (with regards to how they compare to the true genetic effects). As discussed in the following section, this simulation study is performed under the caveat that the issue of model selection (for the spatial model) has been disregarded, due to practicalities of implementation, and the assumed underlying spatial trend takes the form of an ar1(Col).ar1(Row) process. The results therefore are restricted but provide an insight into the comparison of spatial modelling versus fitting the RCB model. A second simulation study is presented to show

the impact of fitting measurement error in the spatial model, on the estimation of genetic effects. The second simulation study is unaffected by the model selection caveat involved in the first study but still assumes an underlying ar1(Col).ar1(Row) spatial process.

The underlying hypotheses being investigated are that, under a range of spatial correlation and genetic variance (and where relevant, measurement error) levels,

- Variety predictions from the spatial model will be significantly closer to the true genetic effects than those from the RCB model
- Fitting a spatial model without including a measurement error term will not result in any significant loss of accuracy in variety predictions than fitting a spatial model with a measurement error term

4.2 Methodology

4.2.1 Models

RCB versus spatial

A restricted simulation study was conducted to examine the importance of spatial modelling on variety predictions and response to selection. The simulations were based on the lucerne yield data from Terry Hie Hie harvest 5. This trial involved 60 genotypes grown in a RCB design with 3 reps, laid out in 6 columns by 30 rows.

A number of different sets of simulations were performed based on different combinations of estimates of the variance components and fixed effects in the linear mixed model (2.2.1). The first set of simulated data (set 0 in Table 4.3) was based on the parameters obtained from fitting Model 4 (in Table 3.1) to the lucerne yield data from Terry Hie Hie harvest 5. The variance parameters are given in Table 4.1. The fixed effects in the model involved a harvest mean and a linear Row term with parameter estimates denoted by $\boldsymbol{\tau_0}$. Based on the parameters in Table 4.1, the genetic variance parameter was taken to be

Table 4.1: Parameter values used for the first set of simulations for RCB vs spatial models

	Component
Variety	0.021
Rep	0.046
Column	0.129
Residual variance (σ^2)	0.097
Column correlation (ϕ_c)	0.003
Row correlation (ϕ_r)	0.255

0.216 times the residual error variance (0.097), the row spatial correlation parameter (ϕ_r) was 0.255, and the spatial correlation parameter for the columns (ϕ_c) was 0.003. Given the values for these parameters, values were simulated for the random effects. For example, the genetic effects were obtained by sampling from the distribution $N(\mathbf{0}, \sigma^2 \boldsymbol{G}_g(\gamma_g))$

where σ^2 is the value of the residual error variance component (0.097), γ_g is the variance component ratio for the genetic effects (0.216), and $\boldsymbol{G}_g = \gamma_g \boldsymbol{I}$. Hence a set of known (true) genetic effects were obtained ($\boldsymbol{g}$). Similarly, values for the other non genetic random effects $\boldsymbol{u}_o$ were generated by sampling from $N(\boldsymbol{0}, \sigma^2 \boldsymbol{G}_o(\boldsymbol{\gamma}_o))$ and $\boldsymbol{e}$ was generated by sampling from $N(\boldsymbol{0}, \sigma^2 \boldsymbol{R}(\boldsymbol{\phi}))$, where $\boldsymbol{R} = \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$ where $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\Sigma}_r$ are ar1 correlation matrices corresponding to the column and row dimensions respectively, with parameters ϕ_c and ϕ_r. Therefore a set of simulated data $\boldsymbol{y}^{(s)}$, was generated as

$$\boldsymbol{y}^{(s)} = \boldsymbol{X}\boldsymbol{\tau_0} + \boldsymbol{Z}_g \boldsymbol{u}_g^{(s)} + \boldsymbol{Z}_o \boldsymbol{u}_o^{(s)} + \boldsymbol{e}^{(s)} \tag{4.2.1}$$

1000 such simulations were performed and hence 1000 data sets generated ($\boldsymbol{y}^{(s)}, s = 1, \ldots, 1000$). The RCB model and the spatial model were fitted to each of these simulated data sets. For each simulation, variety BLUPs obtained from the two methods were compared to the true genetic effects. The same spatial model was fitted to each of the 1000 data sets.

To investigate the more general situation, a further 12 sets of 1000 simulations ($1-12$ in Table 4.3) were performed for a range of combinations of parameters. These sets of simulations covered a range of genetic variances ($0.1, 0.5, 1, 2$ times the residual variance) and a range of spatial correlation scenarios (low ($\phi_r = 0.4$, $\phi_c = 0.2$), medium ($\phi_r = 0.6$, $\phi_c = 0.4$), high ($\phi_r = 0.8$, $\phi_c = 0.6$)). These simulation sets did not include the linear Row or random Column terms which were specific for the data at Terry Hie Hie harvest 5 (set 0). The spatial model fitted to each of these simulation sets included a random Rep effect and an ar1(Col).ar1(Row) residual correlation structure.

The restriction in this study involves the spatial models fitted. In practice model selection is an integral part of the spatial modelling process. That is, the best spatial model is determined for each data set. In this simulation study the same spatial model was fitted to all data sets within a simulation set. That is, for simulation sets $1-12$ the only spatial model fitted was the separable ar1(Col).ar1(Row) residual model (plus Rep effect); this is the model that generated the data. Thus the comparison between spatial models and the RCB model ignores the additional source of variability that model selection brings and assumes an underlying spatial trend in the form of an ar1(Col).ar1(Row) process. Therefore whilst this study is not all encompassing, it does provide restricted information on how the RCB and spatial models compare under varying levels of genetic variance and also different levels of spatial correlation.

Measurement error

The simulation study to assess the effect of including measurement error in the spatial model was also initially based on the lucerne yield data from Terry Hie Hie harvest 5. Subsequent variations of the variance parameter estimates were examined, to cover a wide range of possible scenarios.

The error variance was partitioned into a spatially dependent random vector $\boldsymbol{\xi}$, representing the smooth local trend component and a measurement error component $\boldsymbol{\eta}$,

giving $\boldsymbol{e} = \boldsymbol{\xi} + \boldsymbol{\eta}$, (in the linear mixed model (2.2.1)), where var $(\boldsymbol{\xi}) = \sigma^2 \boldsymbol{\Sigma_c} \otimes \boldsymbol{\Sigma_r}$ and var $(\boldsymbol{\eta}) = \sigma_\eta^2 \boldsymbol{I}_c \otimes \boldsymbol{I}_r$. Hence the total error variance is given by $\boldsymbol{R} = \sigma^2 \boldsymbol{\Sigma_c} \otimes \boldsymbol{\Sigma_r} + \sigma_\eta^2 \boldsymbol{I}_c \otimes \boldsymbol{I}_r$. Simulated data sets were constructed based on (4.2.1) with $\boldsymbol{e}^*$ generated as the sum of $\boldsymbol{\xi}^*$ and $\boldsymbol{\eta}^*$ where $\boldsymbol{\xi}^*$ was generated by sampling from $N(\boldsymbol{0}, \sigma^2 \boldsymbol{R}(\boldsymbol{\phi}))$, where $\boldsymbol{R} = \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$ with $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\Sigma}_r$ assumed to be ar1 correlation matrices corresponding to the column and row dimensions respectively, and $\boldsymbol{\eta}^*$ was generated by sampling from $N(\boldsymbol{0}, \sigma_\eta^2 \boldsymbol{I})$.

The parameter values used in generating the data for the first set of simulations (set 0 in Table 4.5) were based on fitting Model 4 (in Table 3.1) plus a measurement error term, to the yield data from harvest 5 at Terry Hie Hie. The parameter values for the random effects are given in Table 4.2. The fixed effects in the model included a harvest mean and linear Row term. 1000 simulations were based on these parameter values.

Table 4.2: Parameter values used for the first set of simulations for investigating measurement error

	Component
Variety	0.019
Rep	0.040
Column	0.121
units(measurement error) (σ_η^2)	0.064
Residual variance (σ^2)	0.039
Column correlation (ϕ_c)	0.193
Row correlation (ϕ_r)	0.738

To investigate the results for the more general case, a further 18 sets of 1000 simulations were conducted based on a range of parameter values for genetic variance, spatial correlation and measurement error. The parameter value combinations for the simulation sets are detailed in Table 4.5. These simulation sets covered a range of genetic variances ($0.5, 1, 2$ times the total error variance (i.e the sum of the measurement error variance component and the local spatial variation component)), a range of spatial correlation levels, medium ($\phi_r = 0.6$, $\phi_c = 0.4$), high ($\phi_r = 0.9$, $\phi_c = 0.7$), and a range of levels of measurement error ($\frac{1}{3}$, $\frac{1}{2}$ and $\frac{2}{3}$ times the total error variance (i.e. $0.5, 1, 2$ times the local spatial trend component). Note that the total error variance in each of these simulation sets has been arbitrarily set to 1 (i.e. $\sigma_\eta^2 + \sigma^2 = 1$) without loss of generality. These simulation sets did not include the linear Row and random Column terms specific to set 0. The spatial model fitted to each of these simulation sets included a random Rep effect and an ar1(Col).ar1(Row) residual correlation structure plus a measurement error term.

The number of cases where the spatial plus measurement error model could not be fitted (model did not converge within 100 iterations or parameter estimates went to the boundary) was recorded (labelled dnc, for did not converge, in Table 4.5). The spatial models without measurement error converged within 50 iterations for all simulations.

4.2.2 Measures for comparison

To compare the different models for their effectiveness in predicting variety effects, measures were calculated based on the response to selection. The response to selection is a measure used by breeders to quantify the change produced by variety selection.

A key measure for comparing the models is the correlation between the true variety effects and the predicted BLUPs for each variety under each model. This correlation is referred to as accuracy. The stronger the correlation (accuracy), the greater is the response to selection.

In the situation where a breeder selects the top $p\%$ of varieties to be retained for further testing, the response to selection is calculated as the product of the narrow sense heritability and the selection differential (the deviation of the mean of the selected individuals from the population mean) (Falconer & Mackay, 1996).

The realised response to selection for a given trait can be calculated as the mean of the BLUPs of the top $p\%$ of ranked varieties (Cullis et al., 2006). The relative bias and relative mean squared error of the realised response to selection are two measures that were used in the simulation study to compare the spatial and RCB models.

The measures were calculated as

- Mean accuracy

$$MAC = \frac{1}{s}\sum_{i=1}^{s} cor(\boldsymbol{g}_i, \tilde{\boldsymbol{g}}_i)$$

  where $\boldsymbol{g}_i$ is the vector of true genetic effects and $\tilde{\boldsymbol{g}}_i$ is the vector of variety BLUPs for the i^{th} simulation, and s is the number of simulations.

- The relative bias, and relative mean square error of the realised response to selection, where the realised response to selection (RTS) is calculated as the mean of the BLUPs of the top 10% of ranked varieties, and the relative bias (RBRTS) and relative mean square error (RMSERTS) of the RTS are calculated as

$$RBRTS = \frac{1}{s}\sum_{i=1}^{s} \frac{\tilde{RTS}_i - RTS_i}{RTS_i}$$

  and

$$RMSERTS = \frac{1}{s}\sum_{i=1}^{s} \left(\frac{\tilde{RTS}_i - RTS_i}{RTS_i}\right)^2$$

  where $\tilde{RTS}_i$ is the estimated realised response to selection and RTS_i is the true response to selection, for the i^{th} simulation

4.3 Results

4.3.1 RCB versus spatial

The average values of the accuracy of the variety BLUPs from the RCB and spatial models were compared under the different spatial correlation and genetic variance combinations. The relative bias and relative mean squared error of the realised response to selection, based on the RCB model (and similarly the spatial model) were also calculated for each of the different simulation sets. The results are given in Table 4.3.

In all simulation sets the spatial model provided higher mean accuracy, lower RMSE and relative bias values closer to zero, for response to selection, than the RCB model. As expected, this improvement was greatest in the simulations involving the higher spatial correlation parameters.

Figure 4.1 shows the accuracy obtained under the spatial and RCB models under the low, medium and high spatial correlation scenarios, plotted against the genetic variance. This plot shows the increase in accuracy as the genetic variance increases for both the RCB and spatial model and the clear improvement in accuracy of the spatial model over the RCB model as the spatial correlation increases. The improvement is considerable when the spatial correlation is high and the genetic variance is low (accuracy of 0.55 under RCB versus 0.82 under spatial model, when the genetic variance is 0.1 times the residual variance, i.e. heritability = 0.09).

Figure 4.2 presents the realised RTS versus the true RTS for each of the 1000 simulations under the RCB model and the spatial model from simulation set 0, based on the original data from Terry Hie Hie harvest 5. The RCB model performs poorly with many RTS values predicted at zero and generally a low correlation with the true RTS values. While the predicted RTS values for the spatial model are much closer to the true values than the RCB model, the relatively low genetic variance and low spatial correlation in this simulation set results in only moderate correlation with the true values even for the spatial model. Figure 4.3 presents the RTS results of simulation set 10 (in Table 4.3) which has higher spatial correlation and higher genetic variance. This plot shows that there was greater agreement between the predicted RTS under the spatial model and the true RTS and once again the improvement in the predictions from the spatial model as compared to the RCB model.

4.3.2 Measurement error

Before conducting the simulations based on the inclusion of measurement error, the spatial models fitted to the lucerne yield data from the individual harvests in Chapter 3 were re-run with the inclusion of a measurement error term in each analysis. This was done to see how the spatial model and spatial model with measurement error compared in terms of goodness of fit (as compared using non standard REMLRT (Stram & Lee, 1994)). These results are presented in Table 4.4. In 18 of the 28 harvests the addition of measurement error improved the log-likelihood significantly ($P < 0.05$) (or close to significant with 3

Table 4.3: Mean accuracy for BLUPs obtained from RCB and spatial models and correlations between true response to selection and the response to selection obtained from fitting RCB and spatial models based on 1000 simulations. Different sets of simulations were performed based on a range of genetic variances and spatial correlation parameters.

Set	Genetic Var	ϕ_r	ϕ_c	Mean accuracy		RTS			
						Relative bias		RMSE	
				RCB	Spatial	RCB	Spatial	RCB	Spatial
0*	0.02	0.26	0.003	0.53	0.63	-0.64	-0.37	0.71	0.45
1	0.1	0.4	0.2	0.49	0.54	-0.53	-0.48	0.63	0.57
2	0.5	0.4	0.2	0.77	0.81	-0.23	-0.18	0.30	0.25
3	1	0.4	0.2	0.87	0.89	-0.13	-0.10	0.19	0.16
4	2	0.4	0.2	0.93	0.94	-0.07	-0.05	0.13	0.12
5	0.1	0.6	0.4	0.50	0.65	-0.55	-0.36	0.65	0.44
6	0.5	0.6	0.4	0.79	0.88	-0.24	-0.11	0.31	0.17
7	1	0.6	0.4	0.87	0.93	-0.14	-0.06	0.20	0.13
8	2	0.6	0.4	0.93	0.96	-0.08	-0.03	0.13	0.10
9	0.1	0.8	0.6	0.55	0.82	-0.61	-0.18	0.70	0.24
10	0.5	0.8	0.6	0.82	0.95	-0.24	-0.04	0.31	0.11
11	1	0.8	0.6	0.90	0.98	-0.14	-0.02	0.21	0.10
12	2	0.8	0.6	0.94	0.99	-0.07	-0.004	0.13	0.08

* indicates initial set of simulations based on data from harvest 5 at Terry Hie Hie. A linear Row term and random Column term were included in the spatial analysis model for this set of simulations.

harvests having $P = 0.06$). A measurement error term was unable to be fitted in 3 of the 28 analyses.

The results from the simulations for the models with and without measurement error are given in Table 4.5. The mean accuracy and relative bias and relative mean square error for response to selection were calculated for each of the simulation sets, covering a range of genetic variation, spatial correlation and measurement error values. In all cases the mean accuracy was almost exactly the same between the two models (with and without measurement error). The relative bias and relative mean square error for response to selection were also very similar for the two methods. The number of simulations out of 1000 in which the spatial plus measurement error term model was unable to be fitted or did not converge within a reasonable number of iterations was large (ranged from 103 to 566).

It can be seen (Table 4.5) that the number of simulations that did not converge (dnc) increased as the measurement error component increased. It can also be seen that the number that did not converge was higher in the simulations with high spatial correlation, than the simulations with moderate spatial correlation, and also generally increased as the genetic variance increased.

Table 4.4: Results from spatial models fitted with and without measurement error (M2 and M1 respectively) to the 28 lucerne yield harvests. REML log-likelihoods for each model and P values based on adjusted REML likelihood ratio tests comparing the two models, are given, as well as parameter estimates for the spatial correlation parameters, measurement error and residual error variance.

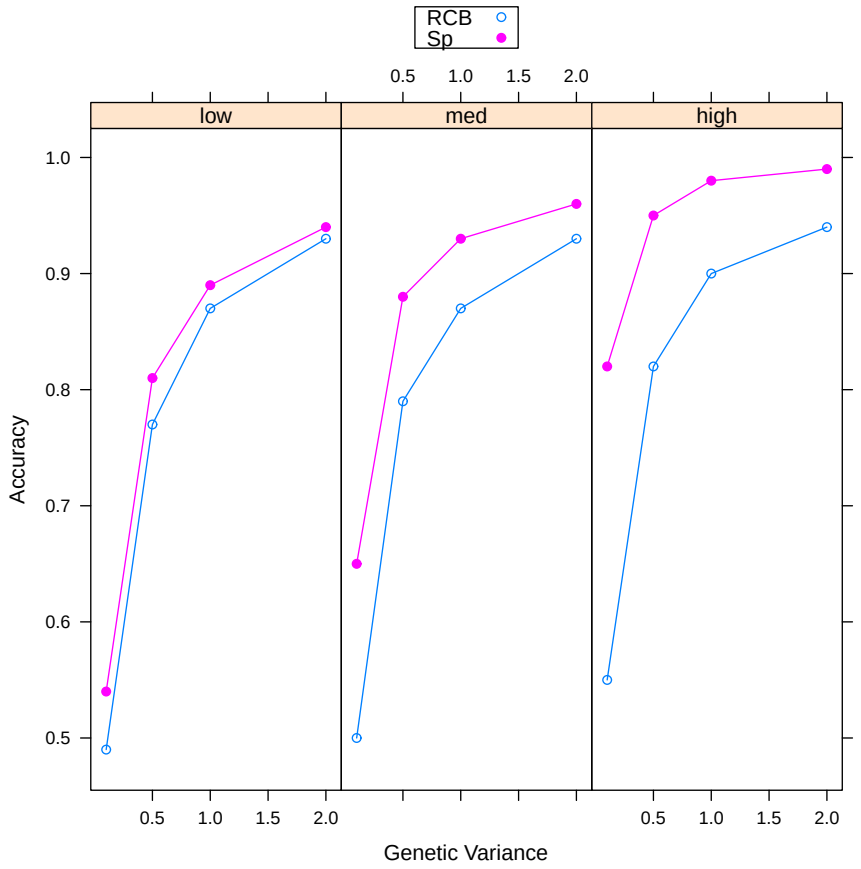
Trial / Harvest	M1 loglik $\sigma^2_\eta = 0$	M2 loglik $\sigma^2_\eta \neq 0$	P	$\hat{\phi}_c$	$\hat{\phi}_r$	$\hat{\sigma}^2_\eta$	$\hat{\sigma}^2$
Euloma 1	-32.414	-32.414	0.500	0.211	0.628	0.000	0.807
Euloma 2	5.568	5.568	0.500	0.300	0.626	0.000	0.589
Euloma 3	-34.086	-33.557	0.152	0.311	0.679	0.062	0.821
Euloma 4	-23.136	-23.136	0.500	0.000	0.587	0.000	0.703
Euloma 5	-12.590	-12.590	0.500	0.259	0.610	0.000	0.660
Euloma 6	-37.133	-37.133	0.498	0.000	0.615	0.000	0.865
Leadville 1	34.703	35.953	0.057	0.000	0.661	0.086	0.184
Leadville 2	-3.845	-0.843	0.007	0.000	0.820	0.106	0.459
Leadville 3	1.578	6.088	0.001	0.625	0.870	0.141	0.456
Leadville 4	12.693	15.917	0.006	0.000	0.770	0.091	0.308
Leadville 5	30.026	34.981	0.001	0.000	0.842	0.154	0.137
Sandigo 1	4.357	6.654	0.016	0.447	0.749	0.081	0.396
Sandigo 2	-109.659	-98.973	0.000	0.857	0.946	0.549	3.958
Tamworth 1	-31.022	-22.345	0.000	0.837	0.723	0.166	1.071
Tamworth 2	-3.342	-0.656	0.010	0.612	0.718	0.156	0.422
Tamworth 3	-3.032	-3.032	0.500	0.181	0.167	0.000	0.384
Tamworth 4	11.178	12.339	0.064	0.652	0.000	0.183	0.106
Tamworth 5	-21.609	-16.661	0.001	0.775	0.485	0.180	0.364
Terry Hie Hie 1	41.560	42.720	0.064	0.000	0.605	0.134	0.108
Terry Hie Hie 2	30.446	33.183	0.010	0.000	0.730	0.128	0.136
Terry Hie Hie 3	80.484	81.753	0.056	0.000	0.679	0.093	0.060
Terry Hie Hie 4	71.044	NA					
Terry Hie Hie 5	93.496	95.082	0.037	0.000	0.706	0.060	0.042
Terry Hie Hie 6	12.010	13.491	0.043	0.000	0.806	0.089	0.089
Terry Hie Hie 7	84.922	NA					
Terry Hie Hie 8	66.775	67.540	0.108	0.000	0.700	0.114	0.056
Terry Hie Hie 9	93.286	NA					
Terry Hie Hie 10	31.376	34.419	0.007	0.000	0.960	0.175	0.076

Table 4.5: Mean accuracy for BLUPs obtained from spatial models with and without measurement error ($\sigma^2_\eta \neq 0$ and $\sigma^2_\eta = 0$ respectively), and relative bias and relative mean square error of realised response to selection obtained from 1000 simulations. The different sets of simulations were based on a range of genetic variances and different spatial correlation and measurement error parameters. The number of simulations that did not converge (dnc) for the spatial plus measurement error model are given in the last column. Note, the results for the spatial plus measurement error model are based on the simulations that converged (i.e. 1000-dnc simulations).

Set	Gen Var	ϕ_r	ϕ_c	σ^2_η	σ^2	Mean accuracy		RTS				dnc
								Relative bias		RMSE		
						$\sigma^2_\eta = 0$	$\sigma^2_\eta \neq 0$	$\sigma^2_\eta = 0$	$\sigma^2_\eta \neq 0$	$\sigma^2_\eta = 0$	$\sigma^2_\eta \neq 0$	
0*	0.02	0.74	0.19	0.06	0.04	0.61	0.61	-0.37	-0.40	0.47	0.52	416
1	0.5	0.6	0.4	0.33	0.67	0.82	0.82	0.23	0.24	-0.16	-0.17	107
2	1.0	0.6	0.4	0.33	0.67	0.90	0.90	0.16	0.16	-0.09	-0.10	103
3	2.0	0.6	0.4	0.33	0.67	0.94	0.94	0.12	0.12	-0.05	-0.05	124
4	0.5	0.9	0.7	0.33	0.67	0.86	0.87	0.18	0.19	-0.10	-0.12	143
5	1.0	0.9	0.7	0.33	0.67	0.92	0.93	0.13	0.13	-0.05	-0.06	168
6	2.0	0.9	0.7	0.33	0.67	0.96	0.96	0.10	0.10	-0.02	-0.03	195
7	0.5	0.6	0.4	0.50	0.50	0.79	0.80	0.25	0.26	-0.19	-0.19	194
8	1.0	0.6	0.4	0.50	0.50	0.88	0.88	0.17	0.16	-0.10	-0.11	230
9	2.0	0.6	0.4	0.50	0.50	0.94	0.94	0.12	0.13	-0.06	-0.06	269
10	0.5	0.9	0.7	0.50	0.50	0.82	0.83	0.21	0.21	-0.14	-0.16	335
11	1.0	0.9	0.7	0.50	0.50	0.90	0.91	0.15	0.14	-0.08	-0.08	352
12	2.0	0.9	0.7	0.50	0.50	0.95	0.95	0.11	0.11	-0.04	-0.04	420
13	0.5	0.6	0.4	0.67	0.33	0.78	0.78	0.28	0.27	-0.22	-0.21	349
14	1.0	0.6	0.4	0.67	0.33	0.87	0.87	0.18	0.18	-0.12	-0.12	385
15	2.0	0.6	0.4	0.67	0.33	0.93	0.93	0.12	0.12	-0.06	-0.06	384
16	0.5	0.9	0.7	0.67	0.33	0.80	0.80	0.26	0.26	-0.19	-0.20	462
17	1.0	0.9	0.7	0.67	0.33	0.88	0.89	0.17	0.17	-0.10	-0.10	515
18	2.0	0.9	0.7	0.67	0.33	0.94	0.94	0.12	0.13	-0.06	-0.06	566

* indicates initial set of simulations based on data from harvest 5 at Terry Hie Hie. A linear Row term and random Column term were included in the spatial analysis model for this set of simulations.

Figure 4.1: Plot of accuracy values for RCB and spatial model for varying levels of spatial correlation (low, med, high) and genetic variance, based on simulation results for sets $1-12$ in Table 4.3

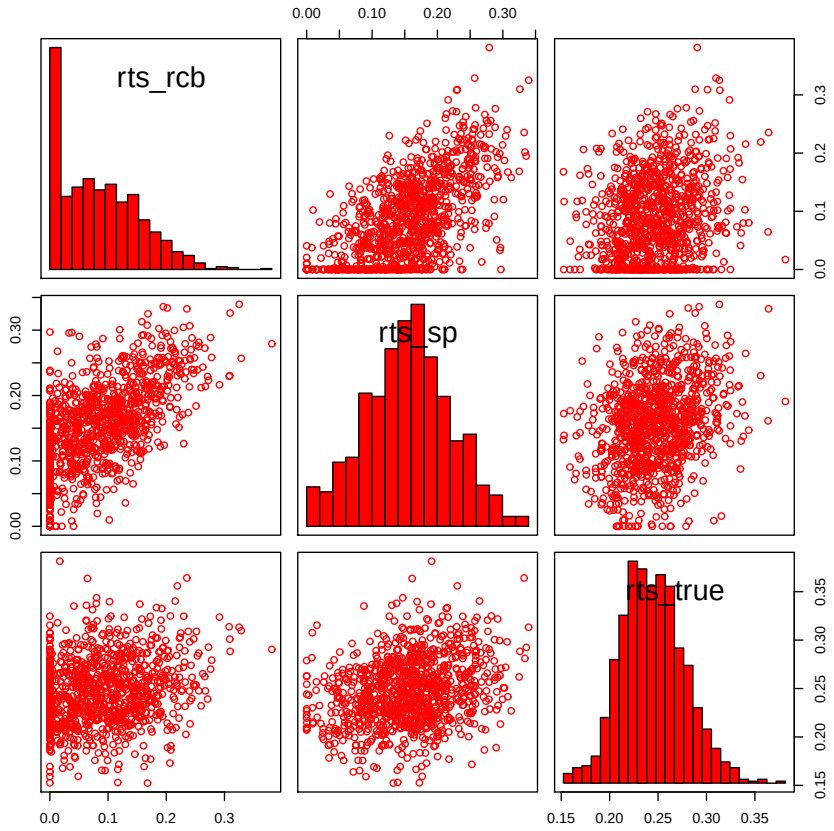


4.4 Summary

The simulation studies in this chapter provide new insight into the value of spatial analysis in perennial crops that is not known to be documented elsewhere in the literature. The simulation results have clearly shown that the spatial model results in better predictions of genetic effects (closer to the true genetic effects) than the RCB model, across a wide range of different genetic variance and spatial correlation levels. As would be expected, this improvement is greater when the spatial correlation is higher. Therefore the spatial analysis methods applied in Chapter 3, not only result in models that provide a better fit to the data but also result in more accurate estimates of genetic effects, than the RCB model, thereby supporting the original hypothesis. It is therefore important to implement these spatial analysis models when analysing perennial crop variety selection data, to obtain accurate variety selections. These spatial analysis methods will be used in subsequent chapters when data is analysed across harvests.

The simulation results for the spatial models with or without fitting a measurement error term, showed variety predictions were very similar under both models and while the addition of a measurement error component to the spatial model resulted in a better model fit in a number of cases, this did not result in an improvement in variety predic-

Figure 4.2: Plots of response to selection for RCB (rts_rcb) and spatial model (rts_sp) vs true RTS (rts_true) for each simulated data set in set 0 (in Table 4.3), based on original data from Terry Hie Hie harvest 5



tions. The original hypothesis was therefore accepted. The number of times where a measurement error term was unable to be fitted was large (ranging from $103 - 566$, out of 1000 simulations). Due to the frequent difficulty in fitting a measurement error term and the fact that the variety predictions were virtually the same under both approaches, it would appear that when analysing perennial crop selection data at individual harvest times the simplest approach is to omit the measurement error term. There are no known published simulation studies investigating the merit of fitting a measurement error term in the spatial analysis models fitted in this chapter.

While a measurement error term may be difficult to fit and not impact on variety predictions at the spatial level, it may be important to account for measurement error in the spatio-temporal sense when data is analysed across harvests. This will be investigated in subsequent chapters.

Figure 4.3: Plots of response to selection for RCB (rts_rcb) and spatial model (rts_sp) vs true RTS (rts_true) for each simulated data set in set 10 (in Table 4.3)

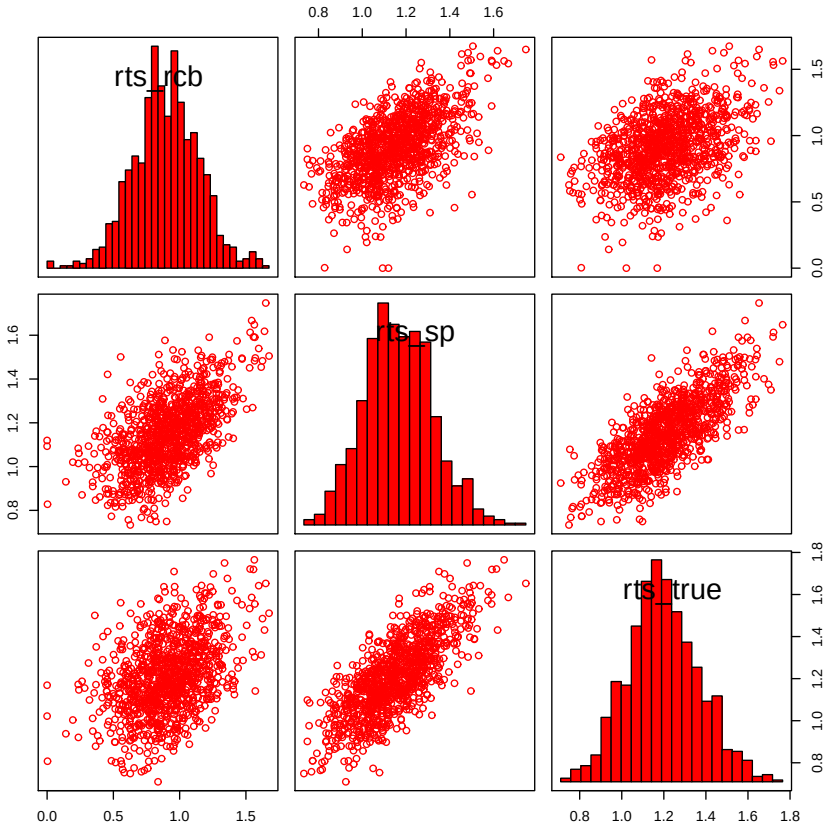
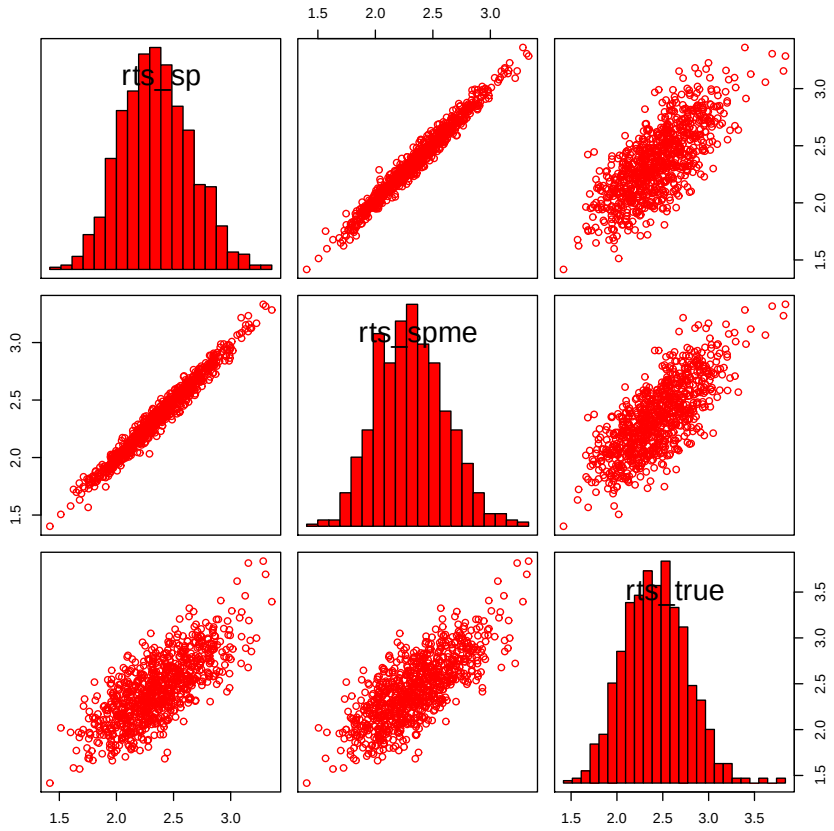


Figure 4.4: Plots of response to selection for spatial models with (rts_spme) and without (rts_sp) measurement error and true RTS (rts_true) for each simulated data set (in which the spatial plus measurement error model converged) in set 6 in Table 4.5



Chapter 5

Analysis of multi-harvest data using separable variance models: Theory

The need for spatial analysis in perennial crops was established in chapters 3 and 4, where the methods of Gilmour et al. (1997) were applied to the data from individual harvests from the motivating data sets. Rather than concentrating on individual harvests in isolation, variety selection in perennial crops is usually based on data from multiple harvests over time, and analysis methods need to incorporate the data across harvests. Such methods need to account for both the spatial correlation and temporal correlation between measurements.

Smith et al. (2007) present a method for analysing data from multiple harvests in perennial crops in the case of a short sequence of balanced repeated measurements in sugarcane variety trials. Smith et al. (2007) use so-called separable residual covariance structures. In this chapter the approach of Smith et al. (2007) is reviewed and then new extensions are presented that may prove suitable for data from longer sequences of multiple measurements, as is more typically the case in perennial crop trials. Firstly, a linear mixed model is introduced for analysis of multi-harvest data, on which the method of Smith et al. (2007) and the new extensions, are built.

5.1 Mixed model for single site multi-harvest data

Consider a perennial crop variety selection trial consisting of n plots in a rectangular array of c columns by r rows ($n = cr$), in which m genotypes are grown and multiple harvests are made. Let h denote the number of harvests (or assessment dates) for the trial and let $\boldsymbol{y}$ be the $hn \times 1$ vector of data observations across all the harvests, ordered as rows within columns within harvests.

A linear mixed model for the data, $\boldsymbol{y}$ may be written, based on (2.2.1) as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{Z}_g\boldsymbol{u}_g + \boldsymbol{Z}_o\boldsymbol{u}_o + \boldsymbol{e} \tag{5.1.1}$$

where $\boldsymbol{\tau}$ is a $(p \times 1)$ vector of fixed effects with design matrix $\boldsymbol{X}^{(hn \times p)}$, $\boldsymbol{u}_g$ is the $hm \times 1$ vector of variety (or genetic) effects for individual harvests, ordered as varieties within harvests, with associated design matrix $\boldsymbol{Z}_g^{(hn \times hm)}$, $\boldsymbol{u}_o$ is a vector of other random effects with associated design matrix $\boldsymbol{Z}_o$ and $\boldsymbol{e}$ is the $hn \times 1$ vector of residuals.

The random effects from the linear mixed model (5.1.1) are assumed to follow a Normal distribution with zero mean vector and variance-covariance matrix

$$\text{var}\left(\begin{bmatrix} \boldsymbol{u}_g \\ \boldsymbol{u}_o \\ \boldsymbol{e} \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{G_g} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G_o} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R} \end{bmatrix} \tag{5.1.2}$$

Therefore the distribution of the data $\boldsymbol{y}$ is Normal with mean $\boldsymbol{X\tau}$ and variance matrix

$$\boldsymbol{H} = \text{var}\,(\boldsymbol{y}) = \boldsymbol{Z_g G_g Z_g}^T + \boldsymbol{Z_o G_o Z_o}^T + \boldsymbol{R}$$

5.2 Multi-harvest analysis approach of Smith et al 2007

Smith et al. (2007) present an approach to analysing multi-harvest variety selection data in perennial crops. Their approach actually models variety selection data from multi-environment trials involving a small number of multiple harvests (balanced across the sites). In this section their approach is summarized for the simplified case of an analysis of multi-harvest data from a single site.

Smith et al. (2007) base their analysis on the linear mixed model (5.1.1). They use a sequential modelling process. The steps involve allowing for non-genetic variation, through design, management and other sources of variation such as spatial trends in the field, accounting for temporal variation and correlation that is inherent in the multiple harvests, and importantly from the breeding point of view, modelling the genetic variation. Their approach involves assuming a simple initial model for $\boldsymbol{G_g}$ in order to determine a suitable residual (non-genetic) model ($\boldsymbol{R}$) and then using this residual model to investigate more complex genetic models.

5.2.1 Modelling non-genetic effects

Smith et al. (2007) recognised that in order to obtain efficient predictions of genetic effects it is essential to suitably model the non-genetic effects such as spatial variation in the field and temporal correlation between repeated measurements. Smith et al. (2007) followed the spatial modelling method of Gilmour et al. (1997), identifying spatial global trend and extraneous variation terms for each harvest and including these terms in $\boldsymbol{\tau}$ or $\boldsymbol{u_o}$. They extended the spatial analysis approach to the spatio-temporal situation by modelling the local spatial correlation and the temporal correlation between the multiple measurements by assuming a three-way separable spatio-temporal process for the residual

variance structure for a trial. Thus in (5.1.2), the structure for $\boldsymbol{R}$ is assumed to be,

$$\boldsymbol{R} = \boldsymbol{R}_h \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r \tag{5.2.3}$$

where $\boldsymbol{R}_h$ is a $h \times h$ covariance matrix that incorporates temporal correlation (between harvests) and possibly heterogeneous variance across harvests, and $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\Sigma}_r$ are the $c \times c$ and $r \times r$ column and row spatial correlation matrices; in this formulation these latter structures are common to all harvests within the trial.

Smith et al. (2007) model $\boldsymbol{R_h}$ using an unstructured matrix, the most general form of covariance matrix, requiring the estimation of $h(h+1)/2$ parameters. For a small number of harvests (as in Smith et al., 2007) this may be suitable but for more extensive sequences of repeated harvests the number of parameters to be estimated may become prohibitive. More parsimonious covariance structures such as the uniform structure (with equal variances and equal correlation between harvests) or heterogeneous covariance model (with differing variances and equal correlation) could be considered but are unlikely to be suitable in practice as correlations between the different harvest times are unlikely to be equal.

In subsequent sections more parsimonious covariance structures for $\boldsymbol{R}$ will be considered.

5.2.2 Modelling genetic effects

Smith et al. (2007) present various models for the genetic effects in the multi-harvest situation where only a small number of harvests (in their case a maximum of three) are involved. In particular they represent the variance matrix $\boldsymbol{G}_g$ in (5.1.2), for such effects by

$$\boldsymbol{G}_g = \boldsymbol{G}_h \otimes \boldsymbol{I}_m$$

where $\boldsymbol{G}_h$ is a $h \times h$ matrix (referred to as the genetic variance matrix) with diagonal elements representing the genetic variance for each harvest and the off diagonal elements representing genetic covariances between the harvests, and $\boldsymbol{I}_m$ (the $m \times m$ identity matrix) is the assumed structure for the varieties. This variety structure ($\boldsymbol{I}_m$) assumes the varieties are unrelated. In Smith et al. (2007) $\boldsymbol{G}_h$ is modelled using an unstructured (us) matrix but they note that factor analytic (fa) models may also be suitable.

In many cases of perennial pasture variety selection trials the number of harvests h is likely to be greater than that in Smith et al. (2007). This makes the use of the unstructured matrix problematic and more parsimonious models need to be used. It is also desirable that the models used allow for investigation into variety by harvest interaction. In the following sections suitable methods for modelling the genetic effects over multiple harvests will be presented.

5.3 Extensions

5.3.1 Contrasts/ deviations or modelling response over time

In the two analyses presented in Smith et al. (2007) of data sets consisting of 2 and 3 harvests the approach was to fit a fixed main effect for harvests in the linear mixed model and hence the analyses were based on the deviations from the harvest means. When there are more harvests involved, the issue arises of whether to model the deviations (or contrasts) from the harvest means, or to model the overall response profile over time. This choice will depend on the aim of the experiment and the trait involved. If the level of the trait is of interest then the ideal approach would be to model the response over time. If interest lies in the differences between varieties more than the actual level of a trait then it may be best to base the analysis on the deviations from the harvest means. Evans & Roberts (1979) show that while the absolute response profile of repeated measurement data on perennial crops may be complex in nature, the sequence of deviations or contrasts from the harvest means may be simpler to model.

For example, in the case of the lucerne persistence data at Terry Hie Hie, the actual level of variety response is of interest as predictions of time to a certain level of persistence are required. The trait response follows a systematic trend over time (Figure 1.4). The ideal approach in this situation is to model the underlying overall trend over time using a smooth curve (e.g. polynomial or cubic smoothing spline) and then investigate the departures from this underlying trend for each variety. These variety departures may be modelled using linear functions or may require more complex models including splines. As discussed in Chapter 1 the cubic smoothing spline may be written as a mixed model and hence it can easily be incorporated into the linear mixed model (5.1.1), and so may be useful in this situation.

Alternatively, in the case of the lucerne yield data at Terry Hie Hie the yield response does not follow a systematic trend over time (see Figure 1.2). This is due to the nature of the trait where the yield data involves growth between cuts, with the cuts occurring at different spaced (time) intervals and the growth being very dependent on the environment and management of the trial during these different time intervals. Knowing the actual level at each time is not imperative, but the differential impact of each variety from the overall performance of all varieties is of most interest. For these reasons the ideal approach for analysing the yield data is to model the variety deviations from the harvest means. In terms of the mixed model (5.1.1) this means that the vector $\boldsymbol{\tau}$ contains the main effects for harvests.

5.3.2 Extensions for modelling non-genetic effects

A general model for R based on Diggle's model

As mentioned in section 5.2.1, the residual variance structure for a trial ($\boldsymbol{R}$), may be modelled using a separable spatio-temporal process, incorporating an unstructured matrix for the temporal correlation component $\boldsymbol{R}_h$ in (5.2.3). This is suitable in some cases where

harvest numbers are small but when more harvests are involved this model is not able to be fitted. A more parsimonious model is proposed, based on the approach of Diggle (1988).

In Chapter 1, the model proposed for longitudinal data by Diggle (1988) was reviewed. An extension of this model is required that incorporates the spatial dependence present in the motivating data sets. The Diggle model for y_{ij} (where y_{ij} denotes the j^{th} measurement on the i^{th} unit (or plot), with t_{ij} denoting the time that measurement y_{ij} was made) is given by

$$y_{ij} = \mu_{ij} + e_{ij}$$

with

$$e_{ij} = \zeta_i + \eta_{ij} + \xi_i(t_{ij})$$

where μ_{ij} is the mean at time t_{ij}, ζ_i is a unit effect (that implies uniform correlation across time), η_{ij} is a measurement error and $\xi_i(t_{ij})$ is a temporally correlated process to account for the serial correlation between measurements on the same unit. The latter three effects are random and for a single unit with h measurements, the variance matrix generated is

$$\text{var}\,(\boldsymbol{e}_i) = \sigma_p^2 \boldsymbol{J}_h + \sigma_m^2 \boldsymbol{I}_h + \sigma^2 \boldsymbol{R}_h^*(\phi) \tag{5.3.4}$$

where σ_p^2 is a between plot or unit variance, $\boldsymbol{J}_h$ is a hxh matrix with all elements equal to 1, σ_m^2 is the measurement error variance, σ^2 is the error scale parameter and $\boldsymbol{R}_h^*(\phi)$ is a smooth (typically) correlation structure over time. If $\boldsymbol{R}_h^*(\phi)$ is a variance-covariance matrix rather than a correlation matrix, then σ^2 must be equal to 1.

Diggle (1988) assumes independent units, however in the case of multi-harvest data the units are plots in the field trial which may be spatially correlated. Our model for multi-harvest data should include terms for plot effects, measurement error and serial dependence and allow for these effects to be spatially correlated. Thus, in the linear mixed model for multi-harvest data (5.1.1), the residual term $\boldsymbol{e}$ can be partitioned into a vector of random plot effects $\boldsymbol{\zeta}$, a temporal correlation process $\boldsymbol{\xi}$ and a vector of measurement errors $\boldsymbol{\eta}$, where each of these three random effects may have their own spatial structure. Hence

$$\boldsymbol{e} = \boldsymbol{\zeta} + \boldsymbol{\eta} + \boldsymbol{\xi}$$

where

($\boldsymbol{\zeta}$,$\boldsymbol{\eta}$,$\boldsymbol{\xi}$) are pairwise independent, mean zero and have variance matrix

$$\text{var}\left(\begin{bmatrix} \boldsymbol{\zeta} \\ \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{bmatrix}\right) = \begin{bmatrix} \sigma_p^2 \boldsymbol{J}_h \otimes \boldsymbol{\Sigma}_c^{(p)} \otimes \boldsymbol{\Sigma}_r^{(p)} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_m^2 \boldsymbol{I}_h \otimes \boldsymbol{\Sigma}_c^{(m)} \otimes \boldsymbol{\Sigma}_r^{(m)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \sigma^2 \boldsymbol{R}_h^*(\phi) \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r \end{bmatrix}$$

We note that if the same spatial correlation structure is assumed for the three random effects $(\boldsymbol{\zeta}, \boldsymbol{\eta}, \boldsymbol{\xi})$, the variance model generalizes to

$$\text{var}(\boldsymbol{e}) = (\sigma_p^2 \boldsymbol{J}_h + \sigma_m^2 \boldsymbol{I}_h + \sigma^2 \boldsymbol{R}_h^*(\phi)) \otimes \boldsymbol{\Sigma}_r \otimes \boldsymbol{\Sigma}_c$$

which is a separable spatial extension of (5.3.4) (the model of Diggle, 1988).

Whilst this separable extension may be theoretically appealing, it may be questionable as to whether the measurement error term should be spatially correlated. For a purely spatial (single harvest time) model the measurement error is assumed to be independently and identically distributed (i.i.d) random "white noise", and similarly in the temporal case with spatially independent measurements, the measurement error is also assumed to be i.i.d. It may be more reasonable to assume the measurement error term in the spatio-temporal context is also independent and not spatially correlated. It may however, be reasonable to assume the same spatial structure for the overall plot (or unit) effect and the plot by harvest effects and hence the variance matrix can be modified to

$$\text{var}(\boldsymbol{e}) = (\sigma_p^2 \boldsymbol{J}_h + \sigma^2 \boldsymbol{R}_h^*(\phi)) \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + \sigma_m^2 \boldsymbol{I}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{I}_r \tag{5.3.5}$$

Model (5.3.5) is still restrictive, and a more desirable model may be to allow $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ to have differing spatial correlation structures and to assume an independent measurement error. Thus we may assume

$$\text{var}(\boldsymbol{e}) = \sigma_p^2 \boldsymbol{J}_h \otimes \boldsymbol{\Sigma}_c^{(p)} \otimes \boldsymbol{\Sigma}_r^{(p)} + \sigma_m^2 \boldsymbol{I}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{I}_r + \sigma^2 \boldsymbol{R}_h^*(\phi) \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r \tag{5.3.6}$$

This model has the advantage of increased flexibility and an independent measurement error component.

Following Diggle (1988), $\boldsymbol{R}_h^*$ may be modelled using a decaying correlation model, which implies the correlation between harvests decreases as the time between harvests increases. For unequally spaced time points, the exponential (or power model) (exp) may be appropriate, while for equally (or close to) spaced measurements the autoregressive (ar1) correlation process may be suitable. This structure may be generalized to a heterogeneous variance process (e.g. ar1h, exph, or antedependence (ante) model) to account for differing variances at each harvest. Details of these models may be found in Appendix B and the notation is defined in Table 5.1.

The above model (based on 5.3.6), with $\boldsymbol{R}_h^*$ as an ar1h process (this requires $\sigma^2 = 1$), enables the spatial and temporal residual correlation structure to be modelled using a maximum of $h + 7$ parameters in comparison to the $h(h+1)/2 + 2$ parameters required for the separable unstructured by autoregressive by autoregressive model (5.2.3) as used in Smith et al. (2007). For a trial with 10 harvests this equates to a difference of 40 parameters.

Table 5.1: Model functions and their associated correlation/covariance matrices (or inverse covariance matrices where simpler) that are used in analysis of the lucerne and chicory breeding trials. $\boldsymbol{I}_h$ is the $h \times h$ identity matrix and $\boldsymbol{J}_h$ is the $h \times h$ matrix of 'ones'.

Name	Model Function	Correlation or variance	Example	correlation or covariance matrix
identity	id	correlation	id(Harvest)	$\boldsymbol{I}_h$
identity variance	idv	variance	idv(Harvest)	$\sigma^2_h \boldsymbol{I}_h$ [1]
diagonal	diag	variance	diag(Harvest)	$\text{diag}\left(\sigma^2_{hj}\right)$ [1,4]
homogeneous correlation	corv	variance	corv(Harvest)	$\sigma^2_h\{\boldsymbol{I}_h + \rho_t(\boldsymbol{J}_h - \boldsymbol{I}_h)\}$ [1,3]
heterogeneous correlation	corh	variance	corh(Harvest)	$\text{diag}\,(\sigma_{hj})\,\{\boldsymbol{I}_h + \rho_t(\boldsymbol{J}_h - \boldsymbol{I}_h)\}\text{diag}\,(\sigma_{hj})$ [2,3]
unstructured correlation	corgh	variance	corgh(Harvest)	$\text{diag}\,(\sigma_{hj})\,\boldsymbol{C}_h\text{diag}\,(\sigma_{hj})$ [2,5]
unstructured	us	variance	us(Harvest)	$\boldsymbol{G}_h$ [6]
factor analytic, order k	fa	variance	fa(Harvest,k)	$\boldsymbol{\Lambda}_h\boldsymbol{\Lambda}_h^T + \boldsymbol{\Psi}_h$ [7]
autoregressive, order 1	ar1	correlation	ar1(Row)	$\boldsymbol{\Sigma}_r^{ar1-1} = \boldsymbol{I}_r + \phi^2\boldsymbol{E}_1 - \phi\boldsymbol{F}_1$ [8,9]
autoregressive, order 2	ar2	correlation	ar2(Row)	$\boldsymbol{\Sigma}_r^{ar2-1} = \boldsymbol{I}_r + \sum_{j=1}^2 \phi_j^2\boldsymbol{E}_j - \sum_{j=1}^2 \phi_j\boldsymbol{F}_j$ [8,9]
autoregressive variance, order 1	ar1v	variance	ar1v(Row)	$\sigma^2_r\boldsymbol{\Sigma}_r^{ar1}$ [1,3,8]
autoregressive heterogeneous variance, order 1	ar1h	variance	ar1h(Row)	$\boldsymbol{D}\boldsymbol{\Sigma}_r^{ar1}\boldsymbol{D}$ [10]
autoregressive heterogeneous variance, order 2	ar2h	variance	ar2h(Row)	$\boldsymbol{D}\boldsymbol{\Sigma}_r^{ar2}\boldsymbol{D}$ [10]
exponential heterogeneous variance	exph	variance	exph(Harvest)	$\boldsymbol{D}\boldsymbol{\Sigma}_h^{exp}\boldsymbol{D}$ [10] where $\boldsymbol{\Sigma}_{\boldsymbol{h}}^{\boldsymbol{exp}}{}_{ij} = \phi^{\lvert t_i - t_j\rvert}$ [11]
antedependence, order s	ante	variance	ante(Harvest,s)	$\boldsymbol{\Sigma}_h^{ante}$ where $\boldsymbol{\Sigma}_h^{ante-1} = \boldsymbol{U}^T\boldsymbol{D}^*\boldsymbol{U}$ [12]

[1] σ^2_h, σ^2_{hj} and σ^2_r are variances ; [2] σ_{hj} are standard deviations; [3] ρ_t,ϕ and ϕ_j are correlations
[4] diag() is a diagonal matrix with elements specified
[5] $\boldsymbol{C}_t$ is a fully parameterized correlation matrix of order t
[6] $\boldsymbol{G}_t$ is a fully parameterized covariance matrix of order t
[7] $\boldsymbol{\Lambda}_t$ is a matrix of factor loadings, $\boldsymbol{\Psi}_t$ is a diagonal matrix of order t
[8] $\boldsymbol{F}_j$ is a matrix which has 'ones' on the jth sub and super-diagonals and zeros elsewhere
[9] $\boldsymbol{E}_j$ is like the identity matrix but with the first and last j 'ones' set to zero
[10] $\boldsymbol{D}$ is a diagonal matrix of standard deviations
[11] t_i and t_j are the harvest times of harvests i and j respectively
[12] $\boldsymbol{U}$ is a lower triangular matrix, $\boldsymbol{D}^*$ is a diagonal matrix as defined in Appendix B

5.3.3 Extensions for modelling genetic effects

When the number of harvests h becomes large, the approach of Smith et al. (2007) of estimating the genetic variance matrix, becomes difficult. It may be possible to estimate $\boldsymbol{G}_h$ using factor analytic models (Smith et al., 2001) and hence obtain variety predictions at each of the h harvest times. However, it may be more desirable for selection purposes to reduce the set of predictions to a smaller number. It may also be desirable to form predictions at times other than the harvest times and to also investigate variety by harvest interactions in more detail.

Two approaches that may may be suitable for modelling the genetic effects from multi-harvest data are detailed below. One approach is the method of random regression (as introduced in Chapter 2). A second approach (also introduced in Chapter 2) is to use factor analytic models to generate a reduced set of factors that describe the covariance structure between the genetic effects at the different harvest times. These factors and their loadings may be interpreted to identify traits that separate out the varieties, with the traits then being used in the modelling and prediction process, or alternatively the harvests may be clustered into groups and predictions formed for these groups.

Random regression

A suitable model for estimating the genetic response over time is the random regression (or random coefficients) model (Laird & Ware, 1982). Random regression analysis is a common approach for the genetic analysis of repeated animal measurements over time (Schaeffer, 2004). Random regression models involve fitting regression coefficients on time (or other explanatory variables), for each variety, as random effects. This allows for variation between varieties in the shape of the response profile over time.

Polynomial random regression

Let g_{ik} denote the random effect for variety i for harvest k (where $k = 1, \ldots, h$), and x_k represent the value for the explanatory variable x for harvest k, then a polynomial random regression model of order p, over x for genetic effects g_{ik} can be formulated as

$$\begin{aligned} g_{ik} &= u_{i0} + u_{i1}x_k + \ldots + u_{ip}x_k^p + \epsilon_{ik} \\ &= \boldsymbol{x}_k^T \boldsymbol{u}_i + \epsilon_{ik} \end{aligned} \tag{5.3.7}$$

where $\boldsymbol{x}_k^T = \begin{bmatrix} 1 & x_k & x_k^2 & \ldots & x_k^p \end{bmatrix}$ and $\boldsymbol{u}_i = \begin{bmatrix} u_{i0} & u_{i1} & u_{i2} & \ldots & u_{ip} \end{bmatrix}^T$. The term ϵ_{ik} represents a residual term for genetic effects, assumed to be independent and identically distributed, with variance σ_ϵ^2. This model can also be written as

$$\boldsymbol{g}_i = \boldsymbol{X}_p \boldsymbol{u}_i + \boldsymbol{\epsilon}_i$$

where

$$\boldsymbol{X}_p = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_h^T \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^p \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_h & x_h^2 & \dots & x_h^p \end{bmatrix} \tag{5.3.8}$$

If $\boldsymbol{u}_i \sim N(\boldsymbol{0}, \boldsymbol{G}_p)$, then

$$\boldsymbol{G}_h = \text{var}\,(\boldsymbol{g}_i) = \boldsymbol{X}_p \boldsymbol{G}_p \boldsymbol{X}_p^T + \sigma_\epsilon^2 \boldsymbol{I}_h$$

The covariance matrix of random polynomial terms, $\boldsymbol{G}_p$ is taken as an unstructured matrix to ensure invariance to translation.

In the case of $p = 1$, (5.3.7) reduces to the linear random regression

$$g_{ik} = u_{i0} + u_{i1} x_k + \epsilon_{ik} \tag{5.3.9}$$

where u_{i0} and u_{i1} are the random intercept and slope terms (respectively) for variety i.

Cubic smoothing spline random regression

While the polynomial random regression models above, may be used in instances to model nonlinear trends, it may be preferable to use natural cubic smoothing splines (Verbyla et al., 1999) to provide a more flexible specification.

The standard approach to fitting cubic smoothing splines in a linear mixed model is to fit a fixed linear component and a random "spline" component, as in (1.3.8). Alternatively a cubic smoothing spline random regression model may be fitted with random linear and spline components.

A random regression model incorporating cubic smoothing splines, for g_{ik} (the random effect for variety i at harvest k, $k = 1, \dots, h$), with x_k denoting the explanatory variable x (e.g. time) at harvest k, can be written as

$$\begin{aligned} g_{ik} &= u_{i0} + u_{i1} x_k + \boldsymbol{z}_{s_k}^T \boldsymbol{u}_{s_i} + \epsilon_{ik} \\ &= \boldsymbol{x}_k^T \boldsymbol{u}_i + \boldsymbol{z}_{s_k}^T \boldsymbol{u}_{s_i} + \epsilon_{ik} \end{aligned}$$

where $\boldsymbol{x}_k^T = \begin{bmatrix} 1 & x_k \end{bmatrix}$ and $\boldsymbol{u}_i = \begin{bmatrix} u_{i0} & u_{i1} \end{bmatrix}^T$, with u_{i0} and u_{i1} denoting the random intercepts and slopes for variety i respectively. For each variety i, $\boldsymbol{u}_{s_i}$ (a $(h-2) \times 1$ vector) is the random spline component of the mixed model formulation of the cubic smoothing spline as presented in (1.3.8). The vector $\boldsymbol{z}_{s_k}^T$ is the k^{th} row of the matrix $\boldsymbol{Z}_s$ in (1.3.8).

Therefore for variety i the random regression spline model is

$$\boldsymbol{g}_i = \boldsymbol{X}_1 \boldsymbol{u}_i + \boldsymbol{Z}_s \boldsymbol{u}_{s_i} + \boldsymbol{\epsilon}_i$$

where

$$\boldsymbol{Z}_s = \begin{bmatrix} \boldsymbol{z}_{s_1}^T \\ \vdots \\ \boldsymbol{z}_{s_h}^T \end{bmatrix}$$

and the matrix $\boldsymbol{X}_1$ is the design matrix for the linear random regression ($p = 1$) of (5.3.8). The distribution of $\boldsymbol{u}_{s_i}$ is assumed to be $\boldsymbol{u}_{s_i} \sim N(0, \sigma_s^2 \boldsymbol{G}_s)$, $\epsilon_i \sim N(0, \sigma_\epsilon \boldsymbol{I})$ and $\boldsymbol{Z}_s = \boldsymbol{Q}(\boldsymbol{Q}^T\boldsymbol{Q})^{-1}$ where $\boldsymbol{G}_s$ and $\boldsymbol{Q}$ are defined in (1.3.6) and (1.3.7) in Chapter 1.

As mentioned above, the usual mixed model formulation of cubic smoothing splines has fixed effects for the intercept and slope. Verbyla et al. (1999) do consider the random linear regression plus independent random spline component but note that there may be issues with this model due to a lack of invariance to a change of basis. This issue is discussed in White et al. (1998) where different formulations of spline models are compared and shown to be equivalent only if the linear trend and spline parameters are correlated and the full set of covariances are estimated.

To incorporate these additional covariances of the cubic smoothing spline it is simpler to consider an alternative formulation (used in the ASReml software). In this formulation $\boldsymbol{u}_{s_i} \sim N(\boldsymbol{0}, \sigma_s^2 \boldsymbol{I})$ and $\boldsymbol{Z}_s = \boldsymbol{Q}(\boldsymbol{Q}^T\boldsymbol{Q})^{-1}\boldsymbol{L}_s$, where $\boldsymbol{G}_s = \boldsymbol{L}_s\boldsymbol{L}_s^T$ is the Cholesky decomposition of $\boldsymbol{G}_s$. Using this form, $\boldsymbol{u}_i$ and $\boldsymbol{u}_{s_i}$ from the random regression model are assumed to follow a Normal distribution with zero mean vector and variance-covariance matrix given by

$$\text{var}\left(\begin{bmatrix} \boldsymbol{u}_i \\ \boldsymbol{u}_{s_i} \end{bmatrix}\right) = \begin{bmatrix} g_{11} & g_{12} & \sigma_{11s} & \sigma_{12s} & \dots & \sigma_{1(h-2)s} \\ g_{12} & g_{22} & \sigma_{21s} & \sigma_{22s} & \dots & \sigma_{2(h-2)s} \\ \sigma_{11s} & \sigma_{21s} & \sigma_s^2 & 0 & \dots & 0 \\ \sigma_{12s} & \sigma_{22s} & 0 & \sigma_s^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ \sigma_{1(h-2)s} & \sigma_{2(h-2)s} & 0 & 0 & \dots & \sigma_s^2 \end{bmatrix}$$

where g_{11} and g_{22} represent the variance of the random intercepts and slopes, respectively and g_{12} the covariance between them, σ_{1js} represents the covariance between the random intercepts and j^{th} element of $\boldsymbol{u}_{s_i}$, σ_{2js} represents the covariance between the random slopes and j^{th} element of $\boldsymbol{u}_{s_i}$, and σ_s^2 represents the variance of the elements of $\boldsymbol{u}_{s_i}$.

Note that while the cubic smoothing spline random regression may model the variety responses over time reasonably well, the approach has limitations in how well it models the underlying covariance structure. The above covariance structure is quite restrictive and there may be alternative models such as the Matern class of covariance functions (Haskard et al., 2007) that are more flexible in their covariance modelling. There are variants of the Matern model that are differentiable and continuous over time and hence may be used to model multi-harvest data over time.

Factor analytic models

The application of factor analytic (fa) models to multi-environment trials is outlined in Chapter 1. Similar principles could apply in the case of multi-harvest data where measurements from different harvests can be regarded as separate traits and variety effects at different harvest times are assumed correlated.

Hence a fa model (of order s) can be fitted to the variety effects at each harvest, with the genetic effects given by

$$g_{ik} = \sum_{r=1}^{s} \lambda_{kr} f_{ir} + \delta_{ik} \tag{5.3.10}$$

where g_{ik} is the random effect for variety $i, i = 1, \ldots, m$ at harvest k, $k = 1, \ldots, h$, f_{ir} is the score for variety i in the r^{th} factor, λ_{kr} is the loading for harvest k for the r^{th} factor and δ_{ik} is a residual.

In vector notation the genetic effects are given by

$$\boldsymbol{g} = (\boldsymbol{\Lambda} \otimes \boldsymbol{I}_m)\boldsymbol{f} + \boldsymbol{\delta}$$

where $\boldsymbol{\Lambda}^{h \times s} = [\boldsymbol{\lambda}_1 \boldsymbol{\lambda}_2 \ldots \boldsymbol{\lambda}_s]$, where $\boldsymbol{\lambda}_r$ is a $t \times 1$ vector of loadings, $\{\lambda_{jr}\}$, $\boldsymbol{f}^{ms \times 1} = [\boldsymbol{f}_1^T \boldsymbol{f}_2^T \ldots \boldsymbol{f}_s^T]^T$, where $\boldsymbol{f}_r$ is a $m \times 1$ vector of factor scores, $\{f_{ir}\}$, and $\boldsymbol{\delta}$ is a $mh \times 1$ vector of residuals, $\{\delta_{ik}\}$.

The joint distribution of $\boldsymbol{f}$ and $\boldsymbol{\delta}$ is assumed to be

$$\begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{\delta} \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{bmatrix} \boldsymbol{I}_s \otimes \boldsymbol{I}_m & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Psi} \otimes \boldsymbol{I}_m \end{bmatrix} \right]$$

where $\boldsymbol{\Psi}^{h \times h}$ is a diagonal matrix of so called specific variances. Hence the variance matrix for the vector $\boldsymbol{g}$ of genetic effects is given by

$$\begin{aligned} \text{var}\,(\boldsymbol{g}) &= (\boldsymbol{\Lambda} \otimes \boldsymbol{I}_m)\text{var}\,(\boldsymbol{f})\,(\boldsymbol{\Lambda}^T \otimes \boldsymbol{I}_m) + \text{var}\,(\boldsymbol{\delta}) \\ &= (\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}) \otimes \boldsymbol{I}_m \\ &= \boldsymbol{G}_h \otimes \boldsymbol{I}_m \end{aligned}$$

In theory, this enables predictions for each variety at each harvest ($\boldsymbol{g}$), to be calculated and these predictions can be combined to form a weighted selection index (see Smith et al., 2007). However in practice, when the number of harvests is large this may not be ideal and it may be desirable to form a selection index based on a smaller number of "traits". Exploiting the regression interpretation of the fa model may provide a method for doing this.

The specification of the fa model in Equation (5.3.10) has the form of a (random) regression on s covariates, $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_s$, where the loading λ_{kr} is the value of the covariate r for harvest k and the random regression coefficient for variety i is given by f_{ir}, and the lack of fit of the model is given by δ_{ik}.

So for example in the case of a 2 factor fa model

$$g_{ik} = \lambda_{k1} f_{i1} + \lambda_{k2} f_{i2} + \delta_{ik} \tag{5.3.11}$$

and so the similarity with the linear random regression model given in Equation (7.4.3) can be seen. If all λ_{k1} are similar in value and of the same sign, the first term in (5.3.11) can be interpreted as an intercept term in a linear random regression with f_{i1} representing the random intercepts for variety i. The second term can be interpreted as comprising of a random slope f_{i2} for each variety i and λ_{k2} the explanatory variable that is regressed upon. For example if the loadings λ_{k2} are related to time of harvest then the model is like a random regression on time. Alternatively if the loadings λ_{k2} are related to the mean at each harvest then the model indicates a random regression on harvest mean yield may be appropriate. If no pattern is clear from the loadings the approach of Cullis et al. (2010) may be used, in which cluster analysis is performed to group the harvests into target groups and variety predictions obtained for each selection group. Hardner et al. (2010) implements a similar clustering approach to investigate genotype by environment interaction, after fitting a factor analytic model, to MET tree breeding data.

5.4 Summary

In this chapter an approach for the analysis of multi-harvest data, based on a linear mixed model has been presented. The approach requires the genetic and non-genetic effects to be modelled. A new model for the residual (non-genetic) effects has been developed based on a spatial extension of the Diggle (1988) model for repeated measurements. The genetic effects can be modelled in various ways depending on the trait of interest. Random regression, splines and factor analytic models have been presented for modelling these genetic effects.

The extension of the Diggle model developed in this thesis for modelling spatially correlated repeated measures data provides a new, more parsimonious approach to modelling multi-harvest data than the models introduced in Smith et al. (2007). The approach is able to handle data from a large number of harvests and accommodates all the major sources of known spatial and temporal correlation.

Like the models in Smith et al. (2007), the models contain separable spatial by temporal structures, which while being attractive for ease of interpretation and computational advantages (Galecki, 1994), are based on the assumption of separability of spatial and temporal processes, which may not hold in all instances (Smith et al., 2007). We have seen in Chapter 3 that the spatial correlation parameters may vary between harvests within a trial and assuming constant spatial correlation across the harvest times may not be sensible. Alternative, non separable models may need to be investigated. In the following chapter the new extended separable residual models will be applied to the lucerne data. In further chapters non-separable models are investigated.

At the genetic level parsimonious models have been presented for modelling the variety

by harvest effects. These methods extend the ideas of Smith et al. (2007). Approaches to modelling variety response profiles over time for traits that are smooth and continuous over time and where interest is in predicting at times other than the harvest times have been presented. Factor analytic models have been investigated and their interpretation as a regression model discussed. As an aid to interpreting the results from the factor analytic models the approach of Cullis et al. (2010) and Hardner et al. (2010) may be used. These approaches allow for investigation into variety by harvest interactions. These approaches are novel to the analysis of multi-harvest data in perennial crop variety selection trials.

Chapter 6

Analysis of multi-harvest data using separable variance models : Examples

6.1 Introduction

In Chapter 5, new approaches for modelling the genetic and residual variation in multi-harvest data were presented that extend the approach of Smith et al. (2007). The issue of whether to model the deviations or contrasts from the harvest means or to model the response over time was discussed. In both approaches the residual models presented were based on the assumption of separability between temporal and spatial correlation processes. In this chapter these new methods are applied to the analysis of data from a single site of the motivating lucerne yield and persistence data.

6.2 Lucerne trial at Terry Hie Hie

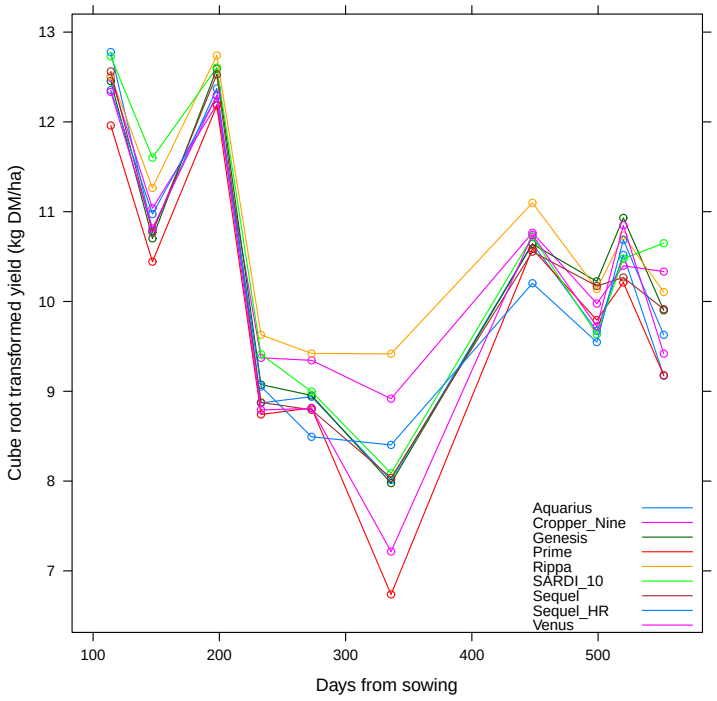
The motivating data considered in this section arises from the lucerne variety assessment trial at Terry Hie Hie, NSW ($30^o48^{'}S150^o09^{'}E$). The data of interest is the yield and persistence of each variety, measured on each plot at multiple times, as detailed in Tables 1.1 and 1.2. The trial was designed as a Randomized Complete Block (RCB) with 3 replicates and was laid out in a rectangular array of 180 plots consisting of 30 rows by 6 columns. The number of varieties tested in the trial was 60 (with eleven being commercial varieties).

6.2.1 Yield

The aim of the analysis of yield data was to identify the best yielding lucerne varieties over the course of the trial and to identify harvest times that discriminate between the varieties and subsequently investigate those varieties that perform best at these times during the trial. For example, winter activity is an important trait in lucerne breeding with varieties ranging from highly winter active to dormant, and hence variety performance over the winter harvests was of interest. It was also of interest how the variety performance changed over harvest times, that is to investigate variety by harvest interaction.

Yield was measured by cutting all trial plots at a consistent defined height at each harvest time and drying the samples to obtain dry matter weights expressed as kg/ha. There were 10 harvest times. The data was transformed prior to analysis using a cube root transformation ($y_{transf} = (y+1)^{\frac{1}{3}}$), to stabilize the variance (as was the case for the analysis of the data from individual harvests, in Chapter 3). A plot of the transformed yield data for the 11 commercial varieties is presented in Figure 6.1.

Figure 6.1: Plot of mean lucerne yield (on transformed scale) for each of the 11 commercial varieties across harvests at Terry Hie Hie



As can be seen in Figure 6.1 the yield response profile over time is not a smooth curve and modelling this profile using a polynomial or cubic smoothing spline is not suitable. There are also good biological reasons (e.g. seasonal influences, variable harvest intervals) why the actual level is not of interest. As discussed in the previous chapter, the ideal approach is to model the variety deviations from the overall means at each harvest time.

The analysis of yield was based on the linear mixed model (5.1.1). This analysis requires appropriate variance models for the genetic and residual effects (that is, suitable models for $\boldsymbol{R}$ and $\boldsymbol{G}_g$). As it is difficult to find optimal models for both $\boldsymbol{R}$ and $\boldsymbol{G}_g$ at the same time, the sequential approach of Smith et al. (2007) was followed by firstly assuming a simple genetic model, in order to determine a suitable residual model. Using this residual model more complex models for the genetic effects were investigated. Care must be taken with this approach. If the genetic effects are not modelled correctly this will have an impact on the residual variation, and terms in the residual model may change in significance given changes in the genetic model. Hence, once more complex genetic models were fitted, the final form of the residual variance model was re-examined. The

initial simple genetic model has genetic variance matrix $\boldsymbol{G}_g$ given by

$$\boldsymbol{G}_g \;=\; diag(\sigma^2_{gk}) \otimes \boldsymbol{I}_m \tag{6.2.1}$$

where σ^2_{gk} is the genetic variance for harvest k. This initial genetic model therefore allows for a different genetic variance for each harvest and assumes the genetic effects are independent between harvests. It also assumes the varieties are unrelated.

The description of the sequential model-building process has been separated into two sections below, namely modelling the non-genetic (residual) effects and modelling the genetic effects, with the residual models given in Table 6.1 and the genetic models given in Table 6.3. A factor for harvests is fitted as a fixed effect in all models to ensure modelling is based on deviations from the harvest means.

The terms used in the residual and genetic models are defined briefly in in Table 5.1 and further in Appendix B.

Modelling non-genetic effects

The first step in modelling the non-genetic (residual) effects was to account for any spatial variation in the data. Initially this involved investigating spatial models for each harvest separately, as discussed in Chapter 3. In this case the residual variance matrix $\boldsymbol{R}$ was given by a block diagonal matrix

$$\boldsymbol{R} = \text{diag}\,(\boldsymbol{R}_k)$$

where $\boldsymbol{R}_k$ denotes the residual variance matrix for harvest k. $\boldsymbol{R}_k$ is given by

$$\boldsymbol{R}_k = \sigma^2_k \boldsymbol{\Sigma}_{kc} \otimes \boldsymbol{\Sigma}_{kr} \tag{6.2.2}$$

where $\boldsymbol{\Sigma}_{kc}$ and $\boldsymbol{\Sigma}_{kr}$ are the spatial correlation matrices in the column and row directions respectively, for harvest k.

The initial model fitted to each of the harvests was a spatial model based on the results of the analyses in Chapter 3. This model modelled the local spatial correlation between plots at each harvest time using a separable autoregressive (ar1(Column).ar1(Row)) process and also included terms to account for global and extraneous spatial variation at each harvest time (identified in Chapter 3). These terms included a random column effect for harvests $2, 3, 4, 5, 6, 8$ (included in $\boldsymbol{u_o}$) and a fixed effect for row 1 for harvest 10.

This represents the baseline model (Y1). In this model (and Y2-Y14, see Table 6.1) the simple genetic model (6.2.1) was included at the genetic level.

Model Y2 fitted a common separable ar1(Column)×ar1(Row) spatial model across the harvests and allowed for heterogeneity of variance across time. This model was not significantly different to model Y1 (REMLRT = 17.17, on 18df P=0.511) but required many less parameters.

Models Y1 and Y2 did not attempt to account for the temporal correlation between

Table 6.1: Summary of residual models fitted for yield at Terry Hie Hie. In all models the simple genetic effects model of (6.2.1) has been fitted and a random Replicate effect has been included for each harvest. All models include global and extraneous spatial effects identified in Chapter 3 (Table 3.2). Residual log-likelihoods (denoted by ℓ), number of parameters in $\boldsymbol{R}$ (npar), number of other non-genetic variance parameters (Other npar) and AIC values (given as differences from the best model) are presented. Model terms are detailed in Table 5.1. Superscripts on ar1 parameters $ar1^1$ denote that within a model the spatial parameters with the same superscript are constrained to be equal.

Model	Residual variance matrix $\boldsymbol{R}$		ℓ	$\boldsymbol{R}$ npar	Other npar	AIC
Y1	$\text{diag}(\sigma^2_k \boldsymbol{\Sigma}_{kc} \otimes \boldsymbol{\Sigma}_{kr})$	at(Harvest).ar1(Col).ar1(Row)	606.156	30	16	714.8
Y2	$\text{diag}(\sigma^2_k \boldsymbol{\Sigma}_{kc} \otimes \boldsymbol{\Sigma}_{kr})$	diag(Harvest).ar1(Col).ar1(Row)	597.859	12	16	x
Y3	$\sigma^2 \Sigma_h \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$	ar1v(Harvest).ar1(Col).ar1(Row)	801.129	4	16	272.8
Y4	$\boldsymbol{D}\boldsymbol{\Sigma}_h\boldsymbol{D} \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$	ar1h(Harvest).ar1(Col).ar1(Row)	834.460	13	16	224.1
Y5	$(\boldsymbol{D}\boldsymbol{\Sigma}_h\boldsymbol{D} + \sigma^2_p \boldsymbol{J}_h) \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$ ar1h(Harvest).ar1^1(Col).ar1^2(Row) + ar1^1(Col).ar1v^2(Row)		916.840	14	16	61.4
Y6	$(\boldsymbol{D}\boldsymbol{\Sigma}_h\boldsymbol{D} + \sigma^2_p \boldsymbol{J}_h + \sigma^2_m \boldsymbol{I}_h) \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$ ar1h(Harvest).ar1^1(Col).ar1^2(Row) + ar1^1(Col).ar1v^2(Row) + Harvest.ar1^1(Col).ar1v^2(Row)		-	15	16	-
Y7	$(\boldsymbol{D}\boldsymbol{\Sigma}^{ar2}_h\boldsymbol{D} + \sigma^2_p \boldsymbol{J}_h) \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$ ar2h(Harvest).ar1^1(Col).ar1^2(Row) + ar1^1(Col).ar1v^2(Row)		925.969	15	16	45.1
Y8	$(\boldsymbol{D}\boldsymbol{\Sigma}^{ar2}_h\boldsymbol{D} + \sigma^2_p \boldsymbol{J}_h + \sigma^2_m \boldsymbol{I}_h) \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$ ar2h(Harvest).ar1^1(Col).ar1^2(Row) + ar1^1(Col).ar1v^2(Row) + Harvest.ar1^1(Col).ar1v^2(Row)		926.080	16	16	46.9
Y9	$(\boldsymbol{D}\boldsymbol{\Sigma}^{ar2}_h\boldsymbol{D} + \sigma^2_p \boldsymbol{J}_h) \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + \sigma^2_m \boldsymbol{I}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{I}_r$ ar2h(Harvest).ar1^1(Col).ar1^2(Row) + ar1^1(Col).ar1v^2(Row) + Harvest.Col.Row		-	16	16	-
Y10	$\boldsymbol{D}\boldsymbol{\Sigma}^{ar2}_h\boldsymbol{D} \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}_r + \sigma^2_p \boldsymbol{J}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}^{(p)}_r + \sigma^2_m \boldsymbol{I}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{I}_r$ ar2h(Harvest).id(Col).ar1(Row) + id(Col).ar1v(Row) + Harvest.Col.Row		927.338	16	16	44.4
Y11	$(\boldsymbol{\Sigma}^{ante2}_h + \sigma^2_p \boldsymbol{J}_h + \sigma^2_m) \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$ ante2(Harvest).ar1^1(Col).ar1^2(Row) + ar1^1(Col).ar1v^2(Row) + Harvest.ar1^1(Col).ar1^2(Row)		960.814	31	16	7.4
Y12	$(\boldsymbol{\Sigma}^{ante2}_h + \sigma^2_p \boldsymbol{J}_h) \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + \sigma^2_m \boldsymbol{I}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{I}_r$ ante2(Harvest).ar1^1(Col).ar1^2(Row) + ar1^1(Col).ar1v^2(Row) + Harvest.Col.Row		962.776	31	16	3.5
Y13	$\boldsymbol{\Sigma}^{ante2}_h \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + \sigma^2_p \boldsymbol{J}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}^{(p)}_r + \sigma^2_m \boldsymbol{I}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{I}_r$ ante2(Harvest).ar1(Col).ar1(Row) + id(Col).ar1v(Row) + Harvest.Col.Row		965.309	32	16	0.5
Y14	$\boldsymbol{\Sigma}^{ante2}_h \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + \sigma^2_p \boldsymbol{J}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}^{(p)}_r$ ante2(Harvest).ar1(Col).ar1(Row) + id(Col).ar1v(Row)		964.532	31	16	0

the multiple harvests. Subsequent models allowed for such correlation at the residual level.

Table 6.2: Lower triangle of the sample estimate of the between harvest variance matrix (variances on the diagonals, correlations below the diagonal) based on predictions of the residuals from from model Y1

	1	2	3	4	5	6	7	8	9	10
1	0.241									
2	0.481	0.253								
3	0.475	0.597	0.155							
4	0.354	0.469	0.393	0.114						
5	0.470	0.420	0.440	0.576	0.096					
6	0.288	0.376	0.280	0.594	0.523	0.115				
7	0.438	0.359	0.405	0.501	0.550	0.409	0.130			
8	0.303	0.280	0.392	0.496	0.557	0.478	0.615	0.153		
9	0.344	0.304	0.372	0.308	0.456	0.245	0.543	0.479	0.095	
10	0.351	0.405	0.373	0.386	0.440	0.174	0.420	0.358	0.273	0.204

To decide on an appropriate model for the residual covariance matrix, the empirical variance and covariance matrix of the residuals from Y1 was calculated. The empirical variances and correlations are presented in Table 6.2. This matrix provides an approximate guide to models for the residual variance structure. From these values it is apparent that there is positive correlation between all harvests with the correlations generally decaying as the time between measurements increases, but that the correlations do not decay to zero.

Based on these observations, a plausible model is to assume an overall average plot effect (Column.Row), a decaying correlation process for the plot by harvest effects and a measurement error effect (similar to the repeated measures model of Diggle, 1988), with spatial extensions to incorporate the spatial correlation between plots, as discussed in the previous chapter.

To build up to this model, firstly a simpler model, similar to that used by Smith et al. (2007) was fitted (Y3), using a 3 way separable process (ar1v(Harvest).ar1(Column).ar1(Row)) for the plot by harvest effects, thereby modelling the temporal correlation and the spatial correlation in each direction (row and column) each with an autoregressive process of order 1, with a common residual variance across the harvests. In this model the spatial correlation model for plots was common for all harvests. The variances for each harvest in Table 6.2 (ranging from 0.095 to 0.253), indicate that it may be more suitable to assume heterogeneous residual variances for each harvest, despite the transformation to cube root. Model Y4 allowed for differing variances across the harvests and was a significant improvement on model Y3. Model Y5 incorporated an overall average plot effect and allowed for spatial correlation at the overall average plot level. In this model the spatial correlation at the overall plot level was constrained to be the same as at the plot by harvest level. It is clear that this model is a significant improvement on previous models based on AIC values.

Model Y6 fitted a fully separable spatial extension of the repeated measures model of Diggle (1988), with an heterogeneous variance autoregressive process (ar1h) for the temporal serial correlation. As discussed in Chapter 5, this separable model imposes the same spatial correlation structure on the plot by harvest effects, the overall plot effect and the measurement error term. This spatial extension of the measurement error term may not be ideal and a more suitable model may be one which also models the temporal serial correlation using an ar1h process and constrains the spatial parameters to be the same for the plot by harvest effects and the overall plot effect, but fits an independent measurement term. Alternatively an unconstrained model which allows the spatial correlation parameters for the overall plot effect to differ from the spatial parameters for the plot by harvest effects may be more suitable. Attempts were made to fit all three above models, but there were problems in fitting the models to the lucerne yield data. These problems are most likely due to a combination of two things, firstly due to the fact that at harvest 9 there was very little spatial correlation (evident from the individual harvest analyses in Chapter 3), which results in the residual variance for harvest 9 going to zero and secondly the temporal correlation structure may not be adequately modelled using the autoregressive model of order 1 (ar1).

On closer inspection of the empirical correlations of residuals in Table 6.2, it appears that while the correlations are decaying, they are not decaying as quickly as what would be expected from an ar1 process and it may be more reasonable to fit a correlation process of higher order (for example ar2).

The next model to be fitted (Y7) fitted a model similar to Y5 but with an ar2 correlation process for the temporal correlation rather than an ar1 process. This model was a significant improvement on Y5 as can be seen by the Residual Maximum Likelihood Ratio Test Statistic (REMLRS) of 18.258 on 1 df ($P<0.001$).

The next model (Y8) fitted a fully separable model similar to Y6 but with an ar2 correlation process for the temporal correlation rather than an ar1 process. This model fitted successfully but was not an improvement on Y7 based on AIC values. As mentioned above a suitable extension to this model may be to allow for an independent measurement error term but still constrain the spatial parameters to be the same for the plot by harvest effects and the overall plot effect. This model was fitted as Y9. Unfortunately there were problems with this model similar to Y5 with the REML estimate of the variance for harvest 9 constrained at the boundary of the parameter space (zero).

The next model (Y10) retained the ar2 correlation process for the temporal correlation but it allowed the spatial correlation parameters for the overall plot effect to differ from the spatial parameters for the plot by harvest effects. In this model the spatial correlation at the overall plot level in the column direction was dropped (set to zero) in order to achieve convergence (this is not surprising as the correlation in the column direction was very small as evidenced in Y8 where the estimated correlation was $\hat{\phi}_c = 0.066$).

Different processes were fitted for the temporal correlation component in subsequent models. The exponential model (exph) was considered (similar to that in Diggle (1988)) but unfortunately similar problems were found as for the ar1 process above. It would seem

reasonable to consider the exponential model as the harvest times are unevenly spaced and the exponential model takes into account the time between measurements. However in models without the measurement error component fitted (which did converge) the exponential model did not fit as well as the ar1 model. An explanation for this may be that while the number of days between harvest times differ, the harvests have been taken after similar growth periods (when the varieties have reached a certain level of growth) and hence may be considered to be evenly spaced with regards to stage of growth.

The antedependence structure was also investigated for modelling the temporal covariance component. An antedependence model of order s assumes that the j^{th} observation ($j > s$), given the s proceeding observations, is independent of all other proceeding observations (Gabriel, 1962). The model is more flexible than the exponential or autoregressive models in that it allows the variances for each harvest time to differ and allows for different antedependence coefficients for each harvest. Initially the antedependence model of order 1 (ante) was tried but resulted in similar problems to the ar1 process above. The antedependence model of order 2 (ante2) proved a better fit and models Y11,Y12,Y13 which incorporate the ante2 model were better than previous models based on AIC values. Model Y11 fitted a fully separable model incorporating an ante2 model for the temporal correlation component and constrained the spatial correlation parameters for the plot by harvest effects, the overall plot effect and the measurement error term to be the same. The next model, Y12, constrained the spatial correlation parameters in the row and column directions to be the same for the overall plot and the plot by harvest effects but allowed for an independent measurement error term. This model had the same number of parameters as model Y11 and resulted in a better fit (based on AIC values). Model Y13 allowed the spatial correlation parameters to differ between the plot and plot by harvest effects. In this model the spatial correlation at the overall plot level in the column direction was set to zero to achieve model convergence. The model therefore required one extra parameter to be estimated than Y12. Despite this extra parameter the model was a significant improvement on model Y12 based on AIC values. The final model in Table 6.1 (Y14) dropped the measurement error term from Y13 but did not result in a significant drop in log-likelihood (P=0.106). Hence the final 'best' residual model was deemed to be model Y14 and this model will be used in the following section where more complex genetic models are incorporated.

Modelling genetic effects

The first attempt at improving the genetic model from the simple model of (6.2.1), was to fit an overall variety main effect plus variety by harvest interaction model (Y15 in Table 6.3). This model allowed for heterogeneous genetic variances across the harvests and assumed a common genetic covariance between each pair of harvests. This model was not a significant improvement on the previous model but it was useful to provide starting values for subsequent models in the model building process.

Attempts were made to fit an antedependence model for the genetic effects (given the time ordering of the genetic effects) but there were problems in fitting the model to this

Table 6.3: Summary of genetic models fitted for yield at Terry Hie Hie. Residual log-likelihoods (denoted by ℓ), number of parameters in $\boldsymbol{G}_g$ (npar), AIC values (given as differences from the best model), are presented for each model.

Model	Genetic Model	$\boldsymbol{G}_g$	$\boldsymbol{G}_g$ npar	ℓ	AIC
Y14	diag(Harvest).Variety	$diag(\sigma^2_{gk}) \otimes \boldsymbol{I}_m$	10	964.532	35.4
Y15	Variety + diag(Harvest).Variety	$(\sigma^2_g \boldsymbol{J}_h + diag(\sigma^2_{gk})) \otimes \boldsymbol{I}_m$	11	964.722	37.0
Y16	fa(Harvest,1).Variety	$(\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1^T + \boldsymbol{\Psi}) \otimes \boldsymbol{I}_m$	20	992.237	0
Y17	fa(Harvest,2).Variety	$(\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_2^T + \boldsymbol{\Psi}) \otimes \boldsymbol{I}_m$	29	998.465	5.5

where
all models have residual model of Y14 in Table 6.1
$\boldsymbol{\Lambda}_1$ is a $h \times 1$ matrix of factor loadings
$\boldsymbol{\Lambda}_2$ is a $h \times 2$ matrix of factor loadings
$\boldsymbol{\Psi} = \text{diag}\,(\psi_1, \ldots, \psi_h)$ is a diagonal matrix of specific variances

data, perhaps due to the very low genetic variance at some harvests.

The next genetic model to be fitted was the factor analytic model. The fitting of the factor analytic model had two purposes: firstly it enables the h by h genetic variance matrix to be estimated and hence variety predictions to be obtained for each of the harvest times. Secondly it may be used to identify a reduced number of "traits" that separate out the varieties. The number of factors required to explain a sufficient amount of the variation and the interpretation of their loadings are both of interest.

In model Y16 a factor analytic model with a single factor (fa1) model was fitted. As model Y15 is nested within the fa1 model a direct comparison can be made using a REMLRT. The fa1 model provided a significant improvement in log-likelihood (REMLRT = 55.030 on 9 df, $P < 0.001$). The percentage variance accounted for (%VAF) by the fa1 model was low for some of the harvests and a more desirable model would be to fit a factor model with two factors (fa2).

Model Y17 fitted a fa2 model, which whilst not providing a significant improvement in log-likelihood to the fa1 model, (REMLRT= 12.456 on 9 df, $P = 0.189$), explained a much greater percentage of variation for many of the harvests which had low %VAF with the fa1 model. The fa1 model also implied a very simplistic and restrictive structure with the first 4 harvests being perfectly correlated (genetic correlation =1) which is unlikely to be the case biologically. In the fa2 model fitted to this data the correlations are shown to be quite different from 1. The fa2 model was therefore chosen as the most suitable model as the resulting genetic correlation structure between harvests made more sense biologically and the model explained more of the total genetic variance. Note Beeck et al. (2010) also use the %VAF as a tool to aid in model selection for factor analytic models.

The loadings and percent variance accounted for (%VAF) by the two factor model is given in Table 6.4. The first factor shows Harvest 6 (the Winter harvest) as being most important, with the highest genetic variance. This factor also shows harvest 9 being

negatively correlated with the other harvests (but with low genetic variance). The second factor is more difficult to interpret but it may be interpreted as a contrast between harvests 1 and 2 and harvests $3, 5, 8$ and 9, which may reflect an establishment effect.

A number of the harvests exhibit low genetic variance, namely harvests 1,3,5 and 7. Two of these harvests (1 and 3) have 100% variance accounted for by the fa2 model, harvest 5 has 99% VAF, while harvest 7 has only 14% VAF. Of the remaining harvests, all which have higher genetic variance, harvests 2,4,6 and 10 have most of their variance explained by the first factor while the second factor contributes to most of the variance accounted for at harvests 8 and 9.

The genetic correlations between harvests and the genetic variances for each harvest from the fa2 model are presented in Table 6.5. As an aid to interpreting the genetic covariance structure between harvests and investigate any variety by harvest interaction the approach of Cullis et al. (2010) was followed. These authors use cluster analysis and heat map representation of the genetic correlation matrix after fitting the fa model in the aim of grouping the harvests into meaningful clusters that may be used for prediction and selection.

A cluster analysis, using the agglomerative (nested) hierarchical clustering algorithm in the agnes package in R (R Development Core Team, 2012), was performed using the average clustering method (Kaufman & Rousseeuw, 1990). The dendrogram of the REML estimates of the dissimilarity matrix $(\boldsymbol{I}_h - \boldsymbol{C}_2)$, where $\boldsymbol{C}_2$ is the REML estimate of the genetic correlation matrix based on the fa2 model), is presented in Figure 6.2.

Cullis et al. (2010) suggest that clusters formed above a cutoff of approximately 0.6 may not be meaningful. Hence the dendrogram (Figure 6.2) suggests possibly 2 main clusters with harvests 7 and 9 in groups of their own (making 4 clusters). The two main clusters include one cluster with most harvests having higher genetic variance (harvests $1, 2, 4, 6, 10$) and the second cluster consisting of harvests 3,8 and 5. These conclusions on the groupings of harvests are supported by the correlations in Table 6.5 and the heat map of correlations (Figure 6.3). From these correlations it can be seen that harvests $1, 2, 4, 6, 10$ are highly correlated with each other, harvests $3, 5, 8$ are highly correlated with each other and also negatively correlated with the establishment harvest 1, while there is little correlation between harvest 9 and the other harvests and also harvest 7 and the other harvests.

The groups obtained from the cluster analysis provide a starting point to identify potential target sets of harvests for which predictions of variety effects may be made. Table 6.6 presents a summary of key variables for each harvest including days between harvests, harvest mean yield and average daily yield (which may reflect rainfall and growth conditions). The table also includes the groupings from the cluster analysis for a dissimilarity level of 0.6 and a general summary of the groups. Cluster group 1 contains the winter harvest 6 (which has substantially higher genetic variance than the other harvests), the establishment harvest 1 (with little genetic variation) and the early harvests with genetic variance $(2, 4)$, plus harvest 10 which also has moderate genetic variance. This would seem reasonable as winter active varieties are known to perform better in early harvests than

their winter dormant counterparts. The varieties were also planted in winter so winter active varieties may establish better and perform better in their first year. The second cluster group is given by harvests $3, 8$ and 5. These harvests have low genetic variance. Harvest 9 is in it's own group possibly due to the very high rainfall occurring near this harvest, with the time between harvest 8 and 9 containing the highest recorded rainfall at nearby weather stations. Harvest 7 is also in a separate group.

It is useful to form selection indices for the main selection groups. Cullis et al. (2010) discuss the issues of conditional and marginal predictions (which may occur if some varieties are not measured at all harvests) and note that selection indices should be based on conditional predictions. As all lucerne varieties are observed at each harvest the predictions of variety effects are conditional predictions and it is suitable to form selection indices for each selection group. In these selection indices, equal weights are assigned to harvests in each group. If u_{ik} represents the genetic effect of variety i at harvest k and $\tilde{u}_{ik}$ is the BLUP of u_{ik}, the predicted selection index for variety i for the two main groups, $\hat{I}_i^{(d)}$, for $d = 1, 2$ may be defined as

$$\begin{aligned} \hat{I}_i^{(1)} &= (\{\tilde{u}_{i1}\} + \{\tilde{u}_{i2}\} + \{\tilde{u}_{i4}\} + \{\tilde{u}_{i6}\} + \{\tilde{u}_{i10}\})/5 \\ \hat{I}_i^{(2)} &= (\{\tilde{u}_{i3}\} + \{\tilde{u}_{i8}\} + \{\tilde{u}_{i5}\})/3 \end{aligned}$$

Figure 6.4 presents the predicted selection indices for each variety for each of the groupings. This figure can be used to see how varieties rank across the groups, for example it can be seen that variety 15 (TL2003%05) performs well with high rankings in groups 1,4 and 2 but is ranked lower in group 3.

Together with the breeder, different selection indices may be formed using different weightings of the harvests within the groups. In the end, how these selection indices are formed and used to make selection decisions for the release of new varieties, is a challenge faced by the breeder.

Table 6.4: REML estimates of rotated factor loadings and percentage variance accounted for (%VAF) from the fa2 model fitted to lucerne yield data (model Y17)

Harvest	$\mathbf{\Lambda}_1$	$\mathbf{\Lambda}_2$	$\mathbf{\Psi}$	%VAF
1	0.028	0.039	0.000	100.000
2	0.125	0.048	0.000	100.000
3	0.019	−0.040	0.000	100.000
4	0.177	−0.016	0.000	100.000
5	0.091	−0.088	0.000	98.984
6	0.545	0.006	0.025	92.271
7	0.027	−0.009	0.005	13.644
8	0.038	−0.120	0.000	100.000
9	−0.012	−0.067	0.015	23.541
10	0.137	0.037	0.005	80.461

The REML estimates of the residual variance parameters in the final model (Y17),

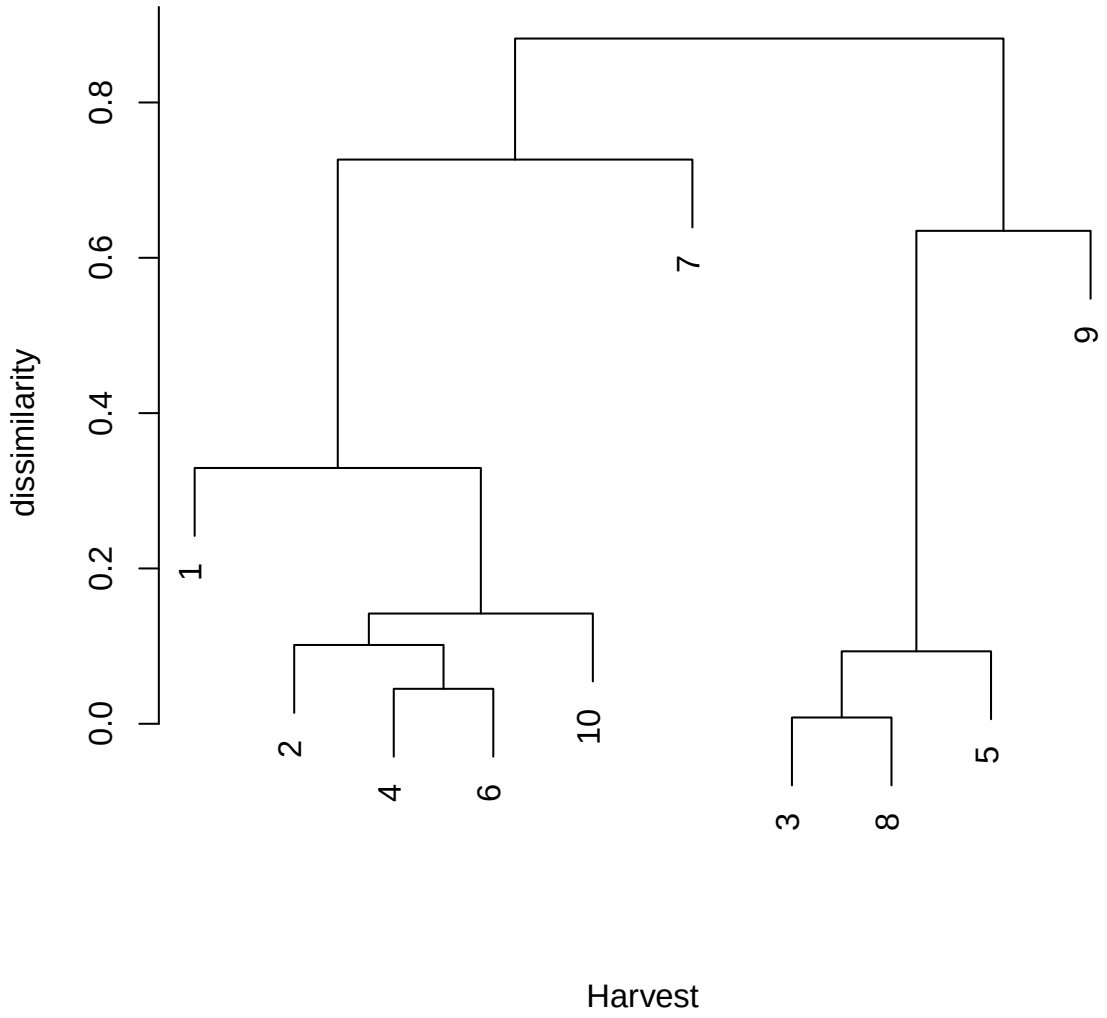
Table 6.5: Lower triangle of the REML estimates of $\boldsymbol{G}_g$, with genetic variances $\times 100$ (on diagonal) and genetic correlations (off diagonals) from fa2 model (Y17) fitted to lucerne yield data

Harvest	1	2	3	4	5	6	7	8	9	10
1	0.229									
2	0.843	1.787								
3	−0.476	0.072	0.191							
4	0.513	0.894	0.511	3.145						
5	−0.133	0.419	0.930	0.778	1.614					
6	0.576	0.899	0.401	0.955	0.681	32.188				
7	0.116	0.286	0.253	0.361	0.330	0.336	0.577			
8	−0.589	−0.061	0.991	0.392	0.875	0.282	0.215	1.594		
9	−0.437	−0.256	0.393	−0.044	0.265	−0.090	0.023	0.427	1.998	
10	0.700	0.892	0.160	0.841	0.460	0.834	0.278	0.042	−0.188	2.490

Table 6.6: Table showing harvest variables including mean yield (yld), days between harvests (dd), average daily yield (ady) and groupings based on the cluster analysis of lucerne yield data

Harvest	Harvest date	yld	dd	ady	group	comments
1	13/11/03	1934	114	16.96	1	establishment and early harvests
2	16/12/03	1314	33	39.83	1	with higher genetic variance,
4	11/03/04	808	35	23.10	1	includes winter harvest
6	22/06/04	624	63	9.91	1	
10	24/01/05	977	32	30.54	1	
3	5/02/04	1948	51	38.19	2	harvests with little genetic variance
8	2/12/04	1012	51	19.86	2	all negatively correlated with harvest 1
5	20/04/04	779	40	19.48	2	
7	12/10/04	1257	112	11.22	3	low average daily yield, long growth period, includes some winter and spring growth
9	23/12/04	1229	21	58.52	4	second year harvest with high rainfall high average daily yield negatively correlated with first two harvests

Figure 6.2: Dendrogram of the dissimilarity matrix from fa2 model (Y17) fitted to the lucerne yield data

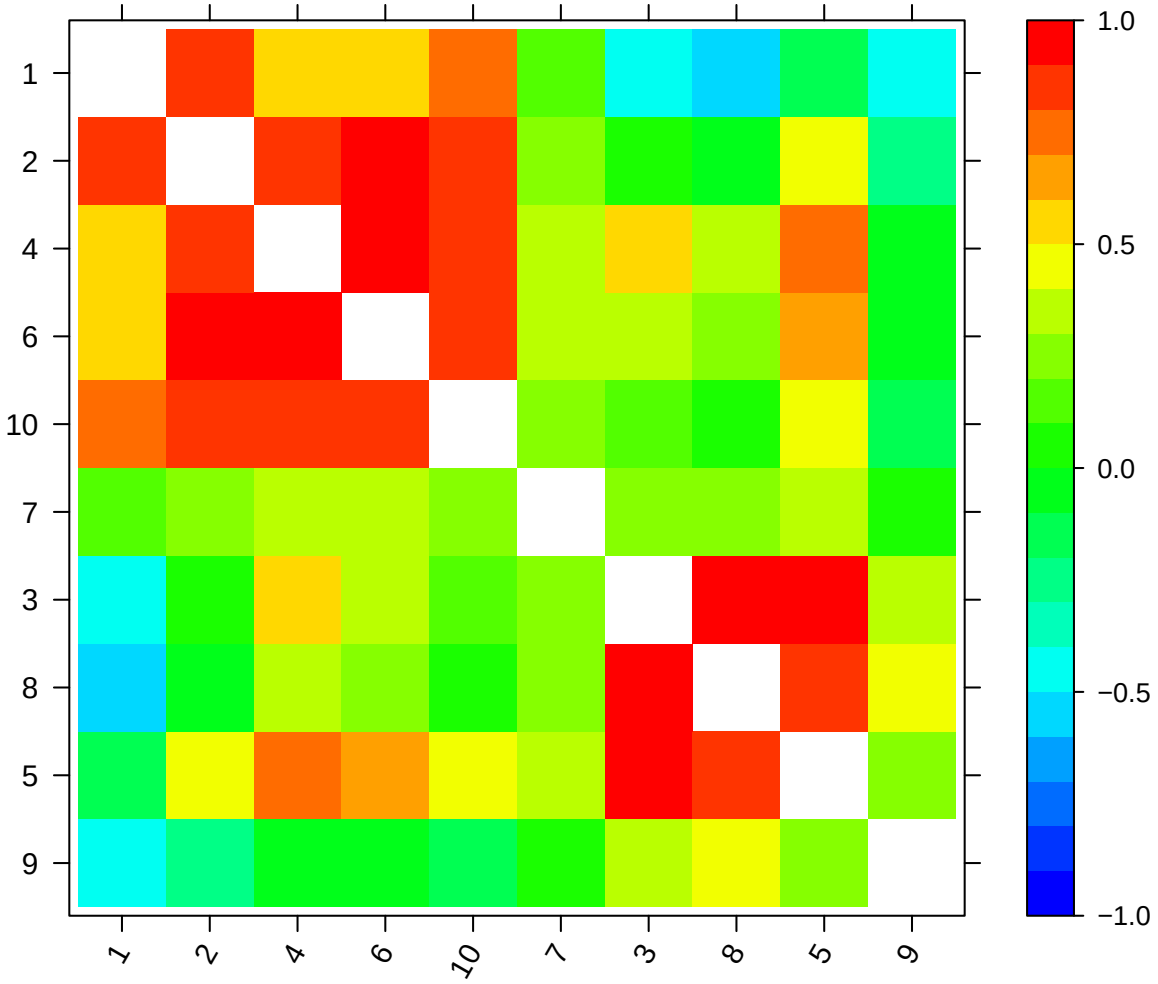


where $\boldsymbol{R}$ is given by

$$\boldsymbol{R} = \sigma_p^2 \boldsymbol{J}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}_r^{(p)} + \boldsymbol{\Sigma}_h^{ante2} \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$$

are given by $\hat{\sigma}_p^2 = 0.073$, $\boldsymbol{\Sigma}_r^{(p)}$ is an ar1 correlation matrix with correlation parameter $\hat{\phi}_r^{(p)} = 0.522$, $\boldsymbol{\Sigma}_c$ is an ar1 correlation matrix with correlation parameter $\hat{\phi}_c = 0.089$, $\boldsymbol{\Sigma}_r$ is an ar1 correlation matrix with correlation parameter $\hat{\phi}_r = 0.170$, and $\boldsymbol{\Sigma}_h^{ante2}$ is an antedependence (of order 2) covariance matrix with parameters given in terms of the

Figure 6.3: Heat map representation of the genetic correlation matrix from the fa2 model (Y17) fitted to the lucerne yield data. Harvests are ordered according to the dendrogram.



inverse variance matrix $\boldsymbol{\Sigma}_{ante2}^{-1} = \boldsymbol{U}^T \boldsymbol{D} \boldsymbol{U}$ where U is a lower triangular matrix given by

$$\boldsymbol{U} = \begin{bmatrix}
1 & & & & & & & & & \\
-0.337 & 1 & & & & & & & & \\
-0.036 & -0.331 & 1 & & & & & & & \\
0 & -0.092 & 0.013 & 1 & & & & & & \\
0 & 0 & -0.016 & -0.196 & 1 & & & & & \\
0 & 0 & 0 & -0.486 & -0.311 & 1 & & & & \\
0 & 0 & 0 & 0 & -0.213 & -0.187 & 1 & & & \\
0 & 0 & 0 & 0 & 0 & -0.188 & -0.456 & 1 & & \\
0 & 0 & 0 & 0 & 0 & 0 & -0.241 & -0.076 & 1 & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.288 & 0.211 & 1
\end{bmatrix}$$

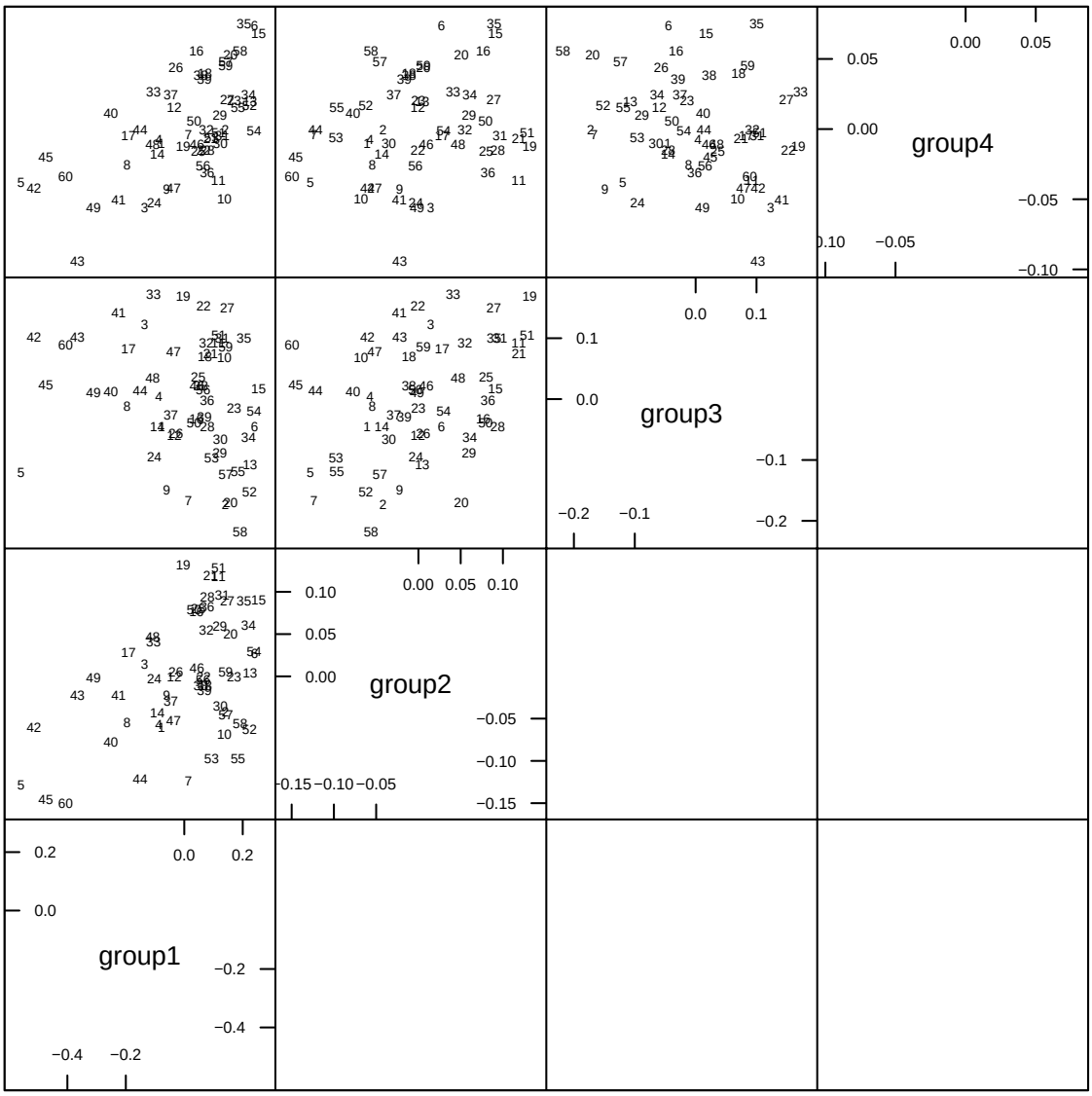
and $\boldsymbol{D}$ is a diagonal matrix with elements
$[6.319, 6.466, 13.141, 13.330, 21.982, 11.117, 14.843, 13.598, 15.411, 6.748]$.

Multiplying these matrices and taking the inverse, we can obtain the antedependence covariance matrix $\boldsymbol{\Sigma} = (\boldsymbol{U^T D U})^{-1}$

$$\boldsymbol{\Sigma} = (\boldsymbol{U^T D U})^{-1}) = \begin{bmatrix} 0.16 & 0.05 & 0.02 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.05 & 0.17 & 0.06 & 0.01 & 0.00 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.02 & 0.06 & 0.10 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.00 & 0.08 & 0.01 & 0.04 & 0.01 & 0.01 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.01 & 0.05 & 0.02 & 0.01 & 0.01 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.00 & 0.04 & 0.02 & 0.12 & 0.03 & 0.03 & 0.01 & 0.01 \\ 0.00 & 0.00 & 0.00 & 0.01 & 0.01 & 0.03 & 0.07 & 0.04 & 0.02 & 0.01 \\ 0.00 & 0.00 & 0.00 & 0.01 & 0.01 & 0.03 & 0.04 & 0.10 & 0.02 & 0.02 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.01 & 0.02 & 0.02 & 0.07 & -0.01 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.01 & 0.01 & 0.02 & -0.01 & 0.16 \end{bmatrix}$$

The residual structure consisted of two main components, namely a Plot effect (which was spatially correlated in the row direction) and a decaying serial dependence component (modelled by a separable antedependence of order two temporal component by an autoregressive of order 1 spatial model in the Column and Row directions($\mathsf{ar1(Column)} \otimes \mathsf{ar1(Row)}$)). The Plot variance (0.073) was reasonably large and of similar magnitude to the variances in the antedependence covariance matrix (diagonals in the matrix $\boldsymbol{U^T D U}^{-1}$ given above which models the serial dependence. The spatial correlation parameter in the Row direction on the overall plot term was high (0.522) while the spatial correlation parameters at the harvest level were lower (Col = 0.078 and Row = 0.172).

Figure 6.4: Pairwise plots of the predicted selection indices for each of the four groups described in the text. Numbers refer to variety names in Table 6.7



6.2.2 Persistence

While in some circumstances it may be sufficient and desirable to model the differences between varieties or variety contrasts over time (as in the previous section for yield), in other instances it may be important to model the actual underlying response over time. For example, with the lucerne persistence data the aim is to predict the time to a certain level of persistence. This prediction cannot be obtained from an analysis based on deviations from the harvest means and requires modelling of the persistence response over time.

In this section the analysis of the lucerne persistence data from Terry Hie Hie, where interest lies in predicting the time until the persistence level of each variety drops to 30%, is presented. The persistence data (based on percentages) from Terry Hie Hie was

Table 6.7: Table of variety names and corresponding numbers for lucerne data

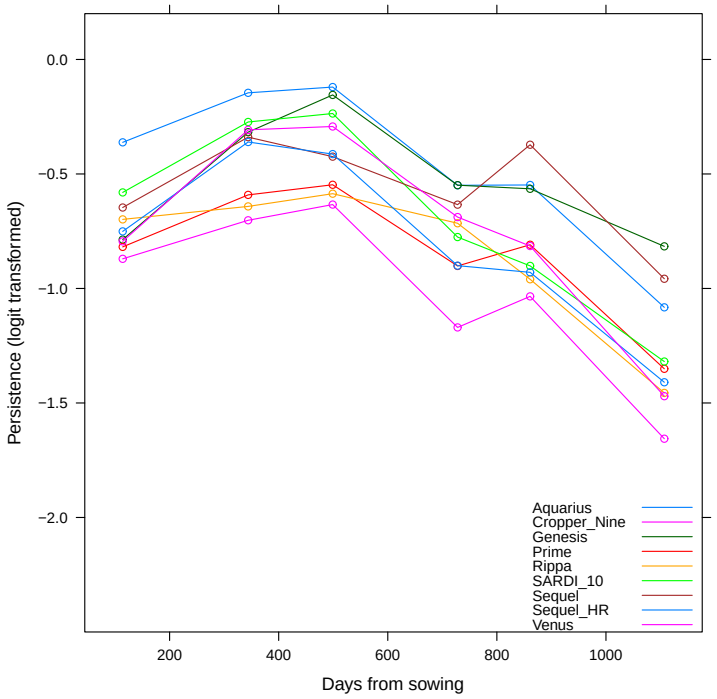
Variety No	Variety Name	Variety No	Variety Name
1	Aquarius	31	TL2003%21
2	Cropper_Nine	32	TL2003%22
3	Genesis	33	TL2003%23
4	Pioneer_L90	34	TL2003%24
5	Prime	35	TL2003%25
6	Rippa	36	TL2003%26
7	SARDI_10	37	TL2003%27
8	Sceptre	38	TL2003%28
9	Sequel	39	TL2003%29
10	Sequel_HR	40	TL2003%30
11	TL2003%01	41	TL2003%31
12	TL2003%02	42	TL2003%32
13	TL2003%03	43	TL2003%33
14	TL2003%04	44	TL2003%34
15	TL2003%05	45	TL2003%35
16	TL2003%06	46	TL2003%36
17	TL2003%07	47	TL2003%37
18	TL2003%08	48	TL2003%38
19	TL2003%09	49	TL2003%39
20	TL2003%10	50	TL2003%40
21	TL2003%11	51	TL2003%52
22	TL2003%12	52	TL2003%53
23	TL2003%13	53	TL2003%54
24	TL2003%14	54	TL2003%55
25	TL2003%15	55	TL2003%56
26	TL2003%16	56	TL2003%57
27	TL2003%17	57	TL2003%58
28	TL2003%18	58	TL2003%59
29	TL2003%19	59	TL2003%60
30	TL2003%20	60	Venus

transformed using a logit transformation prior to analysis (as detailed in section 3.3.2). A plot of the transformed persistence data for the 11 commercial varieties is presented in Figure 6.5. From this plot it is seen that the varieties follow a similar trend over time.

One approach in analysing such data is to model the underlying overall trend over time using a smooth curve (e.g. polynomial or cubic smoothing spline) and then investigate the deviations from this underlying smooth trend for each variety. These variety deviations may be linear and hence a linear random regression approach may be applied. Alternatively more complex random regression models including nonlinear functions or splines at the individual variety level may be used to model the deviations from the overall smooth curve (Verbyla et al., 1999).

Deciding on the most appropriate model to fit to the persistence data involved a complex model building process. A sequential approach similar to that of the yield analysis (in the previous section) was followed. The process commenced with a simple genetic

Figure 6.5: Plot of mean lucerne persistence data (on transformed scale) for each of the 11 commercial varieties at Terry Hie Hie



model in which harvests were treated separately in order to find suitable residual models for modelling the spatial correlation at each harvest. Then the harvests were combined and a residual model incorporating the temporal correlation between harvests was found. At this stage the genetic model was also modified to model the data across the harvests (using a cubic smoothing spline) and in subsequent models the residual model was further refined. Once a final residual model was selected then more complex genetic models were investigated. The presentation of the process is separated into two sections, one focussing on modelling the non-genetic (residual) effects and the second concentrating on modelling the genetic effects. The sequence of residual models fitted is given in Table 6.8, and the genetic models given in Table 6.10.

Modelling non-genetic effects

The analysis commenced by treating the 6 harvest (assessment) times separately, firstly by fitting a RCB model to each harvest (model TP1). In this model and models TP2-TP5 (in Table 6.8) the simple genetic model of (6.2.1) was implemented and for these models an intercept term was included for each harvest. The next step was to account for any local spatial correlation between plots at each harvest. Model TP2 specified a separate spatial model for each harvest using an ar1(Column).ar1(Row) process for the residual structure. This was a significant improvement on TP1 (REMLRS = 38.90 on 12df, P<0.001). Further spatial terms were investigated and included in the model following the approach of Gilmour et al. (1997). The final spatial model included a random

column term for each harvest plus spatial correlation in the row and column directions (ar1(Column).ar1(Row)). This is model TP3.

The residuals from model TP3 were investigated and harvest variances and correlations between harvests calculated (Table 6.9). There are positive correlations between harvests which decay as the time between measurements increases. These correlations do not decay to zero and suggest a plausible model of an overall average plot effect and an autoregressive ar1 process to model the temporal correlation between the harvest measurements (similar to the residual model assumed for the yield data). Spatial correlation between plots can once again be incorporated at the overall plot level and also at the Plot by Harvest level.

As in the yield analysis, the residual spatio-temporal model of the 3 way separable ar1v(Harvest).id(Column).ar1(Row) process for the plot by harvest effects (TP4) provided a much better fit to the data than the separate spatial models for each harvest, based on AIC values. Allowing for heterogeneous residual variances (TP5) provided an improvement on model TP4 where the variances were assumed equal. Note that in all models up to this point a simple diagonal genetic model was fitted which assumes no genetic correlation between the harvests. At this point a genetic model to connect the data across the harvests was implemented (TP6). This model fitted an overall mean cubic smoothing spline (1 + lin(years) + spl(years) + dev(years), where years refers to time (in years) from planting, lin(years) denotes the linear effect of time, spl(years) denotes the random smooth spline component and dev(years) represents non-smooth deviations from the overall smooth spline (Verbyla et al., 1999)) and a term for the variety deviations from the overall spline (idv(Harvest.Variety)). Using this spline genetic model further residual models were investigated.

The next model to be fitted (TP7) incorporated an overall plot effect and allowed for spatial correlation on this effect. In this model the spatial correlation parameter was constrained to be the same at both the overall plot level and the plot by harvest level. This model was an improvement on model TP6 based on an adjusted REMLRT statistic (Stram & Lee, 1994) of 82.124 ($P<0.001$). Next the fully separable residual model of 5.3.5 which provides a spatial extension to the repeated measures model of Diggle (1988) was fitted (TP8). This model was a significant improvement on model TP7 based on an adjusted REMLRT statistic (Stram & Lee, 1994) of 14.4 ($P<0.001$)). This was followed by a model which allowed for an independent measurement error term but still constrained the spatial correlation parameters at the plot and plot by harvest levels to be the same (TP9). This model had the same number of parameters as model TP8 and a similar log-likelihood. Model TP10 fitted an unconstrained model incorporating the overall plot effect and measurement error terms. In this model the spatial correlation parameter in the column direction at the overall plot level was dropped (to achieve convergence). This model was a significant improvement on model TP9 (REMLRT = 9.812 on 1 df, $P=0.002$). Model TP11 further dropped the spatial correlation parameter in the column direction at the plot by harvest level. This model resulted in one less parameter and did not differ significantly to model TP10 (REMLRT=0.54 on 1df, $P=0.5$).

Different models were investigated for the temporal correlation part of the model as

an alternative to the ar1h process; for example TP12 fitted an exponential model (exph) while TP13 fitted an ante-dependence (ante) model. Neither of these models proved an improvement on TP11. Hence the final best residual model was selected as TP11 and this residual model was used for investigating subsequent more complex genetic models as detailed in the following section.

Modelling genetic effects

Table 6.10 presents the genetic models fitted to the persistence response over time. The first model listed is TP14. This model included an overall mean spline (1 + lin(years) + spl(years) +dev(years)) and a diagonal variance model for the variety deviations from this overall spline (diag(Harvest).Variety). The residual variance model was given by $\sigma_p^2 \boldsymbol{J}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}_r^{(p)} + \boldsymbol{D}\boldsymbol{\Sigma}_h\boldsymbol{D} \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}_r + \sigma_m^2 \boldsymbol{I}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{I}_r$. This residual model is common for all models fitted in Table 6.10.

The subsequent models (TP15-18) specified different structures for the variety deviations from the overall underlying spline. Model TP15 specified a factor analytic structure of order 1 to the variety deviations (fa1) in a manner similar to that in the analysis of yield. Model TP16 specified a fa2 model which provided an improvement on the single factor model. Table 6.11 presents the estimated genetic variances for each harvest and genetic correlations between harvests based on the fa2 model. The genetic correlations were very high for successive times (ranging from 0.836 to 0.940) with the highest genetic correlation for persistence occurring between harvest times 4 and 5 (approximately two, and two and a half years after sowing). It is interesting to note that the genetic correlation between the first and final harvest times (six months and three years after sowing) was negative (-0.280), which may indicate a tendency for some varieties that performed well early in the trial to not show as high persistence later.

Subsequent models followed the approach of Verbyla et al. (1999), that is by modelling the overall mean profile over time using a cubic smoothing spline and then random regressions for the variety deviations from this overall mean spline. Using notation based on that of Verbyla et al. (1999) this model can be represented by 1 + lin(years) + spl(years) + dev(years) + Variety + Variety.lin(years) + Variety.spl(years) + Variety.dev(years) where 1 + lin(years) + spl(years) denotes the overall mean spline and Variety, Variety.lin(years) and Variety.spl(years) denote the variety intercepts, slopes and random spline components for the variety deviations from the overall mean. The terms dev(years) and Variety.dev(years) denote lack of fit terms at the mean and Variety by time levels.

Model TP17 fitted a linear random regression over time for the variety deviations. This linear random regression model correlated the intercepts and slopes for varieties. The next model TP18 introduced a spline random regression model for the variety deviations (with correlation between the intercepts and slopes but not correlating these terms with the spline terms). The spline component was statistically significant (REMLRT = 3.671 on 1 df, $P = 0.028$) but the predicted variance component was very small. Attempts were made to correlate the random intercepts, slopes and spline terms in this model (to make the model invariant to a change of basis, as detailed in section 5.3.3) but were not successful. Therefore, because these correlations could not be estimated, and the fact that the spline component was so small relative to the other variance components, it was decided to take the final model as TP17. Table 6.12 presents the predicted variance components for models TP17 and TP18. It can be seen that the predicted variance components for the variety random regression intercept and slope components were approximately 50 and 10 times the variety spline component, respectively. Note that the residual models in Table

Table 6.8: Summary of residual models fitted for persistence at Terry Hie Hie. In all models the simple genetic effects model of 6.2.1 has been fitted and a random Replicate effect has been included for each harvest. Models TP3-TP13 include global and extraneous spatial effects. In models TP1-TP5 a separate mean for each harvest has been fitted while in models TP6-TP13 the overall mean has been modelled over time using the spline model: 1 + lin(years) + spl(years) + dev(years). Residual log-likelihoods (denoted by ℓ), number of parameters in $\boldsymbol{R}$ (npar), number of other non-genetic variance parameters (Other npar) and AIC values (given as differences from the best model) are presented. Groups of models above and below the horizontal line have different fixed effects and are not comparable using residual likelihood ratio tests or AIC values. Model terms are detailed in Table 5.1. Superscripts on ar1 parameters $ar1^1$ denotes that within a model the ar1 spatial parameters with the same superscript are constrained to be equal.

Model	Residual model $\boldsymbol{R}$		ℓ	$\boldsymbol{R}$ npar	Other npar	AIC
TP1	$\text{diag}(\sigma^2_k \boldsymbol{I}_c \otimes \boldsymbol{I}_r)$	at(Harv).id(Col).id(Row)	673.848	6	6	1071.2
TP2	$\text{diag}(\sigma^2_k \boldsymbol{\Sigma}_{kc} \otimes \boldsymbol{\Sigma}_{kr})$	at(Harv).ar1(Col).ar1(Row)	693.421	18	6	1056.0
TP3	$\text{diag}(\sigma^2_k \boldsymbol{\Sigma}_{kc} \otimes \boldsymbol{\Sigma}_{kr})$	at(Harv).ar1(Col).ar1(Row)	711.817	18	12	1031.3
TP4	$\sigma^2 \boldsymbol{\Sigma}_h \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$	ar1v(Harv).ar1(Col).ar1(Row)	1161.726	4	12	103.4
TP5	$\boldsymbol{D}\boldsymbol{\Sigma}_h\boldsymbol{D} \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$	ar1h(Harv).ar1(Col).ar1(Row)	1168.007	9	12	100.9
TP6	$\boldsymbol{D}\boldsymbol{\Sigma}_h\boldsymbol{D} \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$ ar1h(Harv).ar1(Col).ar1(Row)		1170.350	9	14	100.2
TP7	$(\sigma^2_p \boldsymbol{J}_h + \boldsymbol{D}\boldsymbol{\Sigma}_h\boldsymbol{D}) \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$ ar1h(Harv).$ar1^1$(Col).$ar1^2$(Row) + $ar1^1$(Col).ar1v^2(Row)		1211.412	10	14	20.1
TP8	$(\sigma^2_p \boldsymbol{J}_h + \boldsymbol{D}\boldsymbol{\Sigma}_h\boldsymbol{D} + \sigma^2_m \boldsymbol{I}_h) \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$ ar1h(Harv).$ar1^1$(Col).$ar1^2$(Row) + $ar1^1$(Col).ar1v^2(Row) + Harv.$ar1^1$(Col).$ar1^2$(Row)		1218.612	11	14	7.7
TP9	$(\sigma^2_p \boldsymbol{J}_h + \boldsymbol{D}\boldsymbol{\Sigma}_h\boldsymbol{D}) \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + \sigma^2_m \boldsymbol{I}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{I}_r$ ar1h(Harv).$ar1^1$(Col).$ar1^2$(Row) + $ar1^1$(Col).ar1v^2(Row) + Harv.Col.Row		1217.762	11	14	9.4
TP10	$\sigma^2_p \boldsymbol{J}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}^{(p)}_r + \boldsymbol{D}\boldsymbol{\Sigma}_h\boldsymbol{D} \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + \sigma^2_m \boldsymbol{I}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{I}_r$ ar1h(Harv).ar1(Col).ar1(Row) + id(Col).ar1v(Row) + Harv.Col.Row		1222.672	12	14	1.5
TP11	$\sigma^2_p \boldsymbol{J}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}^{(p)}_r + \boldsymbol{D}\boldsymbol{\Sigma}_h\boldsymbol{D} \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}_r + \sigma^2_m \boldsymbol{I}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{I}_r$ ar1h(Harv).id(Col).ar1(Row) + id(Col).ar1v(Row) + Harv.Col.Row		1222.445	11	14	0
TP12	$\sigma^2_p \boldsymbol{J}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}^{(p)}_r + \boldsymbol{D}\boldsymbol{\Sigma}^{exp}_h\boldsymbol{D} \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + \sigma^2_m \boldsymbol{I}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{I}_r$ exph(Harv).ar1(Col).ar1(Row) + id(Col).ar1v(Row) + Harv.Col.Row		1222.267	12	14	2.4
TP13	$\sigma^2_p \boldsymbol{J}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}^{(p)}_r + \boldsymbol{\Sigma}^{ante}_h \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + \sigma^2_m \boldsymbol{I}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{I}_r$ ante(Harv).ar1(Col).ar1(Row) + id(Col).ar1v(Row) + Harv.Col.Row		1223.968	16	14	7.0

Table 6.9: Empirical/sample variances (on diagonal) and correlations (off diagonals) of residuals from model TP3

	1	2	3	4	5	6
1	0.070					
2	0.816	0.064				
3	0.787	0.792	0.070			
4	0.726	0.767	0.794	0.079		
5	0.711	0.756	0.768	0.800	0.078	
6	0.666	0.720	0.708	0.755	0.760	0.059

6.8 were re-fitted with the final genetic model (TP17). This process confirmed that the residual model chosen initially was still the most suitable.

A plot of the predicted splines for each variety from the final model (TP17) is presented in Figure 6.6. This plot shows when each variety reaches 30% persistence as shown by the horizontal line in each plot (at -0.838 on the logit scale). It is evident from this plot that most of the variation is in the overall mean and the curvature is very similar for all varieties. However, there is variation in the starting points of the curves and there is variation between varieties in the rate at which the persistence declines. This variation is more clearly seen in Figure 6.7 which presents the linear deviations from the overall mean spline for each variety together with 95% prediction intervals. It can be seen that varieties such as Genesis, TL2003%04, TL2003%12 have higher positive slopes than most varieties and also medium to high intercepts indicating both good establishment and good longer term persistence while some varieties such as TL2003%53 and TL2003%56 have negative slopes indicating poorer persistence. Variety TL2003%34 has a high positive slope but very low intercept and clearly did not have good persistence or coverage at the start of the trial.

Predicting time to P%

To address the problem of predicting the time taken till the persistence level for each variety declines to $P\%$, the final model TP17 was used. The procedure firstly involved predicting the persistence value for each variety at each of the harvest times (knot points of the splines). This involved prediction for the overall spline component plus the individual linear random regression component for each variety. Denote the overall spline by $f(t)$ and the linear random regression for variety i by $l_i(t) = u_{i0} + u_{i1}t$ where u_{i0} and u_{i1} are the random intercept and slope for variety i. Hence the persistence for variety i at harvest time t_k may be predicted by $f(t_k) + l_i(t_k)$. The next step involved identifying the interval $[t_L, t_R]$ where each variety crosses the $P\%$ level. That is, where $f(t_L) + l_i(t_L) - P > 0$ and $f(t_R) + l_i(t_R) - P < 0$. Once the interval was obtained for each variety, the root of the equation $f(t) + l_i(t) - P = 0$ was required for each variety i. This was achieved using the expression for the cubic smoothing spline $f(t)$ presented in (1.3.8) (Green & Silverman,

Table 6.10: Summary of genetic variance models fitted to transformed persistence at Terry Hie Hie. Residual log-likelihoods (denoted by ℓ), number of parameters in $\boldsymbol{G}_g$ (npar) and AIC values (given as differences from the best model) are presented for each model.

			$\boldsymbol{G}_g$	
Model	Genetic model	ℓ	npar	AIC
TP14	diag(Harvest).Variety	1225.536	6	19.2
TP15	fa(Harvest,1).Variety	1237.473	16	15.3
TP16	fa(Harvest,2).Variety	1241.603	21	17.0
TP17	corh(1,years).id(Variety) + idv(Harvest.Variety)	1237.124	8	0
TP18	corh(1,years).id(Variety)+ spl(years).id(Variety) + idv(Harvest.Variety)	1238.959	9	

All models have residual variance model of TP11 in Table 6.8 ($\boldsymbol{R} = \sigma_p^2 \boldsymbol{J}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}_r^{(p)} + \boldsymbol{D}\boldsymbol{\Sigma}_h\boldsymbol{D} \otimes \boldsymbol{I}_c \otimes \boldsymbol{\Sigma}_r + \sigma_m^2 \boldsymbol{I}_h \otimes \boldsymbol{I}_c \otimes \boldsymbol{I}_r$)
and have the overall mean over time modelled using the spline model : 1 + lin(years) + spl(years) + dev(years).

Table 6.11: Genetic variances (×100) at each harvest (on diagonal) and genetic correlations between harvests (off diagonals) estimated from the fa2 model (TP16) fitted to the lucerne persistence data

	1	2	3	4	5	6
1	0.741					
2	0.842	0.805				
3	0.741	0.864	1.065			
4	0.343	0.625	0.885	1.324		
5	0.077	0.413	0.710	0.940	0.923	
6	-0.280	0.079	0.384	0.732	0.836	0.799

Figure 6.6: Plot showing predicted splines for each variety from model TP17 (in Table 6.10) for logit transformed persistence data over time for Terry Hie Hie

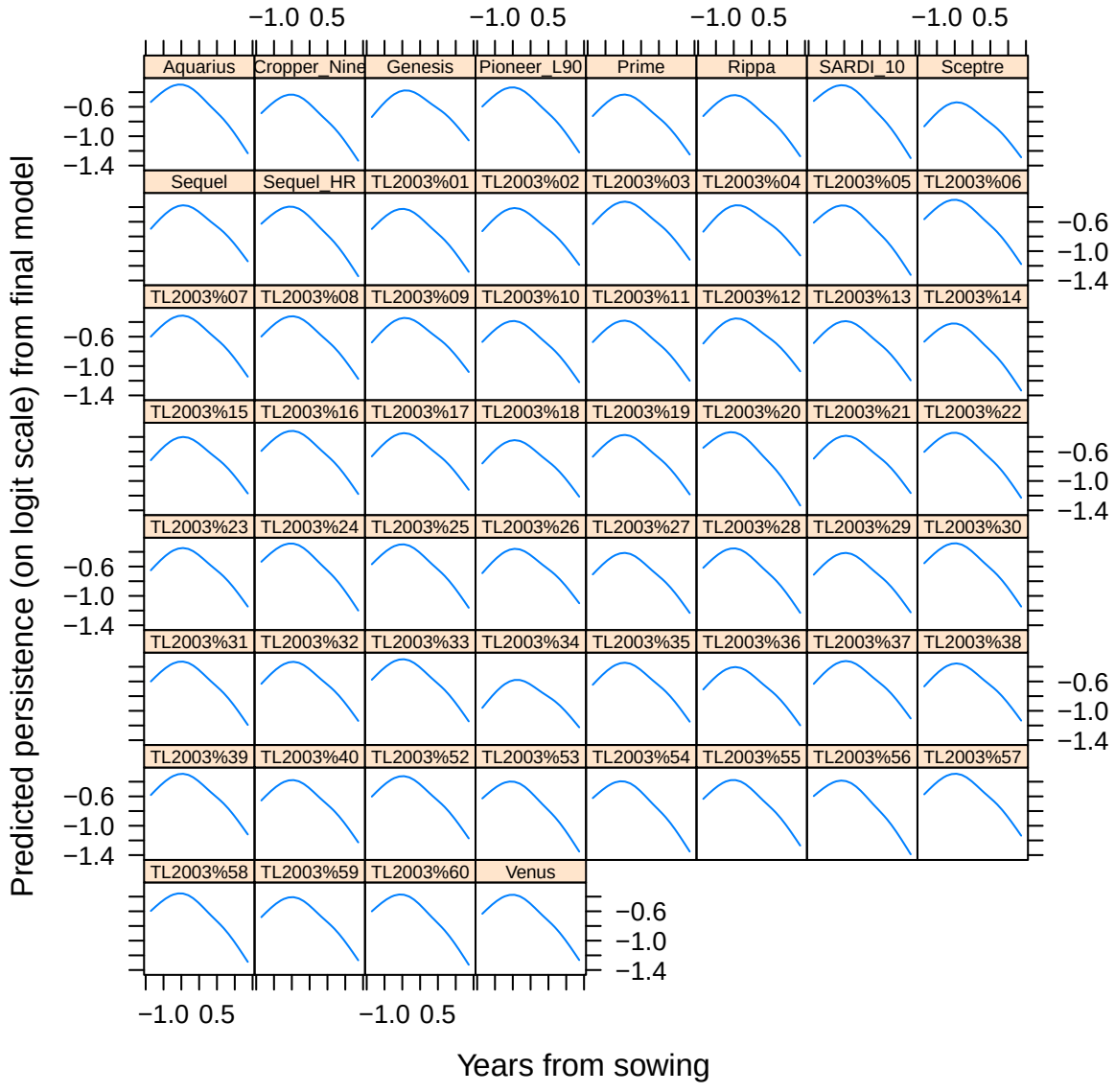


Figure 6.7: Plot showing predicted variety linear deviations (and 95% prediction intervals) from overall mean spline from model TP17 (in Table 6.10) for logit transformed persistence data over time for Terry Hie Hie

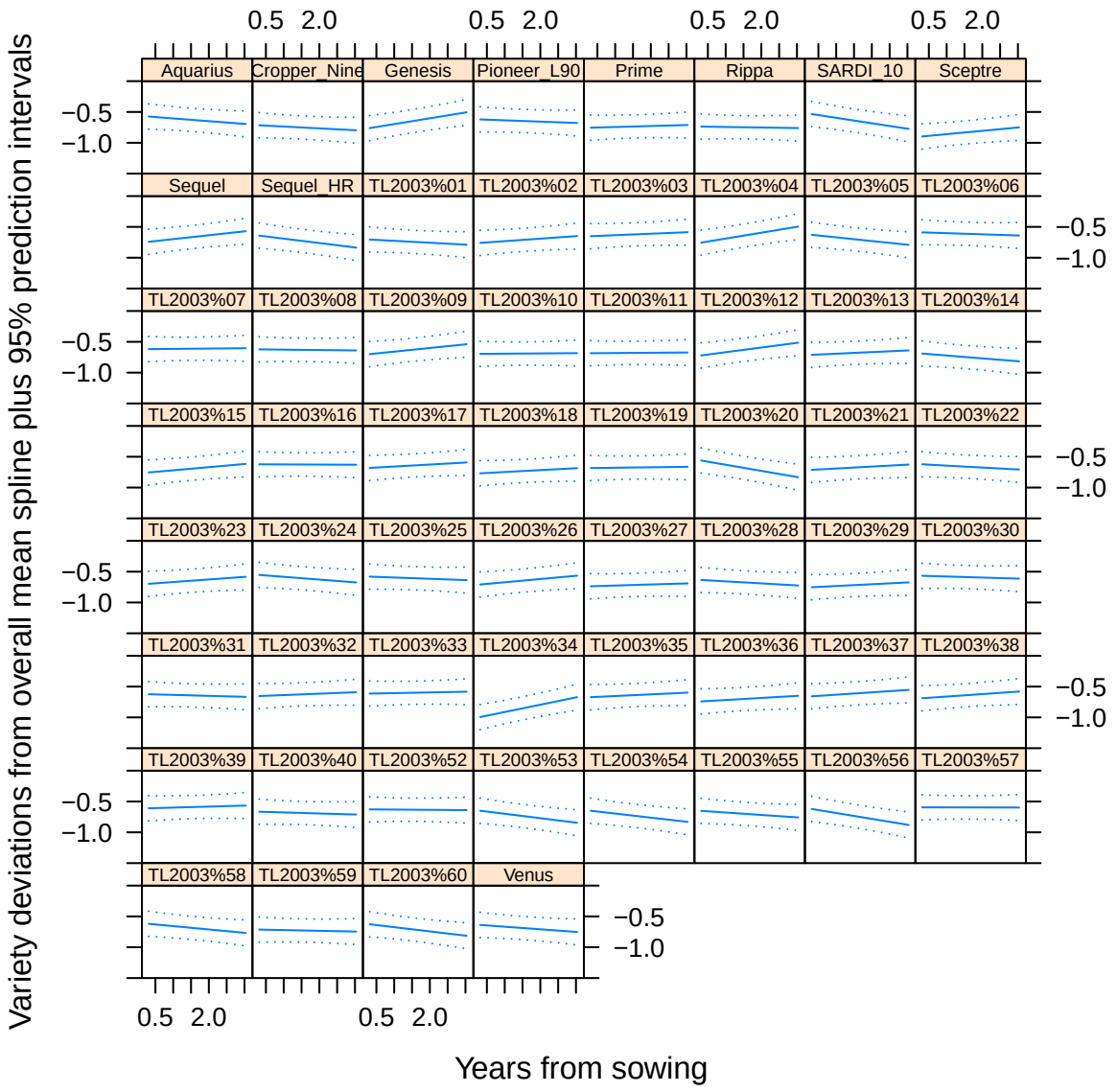


Table 6.12: REML estimates of the variance components ($\times 10^3$ from the random regression models (TP17, TP18) fitted to lucerne persistence data

Term	Var comp TP17	Var comp TP18
spl(years)	10.465	10.454
dev(years)	4.806	4.784
Variety	8.961	9.268
Variety.lin(years)	2.842	2.990
Variety.spl(years)	-	0.204
Variety.dev(years)	0.690	0.000

1994). That is

$$\begin{aligned} f(t) &= \frac{(t-t_L)f_R + (t_R-t)f_L}{h} \\ &- \frac{1}{6}(t-t_L)(t_R-t)[(1+\frac{t-t_L}{h})\gamma_R + (1+\frac{t_R-t}{h})\gamma_L] \end{aligned}$$

where γ_R and γ_L are the second derivatives of the cubic smoothing spline at knot points t_R and t_L respectively, $h = t_R - t_L$, $f_L = f(t_L)$ and $f_R = f(t_R)$ (see Chapter 1 for details).

A number of methods are available to solve the equation, $f(t)+l_i(t)-P = 0$, including the bi-section or Newton Rhapson method.

The results for the predicted times for each variety to decline to 30% (-0.838 on logit scale) were obtained using the R function uniroot to solve the above equation. The results are presented in Figure 6.8, where it can be seen that Genesis took the longest time to decline to 30% persistence. A plot of the predicted splines plus 95% prediction intervals for the top 5 and bottom 5 varieties is given in Figure 6.9. While there is almost a six month difference in predicted time to 30% persistence between the top and bottom varieties, the prediction intervals are quite large and many varieties may not be statistically different.

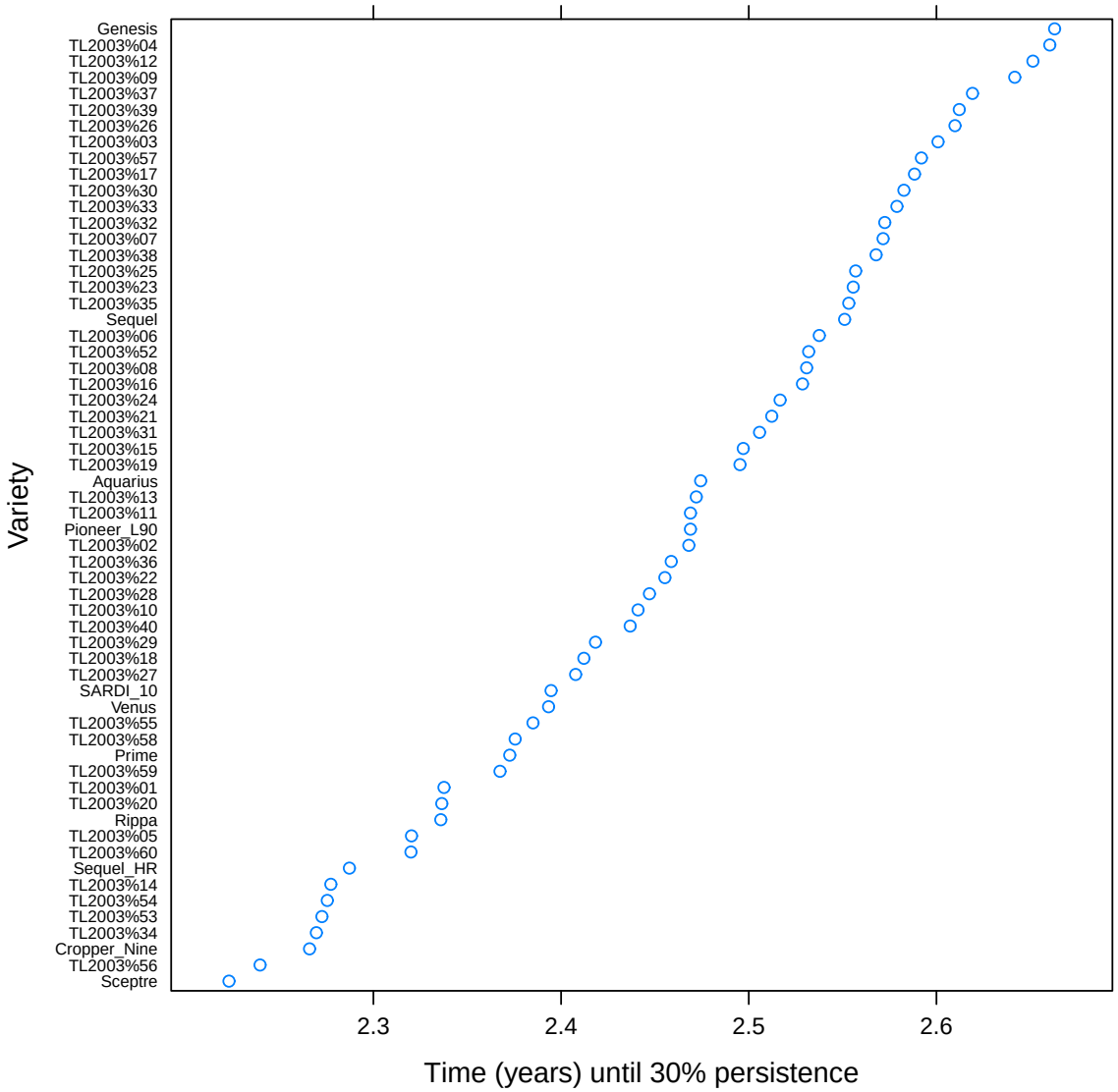
The R code for fitting the model and subsequently predicting the time till 30% persistence is presented in Appendix C.

6.3 Discussion

The methods presented in this chapter provide a new approach for the analysis of multiple harvest variety selection data from a single site, that accounts for both spatial variation between plots and temporal correlation between harvests, and allows the genetic effects to be modelled over time. In both data sets analysed there was substantial spatial and temporal correlation. The three components of temporal correlation identified in Diggle (1988) were found across the two analyses. There were large plot effects and serial correlation in both analyses and significant measurement error in the analysis of the persistence data.

The combined approach of fitting the new extended spatial and temporal residual models (developed in Chapter 5) plus the different models for genetic effects is a new approach that builds on the models of Smith et al. (2007). At the residual level, it has been shown that the models presented provide an approach that is significantly better in fit than assuming independence between harvest times. However the three way separable (harvest by column by row) structure assumed for the spatio-temporal correlation may not always be appropriate (Smith et al., 2007). The model assumes common spatial parameters over harvest times which has been shown in Chapter 3 not to always be the case. It may be expected that the spatial variation between harvest times may differ due to seasonal changes and growth phase of the crop. This may vary between trials. For example in the yield analysis at Terry Hie Hie the spatial correlation parameters did not differ greatly in the individual analyses (see Chapter 3) and this is also shown in the comparison between models Y1 and Y2 (Table 6.1) where separate spatial models are

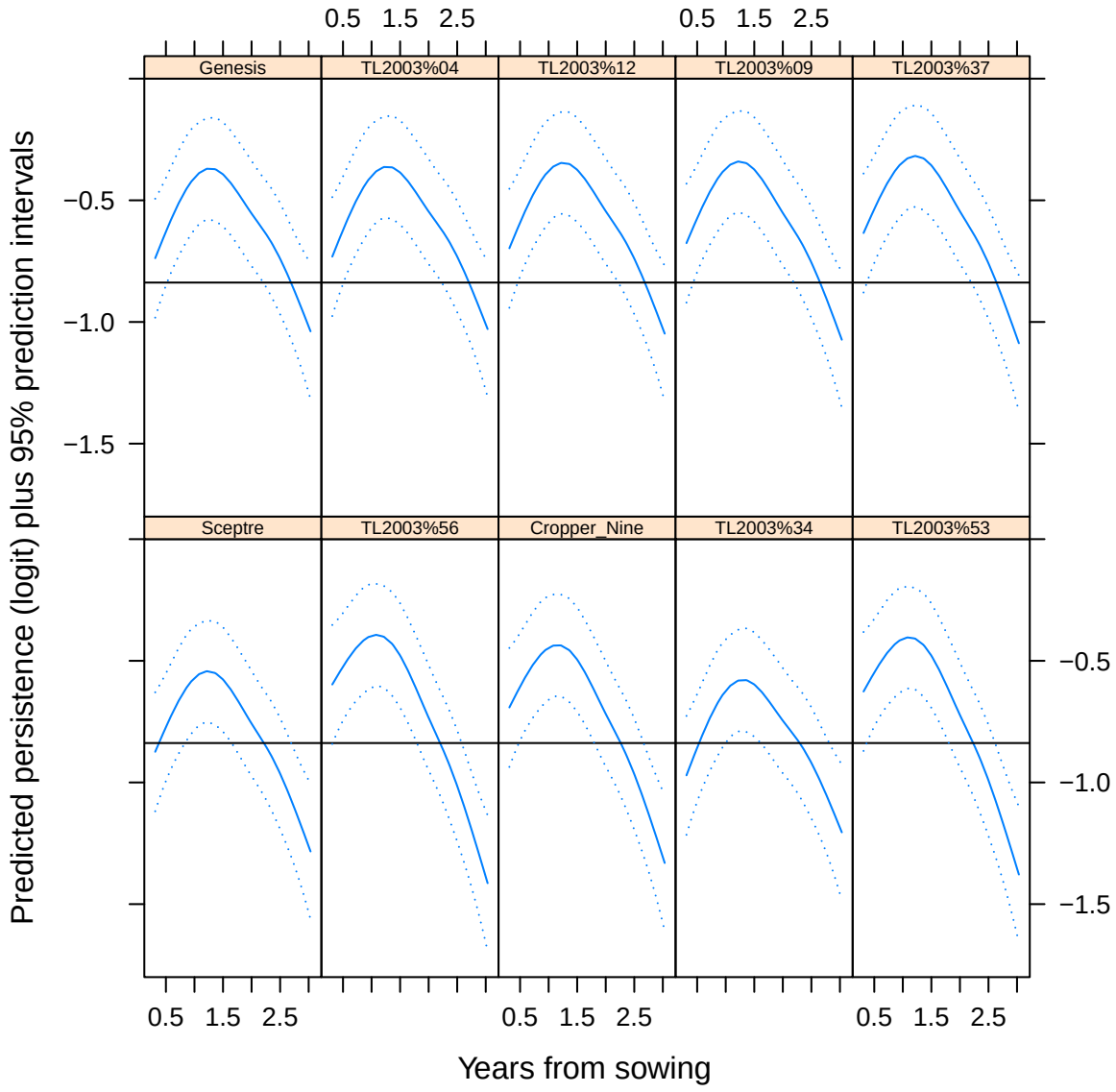
Figure 6.8: Predicted time to persistence level of 30% (-0.838 on logit scale) for each variety based on model TP17 in Table 6.10



shown not to be significantly different to a common spatial model across the harvests. This could indicate that the separable residual model may be reasonable in this case. A similar comparison between individual spatial models and common spatial model for yield data at the other sites (details not presented here) showed a significant difference between models (with the separate spatial models providing a better fit than the common spatial model) for the trials at Leadville ($P=0.022$), Tamworth ($P<0.001$) and Sandigo ($P=0.045$). In these cases it may not be appropriate to assume a separable residual model with common spatial parameters and a more flexible model which allows for different spatial correlation parameters for each harvest may be more suitable.

The genetic models presented relate to the repeated measures nature of the data and performance is modelled at specific times or as a trend over time. The genetic

Figure 6.9: Plot showing predicted splines plus 95% prediction intervals for logit transformed persistence for the top 5 (top panel) and bottom 5 (bottom panel) varieties as ranked on time to reach 30% persistence



models aim to provide a reduced set of genetic effects to enable varietal selection to take place. In some instances the data may warrant a random regression over time approach while in others a different approach, for example a random regression over harvest mean yield, may be desirable, or in other circumstances predictions may be required at each harvest time or for target groups of harvests. Identification of the appropriate approach is facilitated through the implementation of factor analytic (fa) models (Smith et al., 2001), and interpretation of their loadings. The factor analytic models enable predictions to be made at each harvest time, or for target groups of harvests which may be identified using cluster analysis (Cullis et al., 2010). The random regression models for genetic effects provide intercepts and slopes for each variety giving the deviation of the variety from the overall harvest mean profile.

Appropriate selection indices could be developed in collaboration with breeders using the genetic parameters from these models, and hence rankings of varieties presented for selection (Kelly et al., 2007). Smith et al. (2007) presents selection indices based on a weighted sum of the predictions from each harvest time with user supplied weights. This approach can be used to form selection indices for the data in this chapter, based on the predictions obtained at each harvest time, or for groups of harvests, from the Factor analytic model. A similar approach could be used to weight the intercepts and slopes from the random regression models and hence form a selection index.

6.4 Conclusions

The multi-harvest analyses in this chapter have added greater insight into perennial crop variety selection trials than the individual analyses at each harvest time. The temporal correlation between harvests has been identified as an important component to model. The different components of temporal correlation identified by Diggle (1988) have been found and spatial extensions to these components have been able to be fitted. The genetic models have allowed the variety responses to be modelled over time providing further insight into variety by harvest effects. The individual harvest analyses are still important in that they identified global and extraneous spatial trends that needed to be accommodated for each harvest in the multi-harvest analyses thereby resulting in a good starting point for the multi-harvest analyses.

In practice variety selection would normally be based on an analysis of data from multiple sites and methods for across site (MET) analysis are required. In the following chapter the methods used here will be extended to the multi-site situation. It would also be desirable to base selection on a number of key traits, rather than just yield or persistence alone. Ideally a single multivariate analysis of these correlated traits across harvests could be performed and a selection index calculated, based on the multiple traits.

To address the issue of modelling the spatio-temporal correlation at the residual level in a more flexible way that does not assume constant spatial correlation parameters over time, methods incorporating non separable spatio-temporal models need to be investigated. In subsequent chapters such non separable residual models are presented.

Chapter 7

Analysis of multi-environment, multi-harvest trial (MEMHT) data using separable residual models

7.1 Introduction

In Chapters 5 and 6, new methods for modelling the genetic and residual variation in a single multi-harvest trial were presented. Rather than base evaluations purely on results from a single trial, variety selection in perennial crops usually involves a number of trials, conducted across a range of environments, and specific interest lies in investigating variety by environment (location and time) interaction. Therefore, analysis methods for these variety selection trials need to be applicable to the combined data across multiple trials. The approach to analysis is highly motivated by the type of data and how best to connect the data across the trials. In this chapter, the single site, multi-harvest analysis approaches developed in the previous chapters are extended to the multi-environment, multi-harvest trial (MEMHT) situation where varieties are tested across a number of environments. Suitable models are presented for MEMHT data and the approaches applied to the analysis of the lucerne yield and persistence data across trials.

7.2 Model for multi-environment, multi-harvest trial (MEMHT) data

Suppose we have t trials in which m varieties are grown (not all varieties need to be grown in all trials). The j^{th} trial consists of n_j plots in a rectangular array consisting of c_j columns by r_j rows ($n_j = c_j r_j$). Let h_j denote the number of harvests for the j^{th} trial and let h_+ be the total number of trial by harvest combinations ($h_+ = \sum_{j=1}^{t} h_j$). Let $\boldsymbol{y}_j$ be the $h_j n_j \times 1$ vector of observations (for example yield) for trial j, ordered as rows within columns within harvests. The data combined across trials is denoted by $\boldsymbol{y} = (\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, \ldots, \boldsymbol{y}_t^T)^T$; this is an $N \times 1$ vector with $N = \sum_{j=1}^{t} h_j n_j$.

A linear mixed model for the data, $\boldsymbol{y}$ may be written, based on (2.2.1) as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{Z}_g\boldsymbol{u}_g + \boldsymbol{Z}_o\boldsymbol{u}_o + \boldsymbol{e} \tag{7.2.1}$$

where $\boldsymbol{\tau}$ is a vector of fixed effects with design matrix $\boldsymbol{X}$, $\boldsymbol{u}_g$ is the $h_+m \times 1$ vector of variety (or genetic) effects for individual trial by harvest combinations with associated design matrix $\boldsymbol{Z}_g^{(N\times h_+m)}$, $\boldsymbol{u}_o$ is a vector of other random effects with associated design matrix $\boldsymbol{Z}_o$ and $\boldsymbol{e}$ is the $N \times 1$ vector of residuals.

The random effects from the linear mixed model (Equation 7.2.1) are assumed to follow a Normal distribution with zero mean vector and variance-covariance matrix

$$\text{var}\left(\begin{bmatrix} \boldsymbol{u}_g \\ \boldsymbol{u}_o \\ \boldsymbol{e} \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{G}_g & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_o & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R} \end{bmatrix}$$

7.3 Modelling residual effects

In multi-environment (MET) trials, the full residual covariance matrix $\boldsymbol{R}$ is typically given by a block diagonal matrix

$$\boldsymbol{R} = \mathsf{diag}(\boldsymbol{R}_j)$$

where $\boldsymbol{R}_j$ is the residual variance matrix for the j^{th} trial. Therefore each trial has it's own residual covariance structure and residuals are assumed independent between trials. The same approach implemented in the previous chapters in modelling the spatial and temporal residual covariance structure for each trial is therefore applicable in the MEMHT analysis.

7.4 Modelling genetic effects

At the genetic level in MEMHT trials the aim is to model the interaction between trial, harvest within trial and variety. This is variety by environment interaction where environment not only includes location but also harvest or time effects. In general form the variance matrix $\boldsymbol{G}_g$ for these genetic effects across harvests and trials may be represented by

$$\boldsymbol{G}_g = \boldsymbol{G}_{th} \otimes \boldsymbol{I}_m$$

where $\boldsymbol{G}_{th}$ is a $h_+ \times h_+$ genetic variance matrix indexed by all the trial by harvest combinations and $\boldsymbol{I}_m$ is the assumed structure for the varieties.

In the balanced (or near balanced) case of the same number of harvests at the same times at each site Smith et al. (2007) show that $\boldsymbol{G}_{th}$ may be modelled using a separable

form as

$$\boldsymbol{G}_{th} = \boldsymbol{G}_t \otimes \boldsymbol{G}_h \tag{7.4.2}$$

where $\boldsymbol{G}_t$ represents the trial genetic covariance structure (variances and covariances within and between trials), and $\boldsymbol{G}_h$ represents the harvest genetic covariance structure (variances and covariances for the harvests). Smith et al. (2007) recommend using the factor analytic (fa) or unstructured (us) models for $\boldsymbol{G}_t$ and $\boldsymbol{G}_h$ depending on the number of trials and harvests, noting that there may need to be restrictions on $\boldsymbol{G}_t \otimes \boldsymbol{G}_h$ to ensure parameters are identifiable. In unbalanced situations where harvests vary considerably in time and number across the trials, for example in the lucerne trials considered in this thesis, the separable model may not be applicable and alternative approaches need to be considered. Alternative models, for example random regression models, may be considered when the aim is to model the genetic effects using a continuous function over time. Note that while the separable form is desirable for it's ease of interpretation and computing advantages, the separable structure implies that the genetic correlation between pairs of harvests is the same for all trials. This is unlikely to be the case in practice.

7.4.1 Factor analytic models

When the number of harvests is not balanced across the trials and the separable trial by harvest genetic model of (7.4.2) is not able to be used, the full $h_+ \times h_+$ genetic variance matrix $\boldsymbol{G}_{th}$ may be modelled using factor analytic models. The number of trial by harvest combinations may be large and hence many factors may be required to adequately explain the covariance structure. To interpret the results from fitting the factor analytic model the clustering techniques of Cullis et al. (2010) may be implemented. This approach will be illustrated below with the analysis of the lucerne yield data across the five trials.

7.4.2 Random regression models

Where the aim is to model the variety responses over time using a continuous function, the random regression model (Laird & Ware, 1982) provides a suitable approach for MEMHT data. A linear random regression model may be used that allows for random intercepts and slopes (and correlation between these terms) for each variety (as a main effect) and correlated random intercepts and slopes for each variety by environment (trial) response.

Suppose g_{ijk} denotes the random effect for variety i, trial j and harvest k, and t_{jk} represents the harvest time for harvest k at trial j. A linear random regression model for variety by environment interactions g_{ijk} can be formulated as

$$g_{ijk} = u_{i0} + u_{i1}t_{jk} + u_{ij0} + u_{ij1}t_{jk} + \epsilon_{ijk} \tag{7.4.3}$$

where u_{i0} and u_{i1} are the random intercept and slope terms (respectively) for an average variety effect (over environments) for variety i, and u_{ij0} and u_{ij1} are the random intercept and slope terms for variety i at trial j. The term ϵ_{ijk} represents a residual term for genetic

effects, assumed to be independent and identically distributed. For the random regression model, it is assumed

$$\begin{bmatrix} u_{i0} \\ u_{i1} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{G}_{grr} = \begin{bmatrix} g_{g00} & g_{g01} \\ g_{g01} & g_{g11} \end{bmatrix} \right) \tag{7.4.4}$$

and

$$\begin{bmatrix} u_{ij0} \\ u_{ij1} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{G}_{gerr} = \begin{bmatrix} g_{ge00} & g_{ge01} \\ g_{ge01} & g_{ge11} \end{bmatrix} \right) \tag{7.4.5}$$

so that $\boldsymbol{G}_{grr}$ and $\boldsymbol{G}_{gerr}$ are 2×2 covariance matrices for the random intercepts and slopes for each variety (averaged over environments) and for each variety at each site, respectively. Note that the random regression approach could also be extended to encompass semi-parametric regression using cubic smoothing splines as in the previous chapter.

The random regression model provides a balanced genetic variance structure with equal numbers of genetic effects (intercepts and slopes) for each trial, no matter how varied the harvest times or numbers are across trials. Hence the forms for the covariance matrices for the random intercepts and slopes above, are different forms that $\boldsymbol{G}_h$ in (7.4.2) may take.

It may also be desirable to correlate the random coefficients for a variety across the different environments. This may be done by providing a structure for $\boldsymbol{G}_t$ in the separable model (7.4.2). Alternatively, rather than assume a separable structure, the full $2t \times 2t$ matrix of random intercepts and slopes for the t trials may be considered and the covariance structure modelled using an unstructured covariance matrix.

To fully understand the different covariance structures implied by the different random regression models above, it is useful to consider the full $2t \times 2t$ covariance matrix for random intercepts and slopes for the t trials, under each model. These matrices are presented below for a selection of cases (note the parameters are ordered as intercepts for trials 1 to t and then slopes 1 to t).

(a) If the random regression model of (7.4.3) is fitted (corh(1,years):id(Variety) + id(Trial):corh(1,years):id(Variety)), resulting in the two 2×2 covariance matrices $\boldsymbol{G}_{grr}$ and $\boldsymbol{G}_{gerr}$ as defined in (7.4.4) and (7.4.5), the full $2t \times 2t$ covariance matrix is given by

$$\boldsymbol{G}_{2t} = \begin{bmatrix} g_{g00}\boldsymbol{J}_t + g_{ge00}\boldsymbol{I}_t & g_{g01}\boldsymbol{J}_t + g_{ge01}\boldsymbol{I}_t \\ g_{g01}\boldsymbol{J}_t + g_{ge01}\boldsymbol{I}_t & g_{g11}\boldsymbol{J}_t + g_{ge11}\boldsymbol{I}_t \end{bmatrix}$$

where $\boldsymbol{J}_t$ is a $t \times t$ matrix with all elements equal to 1.

This model correlates the genetic effects across trials but it assumes the same correlation for intercepts between trials and similarly for slopes. It also assumes the same correlation between intercepts and slopes within a trial, for all trials and the same variance for intercepts at all trials and the same variance for slopes at all trials.

(b) If the overall random regression for each variety averaged over the trials, is omitted from model (a) and just the individual random regression for each variety at each trial is

fitted (id(Trial).corh(1,years).Variety) the resulting covariance matrix is given by

$$\boldsymbol{G}_{2t} = \begin{bmatrix} g_{ge00}\boldsymbol{I}_t & g_{ge01}\boldsymbol{I}_t \\ g_{ge01}\boldsymbol{I}_t & g_{ge11}\boldsymbol{I}_t \end{bmatrix}$$

This model does not correlate the genetic effects across the trials. It also assumes the same correlation between intercepts and slopes within a trial, for all trials and the same variance for intercepts at all trials and the same variance for slopes at all trials.

(c) If a model is fitted that estimates a different 2×2 $\boldsymbol{G}_{gerr}$ random regression covariance matrix for each trial (at(Trial).corh(1.years).id(Variety)), where these matrices are referred to as $\boldsymbol{G}_{ge1}, \ldots \boldsymbol{G}_{get}$ with

$$\boldsymbol{G}_{gei} = \begin{bmatrix} g_{gei00} & g_{gei01} \\ g_{gei01} & g_{gei11} \end{bmatrix}$$

then the full $2t \times 2t$ covariance matrix is given by

$$\boldsymbol{G}_{2t} = \left[\begin{array}{cccc|cccc} g_{ge100} & 0 & \ldots & 0 & g_{ge101} & 0 & \ldots & 0 \\ 0 & g_{ge200} & \ldots & 0 & 0 & g_{ge201} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & g_{get00} & 0 & 0 & \ldots & g_{get01} \\ \hline g_{ge101} & 0 & \ldots & 0 & g_{ge111} & 0 & \ldots & 0 \\ 0 & g_{ge201} & \ldots & 0 & 0 & g_{ge211} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & g_{get01} & 0 & 0 & \ldots & g_{get11} \end{array}\right]$$

This model does not correlate the intercepts and slopes between trials. It does however allow for different variances for intercepts and slopes from different trials and different correlation between intercepts and slopes within a trial for the different trials.

(d) A combination of the overall random regression and model (c) (corh(1,years):id(Variety) + at(Trial).corh(1.years).id(Variety)) results in

$$\boldsymbol{G}_{2t} = \left[\begin{array}{cccc|cccc} g_{ge100}+g_{ge00} & g_{g00} & \cdots & g_{g00} & g_{g01}+g_{ge101} & g_{g01} & \cdots & g_{g01} \\ g_{g00} & g_{g00}+g_{ge200} & \cdots & g_{g00} & g_{g00} & g_{ge201} & \cdots & g_{g00} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{g00} & g_{g00} & \cdots & g_{g00}+g_{get00} & g_{g00} & g_{g00} & \cdots & g_{get01} \\ \hline g_{ge101} & g_{g00} & \cdots & g_{g00} & g_{g00}+g_{ge100} & g_{g00} & \cdots & g_{g00} \\ g_{g00} & g_{ge201} & \cdots & g_{g00} & g_{g00} & g_{g00}+g_{ge200} & \cdots & g_{g00} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{g00} & g_{g00} & \cdots & g_{get01} & g_{g00} & g_{g00} & \cdots & g_{g00}+g_{get00} \end{array}\right]$$

This model correlates the genetic effects across trials and allows for different correlations between intercepts and slopes at each trial. It is still restrictive in that it restricts the covariances for intercepts between trials to be equal and also the covariances for slopes between trials to be equal.

(e) The genetic effects may be correlated across the trials by providing a structure for $\boldsymbol{G}_t$ in the separable model (7.4.2) for example a common correlation model (cor(Trial).corh(1,years).id(Variety)). This model results in a $t \times t$ correlation matrix modelling the genetic correlation between trials ($\boldsymbol{G}_t$) where

$$\boldsymbol{G}_t = \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}$$

and a 2×2 covariance matrix ($\boldsymbol{G}_{gerr}$)modelling the covariance structure between intercepts and slopes within a trial with

$$\boldsymbol{G}_{gerr} = \begin{bmatrix} g_{ge00} & g_{ge01} \\ g_{ge01} & g_{ge11} \end{bmatrix}$$

Hence the full $2t \times 2t$ covariance matrix is given by

$$\boldsymbol{G}_{2t} = \left[\begin{array}{cccc|cccc} g_{ge00} & \rho g_{ge00} & \dots & \rho g_{ge00} & g_{ge01} & \rho g_{ge01} & \dots & \rho g_{ge01} \\ \rho g_{ge00} & g_{ge00} & \dots & \rho g_{ge00} & \rho g_{ge01} & g_{ge01} & \dots & \rho g_{ge01} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho g_{ge00} & \rho g_{ge00} & \dots & g_{ge00} & \rho g_{ge01} & \rho g_{ge01} & \dots & g_{ge01} \\ \hline g_{ge01} & \rho g_{ge01} & \dots & \rho g_{ge01} & g_{ge11} & \rho g_{ge11} & \dots & \rho g_{ge11} \\ \rho g_{ge01} & g_{ge01} & \dots & \rho g_{ge01} & \rho g_{ge11} & g_{ge11} & \dots & \rho g_{ge11} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho g_{ge01} & \rho g_{ge01} & \dots & g_{ge01} & \rho g_{ge11} & \rho g_{ge11} & \dots & g_{ge11} \end{array}\right]$$

This model assumes genetic effects are correlated between trials but all pairs of intercepts and slopes have the same correlation. It also assumes that all trials have the same correlation between intercepts and slopes within a trial.

(f) If a similar model to (e) is assumed but with an unstructured covariance structure for $\boldsymbol{G}_t$ (us(Trial).corh(1,years).id(Variety)) then the following full covariance matrix is given by

$$\boldsymbol{G}_{2t} = \left[\begin{array}{cccc|cccc} \sigma_{11}g_{ge00} & \sigma_{12}g_{ge00} & \dots & \sigma_{1t}g_{ge00} & \sigma_{11}g_{ge01} & \sigma_{12}g_{ge01} & \dots & \sigma_{1t}g_{ge01} \\ \sigma_{12}g_{ge00} & \sigma_{22}g_{ge00} & \dots & \sigma_{2t}g_{ge00} & \sigma_{12}g_{ge01} & \sigma_{22}g_{ge01} & \dots & \sigma_{2t}g_{ge01} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1t}g_{ge00} & \sigma_{2t}g_{ge00} & \dots & \sigma_{tt}g_{ge00} & \sigma_{1t}g_{ge01} & \sigma_{2t}g_{ge01} & \dots & \sigma_{tt}g_{ge01} \\ \hline \sigma_{11}g_{ge01} & \sigma_{12}g_{ge01} & \dots & \sigma_{1t}g_{ge01} & \sigma_{11}g_{ge11} & \sigma_{12}g_{ge11} & \dots & \sigma_{1t}g_{ge11} \\ \sigma_{12}g_{ge01} & \sigma_{22}g_{ge01} & \dots & \sigma_{2t}g_{ge01} & \sigma_{12}g_{ge11} & \sigma_{22}g_{ge11} & \dots & \sigma_{2t}g_{ge11} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1t}g_{ge01} & \sigma_{2t}g_{ge01} & \dots & \sigma_{tt}g_{ge01} & \sigma_{1t}g_{ge11} & \sigma_{2t}g_{ge11} & \dots & \sigma_{tt}g_{ge11} \end{array}\right]$$

This model assumes the genetic effects are correlated between trials and allows for different

variances for intercepts and slopes at each trial and different covariances between all pairs. However it does impose the restriction that the correlation between intercepts and slopes within a trial is the same for each trial.

(g) The most general of all forms for the $2t \times 2t$ covariance matrix is a fully unstructured covariance matrix (us(Trial+Trial:years).id(Variety)).

The full variance model in all cases is given by

$$\boldsymbol{G}_{th} = \boldsymbol{Z}_{th}\boldsymbol{G}_{2t}\boldsymbol{Z}_{th}^{T}$$

where $\boldsymbol{Z}_{th}$ is a $h_{+} \times 2t$ block diagonal design matrix.

7.5 Analysis of multi-environment, multi-harvest trial (MEMHT) lucerne data

7.5.1 Yield

The lucerne yield data was obtained from five trials (Euloma, Leadville, Sandigo, TCCI, Terry Hie Hie) as detailed in Table 1.1 in Chapter 1. The harvest times were quite varied across these five trials and hence result in an unbalanced trial by harvest data set, with 28 trial by harvest combinations. The yield data was transformed using a cube root transformation (as in Chapter 6) to better approximate the Gaussian distribution.

The residual modelling techniques detailed in Chapter 5 for modelling the spatial and temporal correlation structures within a trial were conducted for each of the five trials and a final residual model was decided upon. This final residual model assumes a 3 way separable temporal by spatial process ante1(Harvest).ar1(Column).ar1(Row) for trials at Euloma, Leadville and TCCI, ante2(Harvest).ar1(Column).ar1(Row) for Terry Hie Hie and us(Harvest).ar1(Column).ar1(Row) for Sandigo (only 2 harvests), together with overall plot (Column.Row) effects for all trials, with spatial correlation on these plot effects in the Row direction for TCCI and Terry Hie Hie. The global and extraneous spatial terms identified in Chapter 3 (Table 3.2) for each trial were also included.

The genetic covariance structure across the 28 trial by harvest combinations was modelled using a factor analytic model (fa). A sequence of genetic models was fitted (as detailed in Table 7.1) to arrive at the final factor analytic model of order 3 (fa3). A model with four factors (fa4 model) was fitted but was not a significant improvement on the fa3 model as based on REMLRT. Reduced rank models were fitted (where specific variances were constrained to zero) at each stage (for example fa1 to fa2) in order to obtain suitable starting values for the next model.

The rotated factor loadings from the fa3 model are presented in Table 7.2. As an aid to interpreting the results from the fa model a cluster analysis was performed based on the correlation matrix estimated from the fa model. A dendrogram showing the cluster analysis results is presented in Figure 7.1 and a heat map representation of the correlations for genetic effects between the harvests and trials is given in Figure 7.2. The genetic

variances for each trial by harvest combination are presented in Table 7.2.

The dendrogram and heat map show three main groups in the trial by harvest combinations and four separate harvests in their own individual groups, as detailed in Table 7.3.

Selection indices have been formed for these groups, using the approach in Cullis et al. (2010) (and as discussed in Chapter 6), using equal weights for harvests within each group. That is, if u_{ij} represents the genetic effect of variety i at trial by harvest combination j, $(j = 1, \ldots, 28)$ (as given in Table 7.2) and $\tilde{u}_{ij}$ is the BLUP of u_{ij}, the predicted selection index for variety i for the 7 groups, $\hat{I}_i^{(d)}$, for $d = 1, \ldots, 7$, may be defined as

$$\begin{aligned}
\hat{I}^{(1)} &= (\{\tilde{u}_{i1}\} + \{\tilde{u}_{i4}\} + \{\tilde{u}_{i26}\} + \{\tilde{u}_{i6}\} + \{\tilde{u}_{i14}\} + \{\tilde{u}_{i9}\} + \{\tilde{u}_{i2}\} \\
&\quad + \{\tilde{u}_{i23}\} + \{\tilde{u}_{i18}\} + \{\tilde{u}_{i21}\} + \{\tilde{u}_{i3}\})/11 \\
\hat{I}^{(2)} &= (\{\tilde{u}_{i5}\} + \{\tilde{u}_{i15}\} + \{\tilde{u}_{i22}\} + \{\tilde{u}_{i24}\} \\
+ &\quad \{\tilde{u}_{i20}\} + \{\tilde{u}_{i28}\} + \{\tilde{u}_{i8}\} + \{\tilde{u}_{i12}\} + \{\tilde{u}_{i19}\} + \{\tilde{u}_{i25}\})/10 \\
\hat{I}^{(3)} &= (\{\tilde{u}_{i7}\} + \{\tilde{u}_{i10}\} + \{\tilde{u}_{i16}\})/3 \\
\hat{I}^{(4)} &= (\{\tilde{u}_{i11}\}) \\
\hat{I}^{(5)} &= (\{\tilde{u}_{i13}\}) \\
\hat{I}^{(6)} &= (\{\tilde{u}_{i17}\}) \\
\hat{I}^{(7)} &= (\{\tilde{u}_{i27}\})
\end{aligned}$$

Figure 7.3 presents the predicted selection indices for each variety for each of the groups. It can be seen that group 3 is negatively correlated with groups 2 and 4. Varieties may be ranked for each of these groups. For example, variety number 51 is the highest ranked variety for groups 1 and 6 but ranked fairly low for groups 4 and 5.

7.5.2 Persistence

The lucerne persistence data was measured at 5 sites at a number of assessment times (varying from 3 times at Euloma to 7 times at Leadville), as detailed in Table 1.2 in Chapter 1. The assessment times varied across the trials and resulted in an unbalanced trial by harvest data set of 25 trial by harvest combinations. The data (percentages) was transformed using a logit transformation to better approximate the Gaussian distribution (as in Chapter 6).

As discussed in Chapters 5 and 6 the lucerne persistence data is continuous over time and the aim is to model the genetic response over time. Random regression is therefore an ideal approach as it may be used to model variety deviations from the underlying overall mean response at each trial and to correlate the genetic effects (through the random intercepts and slopes) across trials. In Chapter 6 the aim was to predict the time when varieties persistence dropped to 30% at Terry Hie Hie, however across the 5 trials, only Terry Hie Hie and Sandigo were measured for sufficient length or had persistence levels

Table 7.1: Summary of genetic models fitted for the MEMHT analysis of lucerne yield data. Residual log-likelihoods (denoted by ℓ), the number of parameters in $\boldsymbol{G}_g$ (denoted by npar) and P values (based on REMLRT with previous model) where appropriate, are presented for each model.

Model	Genetic Model	$\boldsymbol{G}_g$	ℓ	$\boldsymbol{G}_g$ npar	P
MY1	diag(TrialHarv):id(Variety)	$\oplus_{k=1}^{28}\sigma_{gk}^2 \otimes \boldsymbol{I}_m$	1288.028	28	
MY2	Variety+ diag(TrialHarv):id(Variety)	$(\sigma_g^2\boldsymbol{J}_{28} + \oplus_{k=1}^{28}\sigma_{gk}^2) \otimes \boldsymbol{I}_m$	1288.107	29	
MY3	fa(TrialHarv,1):id(Variety)	$(\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1^T + \boldsymbol{\Psi}) \otimes \boldsymbol{I}_m$	1361.335	56	P<0.001
MY4	fa(TrialHarv,2):id(Variety)	$(\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_2^T + \boldsymbol{\Psi}) \otimes \boldsymbol{I}_m$	1390.317	83	P<0.001
MY5	fa(TrialHarv,3):id(Variety)	$(\boldsymbol{\Lambda}_3\boldsymbol{\Lambda}_3^T + \boldsymbol{\Psi}) \otimes \boldsymbol{I}_m$	1411.428	109	P=0.023
MY6	fa(TrialHarv,4):id(Variety)	$(\boldsymbol{\Lambda}_4\boldsymbol{\Lambda}_4^T + \boldsymbol{\Psi}) \otimes \boldsymbol{I}_m$	1415.258	134	n.s

where all models have residual model

at(Trial, c(1,2,4)):ante1(Harvest):ar1(Col):ar1(Row)

+ at(Trial,5):ante2(Harvest):ar1(Col):ar1(Row)

+ at(Trial,3):us(Harvest):ar1(Col):ar1(Row)

+ at(Trial,c(1,2,3)):Col:Row + at(Trial, c(4,5)):Col:ar1v(Row)

where trials (Trial) are numbered 1-5 in alphabetical order (Euloma, Leadville, Sandi, TCCI, Terry). Models also include the spatial global and extraneous terms for each trial as identified in Chapter 3.

Table 7.2: Rotated factor loadings($\mathbf{\Lambda}$), % Variance accounted for(%VAF) and genetic variances estimated from the fa3 factor analytic model (MY5) fitted to the lucerne yield data across all trials

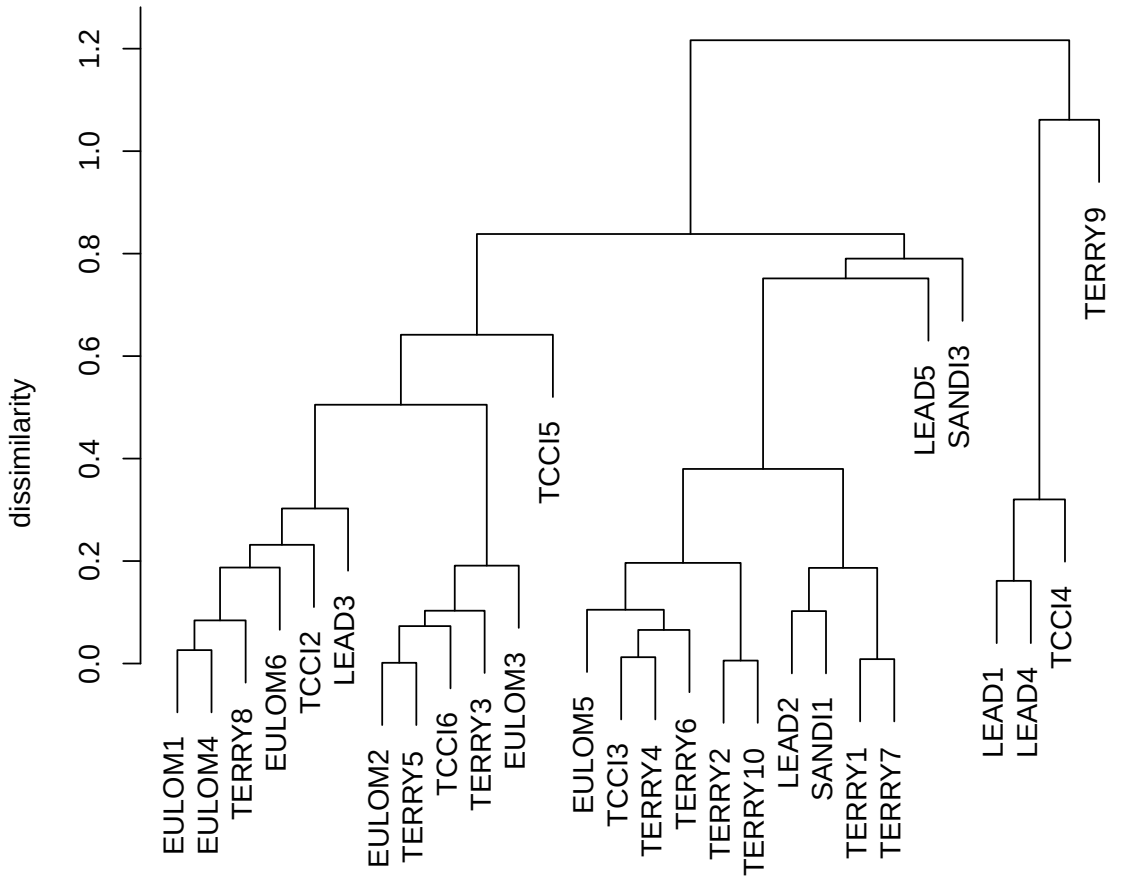
	Trial	Harvest No.	Harvest Date	$\mathbf{\Lambda}_1$	$\mathbf{\Lambda}_2$	$\mathbf{\Lambda}_3$	%VAF	Genetic Variance
1	Euloma	1	1/11/03	0.19	0.18	-0.05	100	0.074
2	Euloma	2	18/12/03	0.07	0.02	0.06	100	0.009
3	Euloma	3	16/02/04	0.02	0.10	0.09	100	0.018
4	Euloma	4	14/04/04	0.13	0.10	0.00	100	0.029
5	Euloma	5	14/10/04	0.15	0.04	0.00	99	0.023
6	Euloma	6	30/11/04	0.03	0.12	-0.04	100	0.018
7	Lead	1	14/10/04	-0.14	0.07	-0.04	100	0.025
8	Lead	2	25/11/04	0.12	-0.02	-0.11	100	0.026
9	Lead	3	21/12/04	0.09	0.11	0.01	62	0.030
10	Lead	4	3/02/05	-0.05	0.10	-0.03	100	0.015
11	Lead	5	24/11/05	0.02	-0.05	0.00	19	0.016
12	Sandi	1	17/11/03	0.16	-0.04	-0.08	89	0.038
13	Sandi	3	04/01/05	0.07	-0.16	-0.00	14	0.225
14	TCCI	2	10/12/03	0.14	0.09	-0.12	100	0.043
15	TCCI	3	16/02/04	0.09	-0.00	0.01	100	0.009
16	TCCI	4	08/10/04	-0.03	0.04	-0.09	100	0.010
17	TCCI	5	19/11/04	0.02	0.10	-0.02	36	0.028
18	TCCI	6	21/12/04	0.12	0.13	0.12	100	0.048
19	Terry	1	13/11/03	0.03	-0.03	-0.03	100	0.002
20	Terry	2	16/12/03	0.10	-0.07	0.00	100	0.015
21	Terry	3	05/02/04	0.02	0.01	0.04	100	0.002
22	Terry	4	11/03/04	0.17	-0.03	0.03	100	0.032
23	Terry	5	20/04/04	0.10	0.04	0.08	100	0.017
24	Terry	6	22/06/04	0.53	-0.09	0.01	90	0.323
25	Terry	7	12/10/04	0.02	-0.03	-0.03	100	0.002
26	Terry	8	02/12/04	0.05	0.09	0.02	100	0.011
27	Terry	9	23/12/04	-0.02	0.01	0.06	21	0.017
28	Terry	10	24/01/05	0.13	-0.09	0.02	100	0.026

above 30% to make this specific problem relevant across trials. It may be of more interest to look at all trial locations and investigate the variety deviations (intercepts and slopes of deviations) from the overall underlying mean profile at each site. By investigating these deviations across the sites variety by environment interactions may be identified. By modelling the variety profiles over time across the sites, predictions may also be obtained and varieties ranked, at times of interest which may not have been an assessment time.

The modelling process commenced with modelling the underlying overall mean at each trial. This was done using a linear model (1 + lin(years)) for trials at Leadville, Sandigo and TCCI and a cubic smoothing spline model (1 + lin(years)+ spl(years) + dev(years)) for Terry Hie Hie (as in Chapter 6). (Note Euloma was excluded from this MET analysis as it was only measured at 3 times and not over a sufficiently long period of time to compare with the other trials).

The residual modelling process for each trial followed that outlined in Chapters 5 and

Figure 7.1: Dendrogram of the dissimilarity matrix from the fa3 model (MY5 in Table 7.1) fitted to the lucerne MEMHT yield data

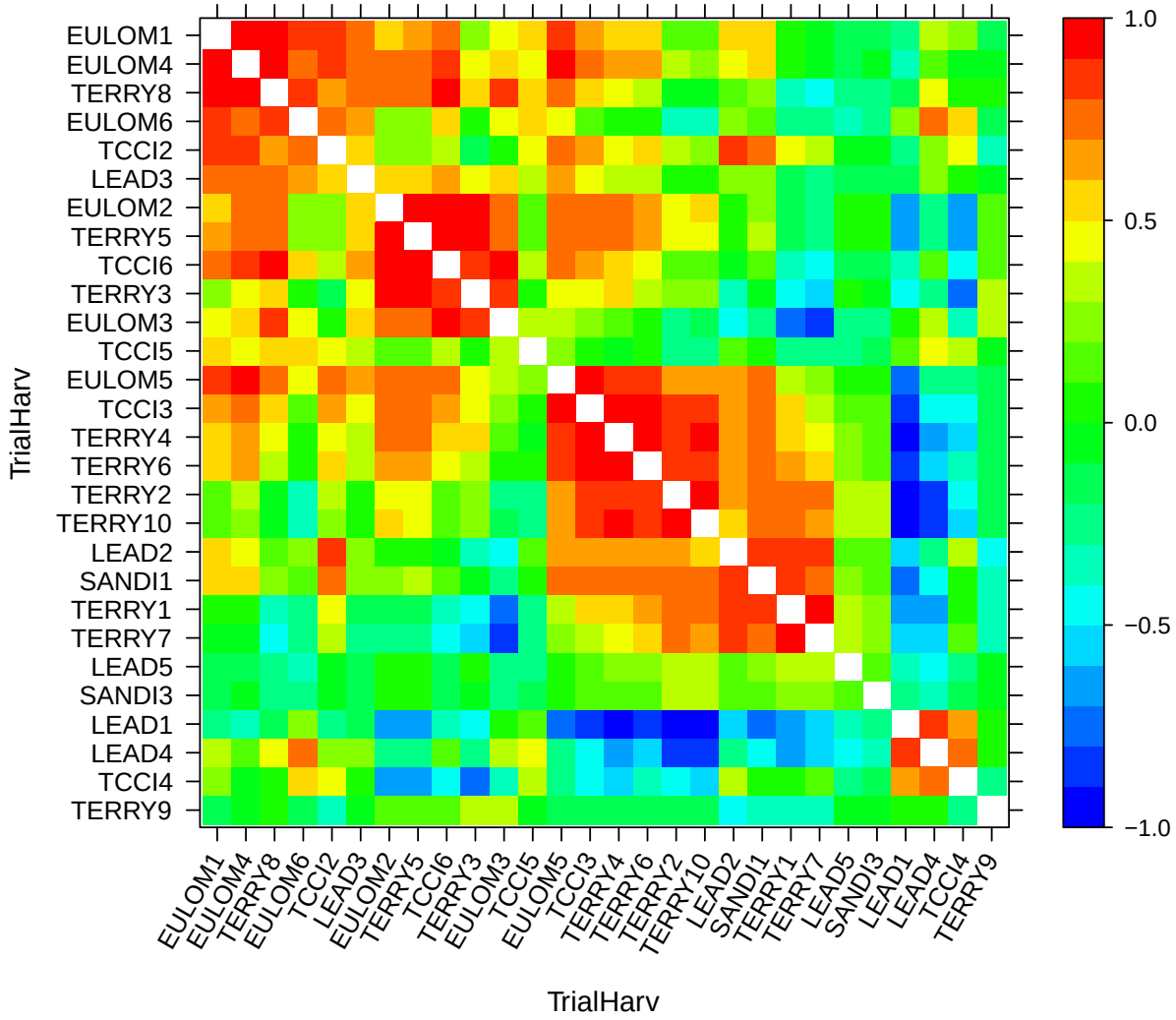


6 in order to model the residual spatial and temporal correlation structure at each trial. The final residual model across all trials included a 3 way separable temporal by spatial process with a heterogeneous variance autoregressive process of order1 (ar1h) for the temporal component and an autoregressive process of order1 for the spatial components in the row and column directions (ar1h(Harvest).ar1(Column).ar1(Row)) for all trials. The model also included an overall plot effect (Column.Row) for each trial and the global and extraneous spatial terms identified for each trial in Chapter 3 (Table 3.4).

The genetic response over time was modelled using the random regression models detailed in section 7.4.2. The sequence of genetic random regression models fitted is presented in Table 7.4.

The first model (MP1) fitted a random regression model with correlated random in-

Figure 7.2: Heat map representation of the genetic correlation matrix from the fa3 model (MY5 in Table 7.1) fitted to the lucerne MEMHT yield data



tercepts and slopes for each variety by environment response but with no correlation of random effects (random intercepts and slopes) between the trials. Hence a single 2×2 covariance matrix $\boldsymbol{G}_{ge}$ was estimated. This is the covariance structure discussed in (b) in section 7.4.2. The next model (MP2) included the random regression of MP1 but with an additional overall random regression for an average variety effect (over environments). This is the model presented in (a) in section 7.4.2. This model correlates the genetic effects across the trials. It was a significant improvement on MP1 (REMLRT = 62.668 on 3 df, $P < 0.001$).

Both models MP1 and MP2 assumed that the correlation between intercepts and slopes within a trial was the same for all trials, and also assumed that the variance for intercepts at all trials was the same and the variance for slopes at all trials was the same. The next

Figure 7.3: Plot of predicted selection indices from the seven groups (in Table 7.3) identified in the cluster analysis based on the genetic correlation matrix from the fa3 model fitted to the lucerne MEMHT yield data. Numbers refer to variety names in Table 6.7

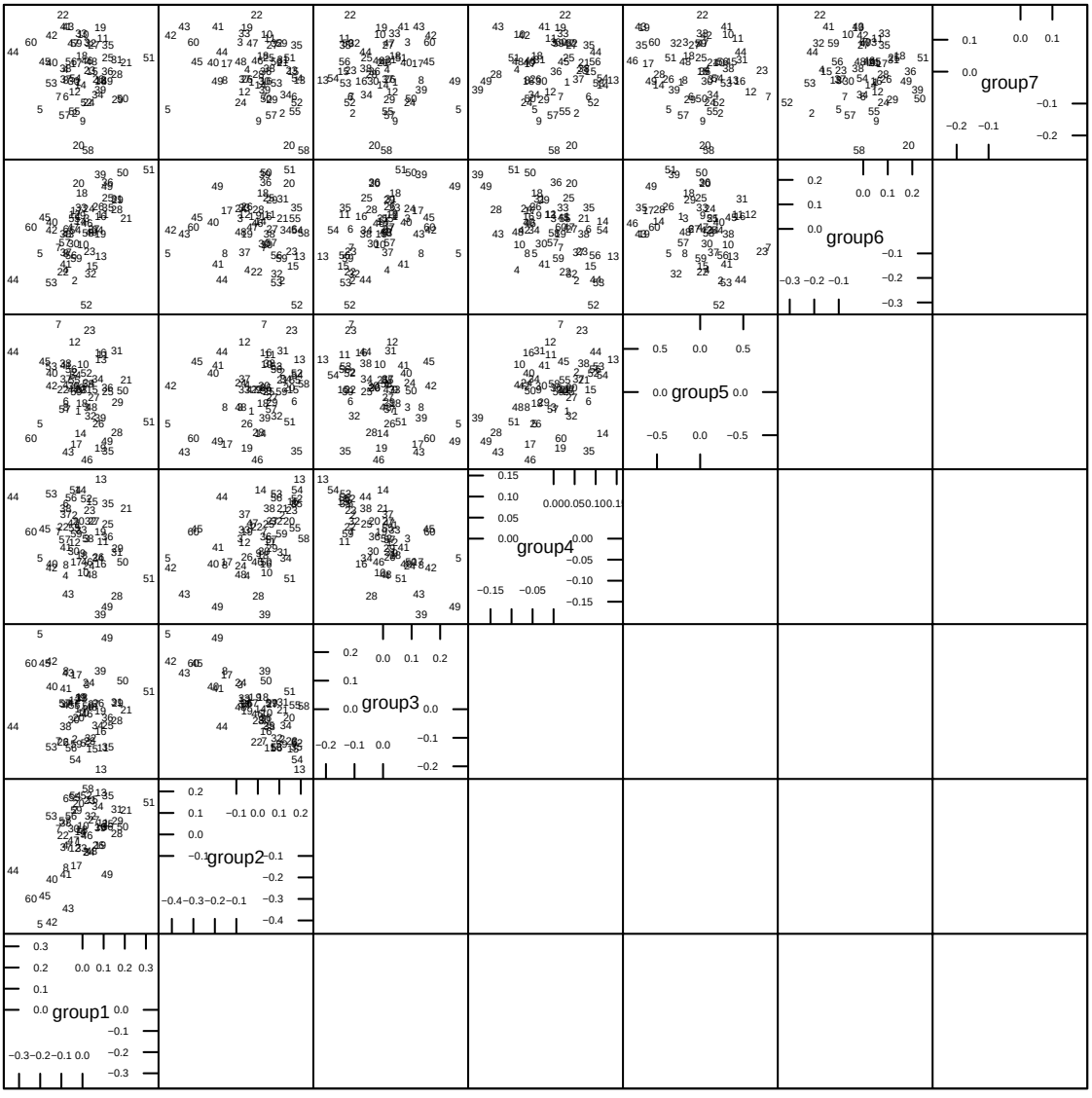


Table 7.3: Table showing groupings of trial by harvest combinations based on the cluster analysis after fitting a fa3 model (MY5 in Table 7.1) to the lucerne MEMHT yield data, as well as harvest variables

TrialHarv	harvest date	yld	group
EULOM1	1/11/03	551	1
EULOM4	14/04/04	884	1
TERRY8	2/12/04	1012	1
EULOM6	30/11/04	771	1
TCCI2	10/12/03	664	1
LEAD3	21/12/04	637	1
EULOM2	18/12/03	910	1
TERRY5	20/04/04	779	1
TCCI6	21/12/04	843	1
TERRY3	5/02/04	1948	1
EULOM3	16/02/04	1084	1
EULOM5	14/10/04	834	2
TCCI3	16/02/04	701	2
TERRY4	11/03/04	808	2
TERRY6	22/06/04	624	2
TERRY2	16/12/03	1314	2
TERRY10	24/01/05	977	2
LEAD2	25/11/04	447	2
SANDI1	17/11/03	381	2
TERRY1	13/11/03	1934	2
TERRY7	12/10/04	1257	2
LEAD1	14/10/04	909	3
LEAD4	3/02/05	538	3
TCCI4	8/10/04	1056	3
LEAD5	24/11/05	1472	4
SANDI3	04/01/05	237	5
TCCI5	19/11/04	954	6
TERRY9	23/12/04	1229	7

model (MP3) allowed for different 2×2 covariance structures for correlated intercepts and slopes for each trial, but this model did not correlate the genetic effects across the trials. This model was not an improvement on MP2 (based on AIC values). The next models (MP4 and MP5) added an overall average variety effect random regression to MP3. Model MP4 did not correlate the random intercepts and slopes within a trial and was fitted more to get starting values for MP5. Unfortunately model MP4 identified some of the slopes and intercepts at some trials as having zero variance (when the overall average random regression was also included) so attempts at fitting MP5 where the intercepts and slopes were correlated, were unsuccessful.

The next two models (MP6 and MP7) followed the separable form of (7.4.2), where a 2×2 covariance matrix was estimated for the covariance structure for the intercepts and slopes within a trial ($\boldsymbol{G}_h$) and a 4×4 covariance matrix was estimated to correlate the random effects across the trials ($\boldsymbol{G}_t$). Model MP6 fitted an equal correlation structure

to $\boldsymbol{G}_t$ (cor) thereby assuming equal correlations between all pairs of trials and equal variances. It is unlikely all pairs of trials will be correlated in the same way and have the same variance, so a more plausible model would be to fit an unstructured model for $\boldsymbol{G}_t$. This was fitted as model MP7. To make this model estimable, a constraint needed to be applied to either $\boldsymbol{G}_h$ or $\boldsymbol{G}_t$. As Smith et al. (2007) point out, the choice of constraint will cause different parameters for $\boldsymbol{G}_h$ and $\boldsymbol{G}_t$ but the Kronecker product $\boldsymbol{G}_h \otimes \boldsymbol{G}_t$ is unique no matter what constraint is chosen. In model MP7 the first element (variance) for $\boldsymbol{G}_t$ was set to 1 so all variances and covariances in $\boldsymbol{G}_h$ and $\boldsymbol{G}_t$ were relative to this. Unfortunately there were problems with this model with some correlations being estimated greater than 1. An attempt to fit a factor analytic (fa1) model to $\boldsymbol{G}_t$ also had problems converging.

The next models (MP8,MP9,MP10,MP11,MP12) fitted a direct covariance structure to the full 8 intercepts and slopes. Model MP8 fitted a diagonal structure to the 8 random effects in order to provide starting values for the subsequent models. Model MP9 fitted an unstructured covariance structure for the full set of intercepts and slopes (thereby allowing for different variances and correlations between intercepts and slopes both within and between trials). This model had a high log-likelihood but required many parameters to be estimated. More parsimonious models (MP10, MP11 and MP12) were fitted using the factor analytic model with 1,2 and 3 factors respectively. The fa2 model (MP11) was a significant improvement on MP10 (REMLRT= 32.574 on 7 df, $P < 0.001$). The fa3 model (MP12)was not a significant improvement on the fa2 model (MP11).

Comparing AIC values for all models in Table 7.4 it can be seen that models MP2 and MP6 are very similar in goodness-of-fit to MP11. Models MP2 and MP6 are much more restrictive than MP11 in that they assume the same correlation between intercepts and slopes within a trial for all trials, and they also assume the same correlation for intercepts between trials and the same correlation between slopes between trials, whereas MP11 allows for different correlations between all pairs of intercepts and slopes.

Given that it has been suggested (see Vaida & Blanchard, 2005) that a difference of 2 or less in AIC may not be reliable in ranking two models, and the fact that model MP11 provides greater insight into the genetic covariance structure across the harvests and trials, the final model has been chosen as MP11.

The results from fitting model MP11 are presented in Table 7.5 and Figures 7.4 and 7.5. These figures show the intercepts and slopes from the random regression model for each variety across the 4 trials while the table presents the genetic correlations (between intercepts and slopes) for the 4 trials. The data has been adjusted so that time zero is actually 6 months after sowing. Hence the intercepts are predicted deviations from trial harvest means at time 6 months and the intercepts may be interpreted to reflect an establishment effect with a positive intercept indicating an above average establishment. The slopes reflect whether the varieties rate of persistence is above or below average after establishment. It is clearly desirable to find varieties with a large positive intercept and large positive slope (hence varieties in the top right hand corner of each panel in Figure 7.5).

It is apparent from Figure 7.4 and Table 7.5 that intercepts were highly positively

correlated between trials for all pairs of trials. This indicates generally consistent results in establishment across the trials. The slopes were also very consistent between trials, with the possible exception between Terry Hie Hie and Sandigo. At all trials (except Sandigo) the intercepts and slopes were negatively correlated within a trial.

From Figure 7.5 it can be seen that varieties 8 and 44 perform quite differently to other varieties. Variety 44 has a high slope for all trials indicating a high rate of persistence but while the intercept for this variety at Sandigo is average, it is below average at the other trials, being very low at Terry Hie Hie. This variety therefore shows good establishment and persistence at Sandigo but did not establish well at the other trials. Variety 45 performed consistently well with both moderately high intercepts and slopes across all trials.

Figure 7.4: Pairwise plots of intercepts(i) and slopes(sl) for the four trials from random regression model (MP11) fitted to the lucerne MEMHT persistence data

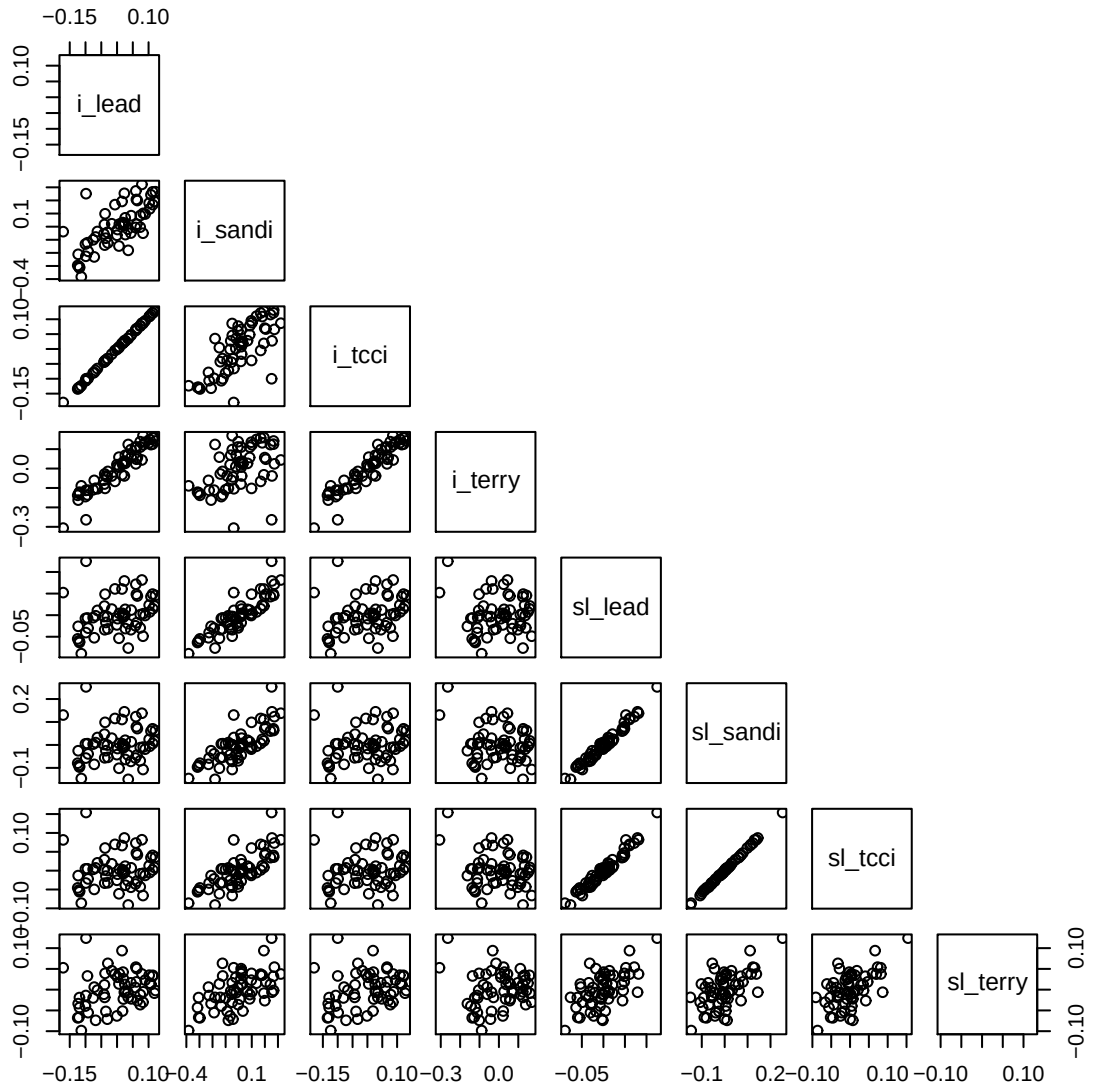


Table 7.4: Summary of genetic models fitted for the MEMHT analysis of lucerne persistence data. Residual log-likelihoods (denoted by ℓ), the number of parameters in the genetic model (denoted by npar) and AIC values (given as differences from the final model MP11) are presented for each model

Model	Genetic model	ℓ	npar	AIC
MP1	id(Trial):corh(1,years):id(Variety) + diag(Trial).Harvest.Variety	3729.431	7	56
MP2	corh(1,years):id(Variety)+ id(Trial):corh(1,years):id(Variety) + diag(Trial).Harvest.Variety	3760.765	10	0
MP3	at(Trial):corh(1,years):id(Variety)+ diag(Trial).Harvest.Variety	3743.696	16	46
MP4	diag(1,years):id(Variety)+ at(Trial):diag(1,years):id(Variety) + diag(Trial).Harvest.Variety	3763.486	14	2
MP5	corh(1,years):id(Variety)+ at(Trial):corh(1,years):id(Variety) + diag(Trial).Harvest.Variety	–	19	
MP6	cor(Trial).corh(1,years):id(Variety) + diag(Trial).Harvest.Variety	3759.976	8	-2
MP7	us(Trial).corh(1,years):id(Variety) + diag(Trial).Harvest.Variety	–	16	
MP8	diag(Trial, Trial:years):id(Variety) + diag(Trial).Harvest.Variety	3734.110	12	57
MP9	us(Trial, Trial:years):id(Variety) + diag(Trial).Harvest.Variety	3784.576	40	12
MP10	fa1(Trial, Trial:years):id(Variety) + diag(Trial).Harvest.Variety	3761.314	20	18
MP11	fa2(Trial, Trial:years):id(Variety) + diag(Trial).Harvest.Variety	3777.601	27	0
MP12	fa3(Trial, Trial:years):id(Variety) + diag(Trial).Harvest.Variety	3780.527	33	6

where all models include the residual model: at(Trial):ar1h(Harvest).ar1(Col).ar1(Row) + at(Trial):Col:Row
and include the spatial global and extraneous terms for each site as identified in Chapter 3. The term years refers to time in years from sowing (centred across all trials).
The overall mean for trial Terry Hie Hie was modelled using a spline (1+lin(years)+spl(years)+dev(years))
and the overall mean for the other trials was modelled using a linear regression (1+lin(years)) for each trial.

Figure 7.5: Plot of predicted intercepts and slopes for each variety at each trial from random regression model (MP11) fitted to the lucerne MEMHT persistence data. Numbers refer to variety names in Table 6.7

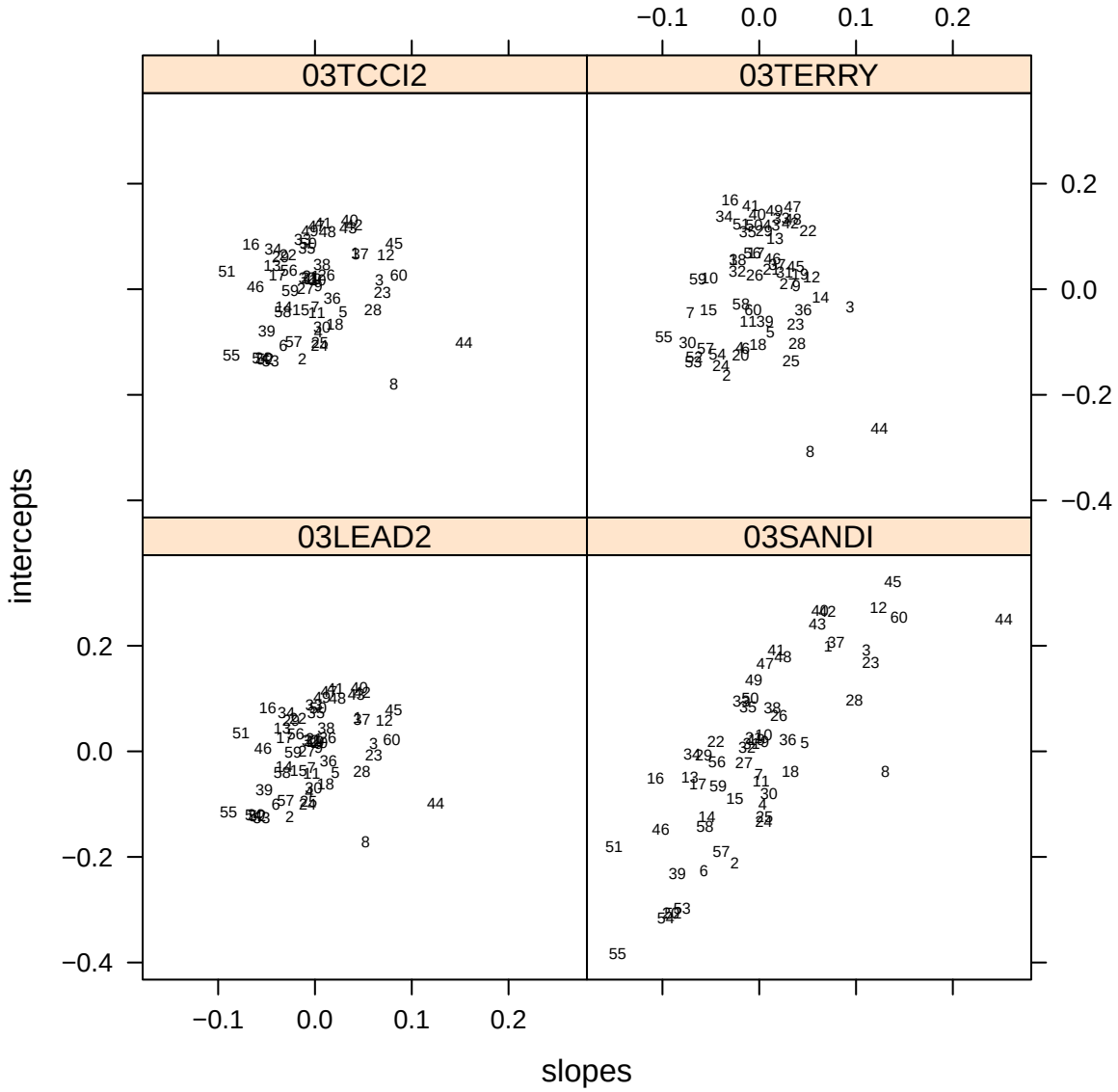


Table 7.5: REML estimates of variances (on diagonal) and correlations (off diagonals) for intercepts (int) and slopes (sl) from model MP11 fitted to the lucerne MEMHT persistence data

	int lead	int sandi	int tcci	int terry	sl lead	sl sandi	sl tcci	sl terry
int lead	0.010							
int sandi	0.637	0.016						
int tcci	0.981	0.775	0.012					
int terry	0.983	0.768	1.000	0.020				
sl lead	-0.485	0.364	-0.307	-0.316	0.003			
sl sandi	-0.462	0.166	-0.338	-0.344	0.746	0.009		
sl tcci	-0.657	0.162	-0.498	-0.507	0.978	0.754	0.003	
sl terry	-0.363	0.187	-0.251	-0.257	0.650	0.491	0.647	0.002

7.6 Summary

The methods presented in this chapter provide a new approach for the analysis of multi-harvest variety selection data from multiple trials. The approach allows variety predictions to be made across environments and gives an insight into variety by environment interactions. The approach extends that of Smith et al. (2007) by including the new extended spatial by temporal residual models for each trial, enabling trials with large numbers of harvests to be analysed, as well as being able to model the genetic effects over time for each trial.

At the residual level, the models assumed a three way separable harvest by column by row Kronecker product to model the spatial and temporal correlation within a trial. The residual models fitted assume constant row and column spatial correlations across all harvests within each trial. As we have found in Chapter 3 and Chapter 6 this assumption may not hold for some sites and it may be more sensible to assume more flexible non-separable residual models allowing for differing spatial correlation parameters across the harvests.

The genetic models for yield are quite general with the fa model fitted to all trial by harvest combinations with no harvest within trial structure. The clustering aids in interpretation from this fa model however the groups obtained from the cluster analysis are slightly difficult to interpret. The random regression models fitted to persistence across trials result in intercepts and slopes (from the linear deviation random regressions) for each variety for each trial. It is possible to rank the varieties on their intercepts (which reflect establishment) and slopes however it may not be straightforward to make variety selections based on these two traits. The differences between intercepts and slopes for a variety across the trials gives insight into genotype by environment (gxe) interaction.

In the following chapters such non-separable residual models will be investigated.

Chapter 8

Non-separable residual models: Multivariate AR1 in 1 spatial dimension

8.1 Introduction

In Chapters 5, 6 and 7, models were presented for the analysis of multi-harvest data using separable residual covariance structures. In these models the residual covariance structure can be written as the product of a temporal and a spatial component as in (5.2.3), where the spatial correlation parameters are common to all harvests in the trial. This assumption of separability of the temporal and spatial covariance processes is a strong assumption and may not hold in many situations. In Chapter 3 the analysis of the example data sets at each individual harvest time revealed some sites having similar spatial correlation parameters across the harvest times while other sites had quite different spatial correlation over the harvests. It may be more reasonable in some situations to assume a non-separable residual covariance model that allows for varying spatial correlation parameters across the harvests.

In this chapter the multivariate autoregressive model (MVAR1) is investigated as a suitable non-separable model for modelling the residual covariance structure from multi-harvest (or multi-trait) data collected from plots in a single spatial dimension (e.g. along a row) or for data from plots in a row column lattice, but where spatial correlation is only evident in one direction (e.g. row) and the residuals are assumed to be uncorrelated in the column direction.

While the multivariate autoregressive model can be found in the literature to model time series data, there are no known published examples of the MVAR1 model being used to model multi-harvest data on spatially correlated plots in the field. This chapter investigates the suitability of these models for the analysis of perennial crop variety selection data and develops the framework for their implementation in the linear mixed model with estimation using REML.

The chapter begins with the general form of the multivariate autoregressive model

and then looks at the conditions necessary for this model to be suitable for multivariate data measured on plots in a row or column in the field. Different forms of the MVAR1 model are presented and investigated for different situations. Code has been written in R (R Development Core Team, 2012) and these models are then applied to two multi-trait examples (using the lucerne yield and persistence data from two sites) and a multi-harvest example (using the lucerne yield data from Terry Hie Hie). In these examples spatial correlation was only evident in the row direction and independence between plots assumed in the column direction.

8.2 Multivariate AR1 model (MVAR1)

Consider a multivariate autoregressive model for the residuals from a linear mixed model (2.2.1) modelling data from t measurements (harvests or multiple traits) on each of n plots in a row in the field. Let $\boldsymbol{e}_i$ denote the $t \times 1$ vector of residuals at plot i, $i = 1, \ldots, n$ and e_{ik} denote the kth residual on the ith plot, $k = 1, \ldots, t$. As well as the residuals being spatially correlated the t residuals are also likely to be related across time (or traits).

The multivariate analogue of the first order autoregressive process (hereafter referred to as MVAR1) can be written as

$$\boldsymbol{e}_1 = \boldsymbol{\epsilon}_1 \tag{8.2.1}$$

$$\boldsymbol{e}_{i+1} = \boldsymbol{\Omega}\boldsymbol{e}_i + \boldsymbol{\epsilon}_{i+1} \tag{8.2.2}$$

for $i = 1, \ldots, n-1$, where $\boldsymbol{e}_i$ represents the $t \times 1$ vector of residuals for plot i, $\boldsymbol{\Omega}$ is a $t \times t$ matrix of spatial dependency parameters (a multivariate version of ϕ introduced in Chapter 1) which has spatial dependency parameters between neighbouring plots at the same time (or same trait) on the diagonals and the spatial dependency parameters between neighbouring plots at different times (or different traits) on the off-diagonals, and the $\boldsymbol{\epsilon}_i$'s are independent $t \times 1$ vectors with $\boldsymbol{\epsilon}_1 \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\epsilon}_{i+1} \sim N(\mathbf{0}, \boldsymbol{\Sigma} - \boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T)$.

Therefore

$$\boldsymbol{e}_1 \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \tag{8.2.3}$$

$$\boldsymbol{e}_{i+1}|\boldsymbol{e}_i \sim N(\boldsymbol{\Omega}\boldsymbol{e}_i, \boldsymbol{\Sigma} - \boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T) \tag{8.2.4}$$

Now,

$$\begin{aligned} E(\boldsymbol{e}_{i+1}) &= E(E(\boldsymbol{e}_{i+1}|\boldsymbol{e}_i)) \\ &= E(\boldsymbol{\Omega}\boldsymbol{e}_i) \\ &= \boldsymbol{\Omega}E(\boldsymbol{e}_i) \\ &= \boldsymbol{\Omega} \times \mathbf{0} \\ &= \mathbf{0} \end{aligned}$$

and

$$\begin{aligned}\text{var}\,(\boldsymbol{e}_{i+1}) &= E(\text{var}\,(\boldsymbol{e}_{i+1}|\boldsymbol{e}_i)) + \text{var}\,(E(\boldsymbol{e}_{i+1}|\boldsymbol{e}_i)) \\ &= E(\boldsymbol{\Sigma} - \boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T) + \text{var}\,(\boldsymbol{\Omega}\boldsymbol{e}_i) \\ &= \boldsymbol{\Sigma} - \boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T + \boldsymbol{\Omega}\text{var}\,(\boldsymbol{e}_i)\,\boldsymbol{\Omega}^T\end{aligned}$$

Thus,

$$\begin{aligned}\text{var}\,(\boldsymbol{e}_2) &= \boldsymbol{\Sigma} - \boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T + \boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T \\ &= \boldsymbol{\Sigma}\end{aligned}$$

and by induction $\text{var}\,(\boldsymbol{e}_i) = \boldsymbol{\Sigma}$, for all i. Thus $\boldsymbol{e}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$.
The $\boldsymbol{e}_i$ are correlated. Note firstly that

$$\begin{aligned}\text{cov}\,(\boldsymbol{e}_{i+1},\ \boldsymbol{e}_i) &= \text{cov}\,(\boldsymbol{\Omega}\boldsymbol{e}_i + \boldsymbol{\epsilon}_{i+1},\ \boldsymbol{e}_i) \\ &= \text{cov}\,(\boldsymbol{\Omega}\boldsymbol{e}_i,\ \boldsymbol{e}_i) + \text{cov}\,(\boldsymbol{\epsilon}_{i+1},\ \boldsymbol{e}_i) \\ &= \boldsymbol{\Omega}\text{cov}\,(\boldsymbol{e}_i,\ \boldsymbol{e}_i) + \boldsymbol{0} \\ &= \boldsymbol{\Omega}\text{var}\,(\boldsymbol{e}_i) \\ &= \boldsymbol{\Omega}\boldsymbol{\Sigma}\end{aligned}$$

and hence

$$\text{cov}\,(\boldsymbol{e}_i,\ \boldsymbol{e}_{i+1}) = (\boldsymbol{\Omega}\boldsymbol{\Sigma})^T = \boldsymbol{\Sigma}\boldsymbol{\Omega}^T$$

by definition. Furthermore,

$$\begin{aligned}\text{cov}\,(\boldsymbol{e}_{i+2},\ \boldsymbol{e}_i) &= \text{cov}\,(\boldsymbol{\Omega}\boldsymbol{e}_{i+1} + \boldsymbol{\epsilon}_{i+2},\ \boldsymbol{e}_i) \\ &= \boldsymbol{\Omega}\text{cov}\,(\boldsymbol{e}_{i+1},\ \boldsymbol{e}_i) \\ &= \boldsymbol{\Omega}\boldsymbol{\Omega}\boldsymbol{\Sigma} \\ &= \boldsymbol{\Omega}^2\boldsymbol{\Sigma}\end{aligned}$$

and so $\text{cov}\,(\boldsymbol{e}_i,\ \boldsymbol{e}_{i+2}) = \boldsymbol{\Sigma}(\boldsymbol{\Omega}^T)^2$. Thus $\text{cov}\,(\boldsymbol{e}_{i+j},\ \boldsymbol{e}_i) = \boldsymbol{\Omega}^j\boldsymbol{\Sigma}, j = 1, \ldots, n-1$. Therefore the variance covariance matrix of $\boldsymbol{e}$ can be written as

$$\text{var}\,(\boldsymbol{e}) = \boldsymbol{R} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}\boldsymbol{\Omega}^T & \boldsymbol{\Sigma}(\boldsymbol{\Omega}^T)^2 & \ldots & \boldsymbol{\Sigma}(\boldsymbol{\Omega}^T)^{n-1} \\ \boldsymbol{\Omega}\boldsymbol{\Sigma} & \boldsymbol{\Sigma} & \boldsymbol{\Sigma}\boldsymbol{\Omega}^T & \ldots & \boldsymbol{\Sigma}(\boldsymbol{\Omega}^T)^{n-2} \\ \boldsymbol{\Omega}^2\boldsymbol{\Sigma} & \boldsymbol{\Omega}\boldsymbol{\Sigma} & \boldsymbol{\Sigma} & \ldots & \boldsymbol{\Sigma}(\boldsymbol{\Omega}^T)^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Omega}^{n-1}\boldsymbol{\Sigma} & \boldsymbol{\Omega}^{n-2}\boldsymbol{\Sigma} & \boldsymbol{\Omega}^{n-3}\boldsymbol{\Sigma} & \ldots & \boldsymbol{\Sigma} \end{bmatrix}$$

and $\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{R})$.

8.2.1 Inverse variance matrix for MVAR1

The joint density function of $\boldsymbol{e}$, $f(\boldsymbol{e})$, can be written as the product of conditional densities and the marginal density of one or more variables. Hence

$$f(\boldsymbol{e}) = f(\boldsymbol{e}_1)f(\boldsymbol{e}_2|\boldsymbol{e}_1)f(\boldsymbol{e}_3|\boldsymbol{e}_2,\boldsymbol{e}_1).....f(\boldsymbol{e}_n|\boldsymbol{e}_{n-1}...\boldsymbol{e}_1)$$

Using (8.2.3) and (8.2.4), the log density of $\boldsymbol{e}$ may be written as

$$\begin{aligned}\log f(\boldsymbol{e}) = \log\det\boldsymbol{\Sigma}^{-1} - \frac{1}{2}\boldsymbol{e}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_1 + \frac{n-1}{2}\log\det\boldsymbol{U} \\ -\frac{1}{2}\sum_{i=1}^{n-1}(\boldsymbol{e}_{i+1}-\boldsymbol{\Omega}\boldsymbol{e}_i)^T\boldsymbol{U}(\boldsymbol{e}_{i+1}-\boldsymbol{\Omega}\boldsymbol{e}_i) \qquad (8.2.5)\end{aligned}$$

where $\boldsymbol{U} = (\boldsymbol{\Sigma} - \boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T)^{-1}$.

From this joint density function the elements of the inverse variance covariance matrix could be derived by taking the negative of the second partial derivatives as in Verbyla (1985). Alternatively, by simply expanding the quadratic forms in (8.2.5) the exponent of the full multivariate normal distribution of $\boldsymbol{e}$ can be obtained. Thus the quadratic forms can be written as

$$\begin{aligned}
& \boldsymbol{e}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_1 + \sum_{i=1}^{n-1}(\boldsymbol{e}_{i+1}-\boldsymbol{\Omega}\boldsymbol{e}_i)^T\boldsymbol{U}(\boldsymbol{e}_{i+1}-\boldsymbol{\Omega}\boldsymbol{e}_i) \\
= \; & \boldsymbol{e}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_1 + \boldsymbol{e}_1^T\boldsymbol{\Omega}^T\boldsymbol{U}\boldsymbol{\Omega}\boldsymbol{e}_1 - \boldsymbol{e}_1^T\boldsymbol{\Omega}^T\boldsymbol{U}\boldsymbol{e}_2 - \boldsymbol{e}_2^T\boldsymbol{U}\boldsymbol{\Omega}\boldsymbol{e}_1 + \boldsymbol{e}_2^T\boldsymbol{U}\boldsymbol{e}_2 + \\
& \boldsymbol{e}_2^T\boldsymbol{\Omega}\boldsymbol{U}\boldsymbol{\Omega}\boldsymbol{e}_2 - \boldsymbol{e}_2^T\boldsymbol{\Omega}\boldsymbol{U}\boldsymbol{e}_3 - \boldsymbol{e}_3^T\boldsymbol{U}\boldsymbol{\Omega}\boldsymbol{e}_2 + \ldots + \\
= \; & \boldsymbol{e}_1^T(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Omega}^T\boldsymbol{U}\boldsymbol{\Omega})\boldsymbol{e}_1 - \boldsymbol{e}_1^T\boldsymbol{\Omega}^T\boldsymbol{U}\boldsymbol{e}_2 - \boldsymbol{e}_2^T\boldsymbol{U}\boldsymbol{\Omega}\boldsymbol{e}_1 + \boldsymbol{e}_2^T(\boldsymbol{U} + \boldsymbol{\Omega}^T\boldsymbol{U}\boldsymbol{\Omega})\boldsymbol{e}_2 \\
& -\boldsymbol{e}_2^T\boldsymbol{\Omega}^T\boldsymbol{U}\boldsymbol{e}_3 - \boldsymbol{e}_3^T\boldsymbol{U}\boldsymbol{\Omega}\boldsymbol{e}_3 + \ldots + \boldsymbol{e}_n^T\boldsymbol{U}\boldsymbol{e}_n \\
= \; & \boldsymbol{e}^T\boldsymbol{R}^{-1}\boldsymbol{e}
\end{aligned}$$

where

$$\boldsymbol{R}^{-1} = \begin{bmatrix}
\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Omega}^T\boldsymbol{U}\boldsymbol{\Omega} & -\boldsymbol{\Omega}^T\boldsymbol{U} & 0 & \ldots & 0 \\
-\boldsymbol{U}\boldsymbol{\Omega} & \boldsymbol{U} + \boldsymbol{\Omega}^T\boldsymbol{U}\boldsymbol{\Omega} & -\boldsymbol{\Omega}^T\boldsymbol{U} & \ldots & 0 \\
0 & -\boldsymbol{U}\boldsymbol{\Omega} & \boldsymbol{U} + \boldsymbol{\Omega}^T\boldsymbol{U}\boldsymbol{\Omega} & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & -\boldsymbol{U}\boldsymbol{\Omega} & \boldsymbol{U}
\end{bmatrix}$$

Recall that $\boldsymbol{R} = \mathrm{var}\,(\boldsymbol{e})$, so that $\boldsymbol{R}^{-1} = \mathrm{var}\,(\boldsymbol{e})^{-1}$ is the inverse covariance matrix. Note that the first and last diagonal elements of $\boldsymbol{R}^{-1}$ are not the same. In some examples where there is a defined ordering of the "units" on which the multivariate measurements are made, for instance in the case of multiple variables measured on an animal at a number of times (time is ordered) there may be no issue with this difference in the first and last elements of $\boldsymbol{R}^{-1}$. However in the case here, of multivariate data measured on plots in a

row (or column) in a field, the same results are required to hold no matter if observations are observed from left to right along the row or from right to left. In this case the inverse covariance matrix is required to be directionally invariant and hence the the first diagonal element must be equal to the last. This will mean constraints will need to be applied to the elements of $\boldsymbol{R}$ or $\boldsymbol{R}^{-1}$.

8.2.2 Constraint for directional invariance of inverse covariance matrix of MVAR1

For the inverse variance covariance matrix $\boldsymbol{R}^{-1}$ to be directionally invariant, the first and last diagonal elements are required to be equal. i.e $\boldsymbol{U} = \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Omega}^T\boldsymbol{U}\boldsymbol{\Omega}$. To investigate the condition for this to hold, the following may be written:

$$\begin{aligned} \boldsymbol{U} &= (\boldsymbol{\Sigma} - \boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T)^{-1} \\ &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T\boldsymbol{\Sigma}^{-1})^{-1}\boldsymbol{\Sigma}^{-1} \end{aligned} \tag{8.2.6}$$

Now

$$\begin{aligned} \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Omega}^T\boldsymbol{U}\boldsymbol{\Omega} &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\boldsymbol{\Omega}^T(\boldsymbol{\Sigma} - \boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T)^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1} \\ &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{\Omega}^T(\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T - \boldsymbol{\Sigma})^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1} \\ &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Omega}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega})^{-1}\boldsymbol{\Sigma}^{-1} \end{aligned} \tag{8.2.7}$$

Thus for (8.2.6) to equal (8.2.7) it is sufficient that

$$\boldsymbol{\Omega}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T\boldsymbol{\Sigma}^{-1}$$

This can be achieved if

$$\boldsymbol{\Omega}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{\Omega}^T \tag{8.2.8}$$

This condition is sufficient, but may not be necessary. The condition implies

$$\begin{aligned} \boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma} &= \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T \\ \boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1} &= \boldsymbol{\Omega}^T\boldsymbol{\Sigma}^{-1} \\ \boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega} &= \boldsymbol{\Omega}^T\boldsymbol{\Sigma}^{-1} \end{aligned}$$

Now as $(\boldsymbol{\Omega}\boldsymbol{\Sigma})^T = \boldsymbol{\Sigma}^T\boldsymbol{\Omega}^T = \boldsymbol{\Sigma}\boldsymbol{\Omega}^T$ this would mean $(\boldsymbol{\Omega}\boldsymbol{\Sigma})^T = \boldsymbol{\Omega}\boldsymbol{\Sigma}$. Thus $\boldsymbol{\Omega}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{\Omega}^T$ means that $\text{cov}\,(\boldsymbol{e}_i,\ \boldsymbol{e}_{i+1}) = \text{cov}\,(\boldsymbol{e}_{i+1},\ \boldsymbol{e}_i)$ for $i = 1, \ldots, n-1$, or symmetry in the relationship

between successive pairs. Notice also that

$$\begin{aligned}\boldsymbol{\Omega}^2\boldsymbol{\Sigma} &= \boldsymbol{\Omega}\boldsymbol{\Omega}\boldsymbol{\Sigma} \\ &= \boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T \\ &= (\boldsymbol{\Omega}\boldsymbol{\Sigma})\boldsymbol{\Omega}^T \\ &= \boldsymbol{\Sigma}\boldsymbol{\Omega}^T\boldsymbol{\Omega}^T \\ &= \boldsymbol{\Sigma}(\boldsymbol{\Omega}^T)^2\end{aligned}$$

a result that carries over to all powers.

Therefore, under the constraint $\boldsymbol{\Omega}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{\Omega}^T$ the variance covariance matrix for $\boldsymbol{e}$ is a symmetric matrix as follows:

$$\begin{aligned}\text{var}\,(\boldsymbol{e}) = \boldsymbol{R} &= \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Omega}\boldsymbol{\Sigma} & \boldsymbol{\Omega}^2\boldsymbol{\Sigma} & \dots & \boldsymbol{\Omega}^{n-1}\boldsymbol{\Sigma} \\ \boldsymbol{\Omega}\boldsymbol{\Sigma} & \boldsymbol{\Sigma} & \boldsymbol{\Omega}\boldsymbol{\Sigma} & \dots & \boldsymbol{\Omega}^{n-2}\boldsymbol{\Sigma} \\ \boldsymbol{\Omega}^2\boldsymbol{\Sigma} & \boldsymbol{\Omega}\boldsymbol{\Sigma} & \boldsymbol{\Sigma} & \dots & \boldsymbol{\Omega}^{n-3}\boldsymbol{\Sigma} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Omega}^{n-1}\boldsymbol{\Sigma} & \boldsymbol{\Omega}^{n-2}\boldsymbol{\Sigma} & \boldsymbol{\Omega}^{n-3}\boldsymbol{\Sigma} & \dots & \boldsymbol{\Sigma} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{I}_t & \boldsymbol{\Omega} & \boldsymbol{\Omega}^2 & \dots & \boldsymbol{\Omega}^{n-1} \\ \boldsymbol{\Omega} & \boldsymbol{I}_t & \boldsymbol{\Omega} & \dots & \boldsymbol{\Omega}^{n-2} \\ \boldsymbol{\Omega}^2 & \boldsymbol{\Omega} & \boldsymbol{I}_t & \dots & \boldsymbol{\Omega}^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Omega}^{n-1} & \boldsymbol{\Omega}^{n-2} & \boldsymbol{\Omega}^{n-3} & \dots & \boldsymbol{I}_t \end{bmatrix} (\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})\end{aligned}$$

and the inverse of the variance covariance matrix is:

$$\boldsymbol{R}^{-1} = \begin{bmatrix} \boldsymbol{U} & -\boldsymbol{\Omega}^T\boldsymbol{U} & 0 & \dots & 0 \\ -\boldsymbol{U}\boldsymbol{\Omega} & \boldsymbol{U} + \boldsymbol{\Omega}^T\boldsymbol{U}\boldsymbol{\Omega} & -\boldsymbol{\Omega}^T\boldsymbol{U} & \dots & 0 \\ 0 & -\boldsymbol{U}\boldsymbol{\Omega} & \boldsymbol{U} + \boldsymbol{\Omega}^T\boldsymbol{U}\boldsymbol{\Omega} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -\boldsymbol{U}\boldsymbol{\Omega} & \boldsymbol{U} \end{bmatrix}$$

If we further consider the constraint and it's implications on the above form of $\boldsymbol{R}^{-1}$, we may be able to write the inverse variance covariance matrix in more simple terms involving $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$. To see this consider the following: $\boldsymbol{U} = (\boldsymbol{\Sigma} - \boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T)^{-1}$ and as $\boldsymbol{\Omega}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{\Omega}^T$ under the constraint, then

$$\begin{aligned}\boldsymbol{U} &= (\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{\Omega}^T\boldsymbol{\Omega}^T)^{-1} \\ &= (\boldsymbol{I} - (\boldsymbol{\Omega}^T)^2)^{-1}\boldsymbol{\Sigma}^{-1}\end{aligned}$$

or

$$\begin{aligned} \boldsymbol{U} &= (\boldsymbol{\Sigma} - \boldsymbol{\Omega}\boldsymbol{\Omega}\boldsymbol{\Sigma})^{-1} \\ &= ((\boldsymbol{I} - \boldsymbol{\Omega}^2)\boldsymbol{\Sigma})^{-1} \\ &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{I} - \boldsymbol{\Omega}^2)^{-1} \end{aligned}$$

Now

$$\boldsymbol{U}^{-1} = \boldsymbol{\Sigma} - \boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^T \tag{8.2.9}$$

but also

$$\boldsymbol{U}^{-1} = \boldsymbol{\Sigma} - \boldsymbol{\Omega}^2\boldsymbol{\Sigma} \tag{8.2.10}$$

Hence using (8.2.10)

$$\begin{aligned} \boldsymbol{\Omega}\boldsymbol{U}^{-1} &= \boldsymbol{\Omega}\boldsymbol{\Sigma} - \boldsymbol{\Omega}^3\boldsymbol{\Sigma} \\ &= \boldsymbol{\Sigma}\boldsymbol{\Omega}^T - \boldsymbol{\Sigma}(\boldsymbol{\Omega}^T)^3 \end{aligned}$$

and using (8.2.9)

$$\boldsymbol{U}^{-1}\boldsymbol{\Omega}^T = \boldsymbol{\Sigma}\boldsymbol{\Omega}^T - \boldsymbol{\Sigma}(\boldsymbol{\Omega}^T)^3$$

Therefore

$$\boldsymbol{\Omega}\boldsymbol{U}^{-1} = \boldsymbol{U}^{-1}\boldsymbol{\Omega}^T$$

and

$$\boldsymbol{\Omega}^T\boldsymbol{U} = \boldsymbol{U}\boldsymbol{\Omega}$$

Note also that

$$\boldsymbol{U}\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{I} - \boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}$$

so that

$$(\boldsymbol{I} - \boldsymbol{\Omega}^2)^{-1}\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{I} - \boldsymbol{\Omega}^2)^{-1}$$

Now,

$$\begin{aligned} \boldsymbol{U} + \boldsymbol{\Omega}^T \boldsymbol{U} \boldsymbol{\Omega} &= \boldsymbol{U}(\boldsymbol{I} + \boldsymbol{\Omega}^2) \\ &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{I} - \boldsymbol{\Omega}^2)^{-1}(\boldsymbol{I} + \boldsymbol{\Omega}^2) \end{aligned}$$

Therefore the inverse variance covariance matrix of $\boldsymbol{e}$ can be written as $\text{var}(\boldsymbol{e}^{-1}) =$

$$\begin{aligned} \boldsymbol{R}^{-1} &= \boldsymbol{I}_n \otimes (\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1}) \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{\Omega} & 0 & \dots & 0 \\ -\boldsymbol{\Omega} & \boldsymbol{I} + \boldsymbol{\Omega}^2 & -\boldsymbol{\Omega} & \dots & 0 \\ 0 & -\boldsymbol{\Omega} & \boldsymbol{I} + \boldsymbol{\Omega}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \boldsymbol{I} \end{bmatrix} \\ &= [\boldsymbol{I}_n \otimes (\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1})][\boldsymbol{I}_{nt} - \boldsymbol{F}_{1n} \otimes \boldsymbol{\Omega} + \boldsymbol{E}_{1n} \otimes \boldsymbol{\Omega}^2] \end{aligned} \tag{8.2.11}$$

where $\boldsymbol{F}_{1n}$ is a $n \times n$ matrix which has 'ones' on the first sub and super-diagonals and zeroes elsewhere, and $\boldsymbol{E}_{1n}$ is a $n \times n$ diagonal matrix of 'ones' with the exception of the first and last diagonal elements which equal zero. Note $\boldsymbol{F}_{1k}$ and $\boldsymbol{E}_{1k}$ are similarly defined $k \times k$ matrices for any k. It can be seen that this inverse covariance matrix resembles the univariate ar1 inverse covariance matrix.

8.2.3 Constraints: Special cases

Given specific forms for $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$, there are a number of situations where $\boldsymbol{\Omega}\boldsymbol{\Sigma}$ is automatically symmetric. For example

- If $\boldsymbol{\Sigma} = \text{diag}(\sigma_i^2)$ and $\boldsymbol{\Omega} = \text{diag}(\omega_i)$, the resulting MVAR1 residual model is equivalent to the residual structure resulting from a separate analysis of each harvest time (or trait), assuming an ar1 correlation structure between plots, with ω_i the spatial correlation parameter for time i. This model is similar to model Y1 in Table 6.1 but with spatial correlation in only one direction.

- If $\boldsymbol{\Omega} = \omega \boldsymbol{I}_t$ then any suitable covariance matrix for $\boldsymbol{\Sigma}$ (e.g. corv, corh, us, ar1v, ar1h, as defined in Table 5.1) will result in a symmetric $\boldsymbol{\Omega}\boldsymbol{\Sigma}$. This MVAR1 model is equivalent to the separable residual covariance structure in (5.2.3) but in one spatial dimension e.g. row (i.e. $\text{ar1}_r \otimes \boldsymbol{R}_h$ where $\boldsymbol{R}_h = \boldsymbol{\Sigma}$), with ω the spatial correlation parameter in the row direction. To see this consider (8.2.11) and the following:

$$\begin{aligned} \boldsymbol{R}^{-1} &= [\boldsymbol{I}_n \otimes (\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_t - \omega^2 \boldsymbol{I}_t)^{-1})][\boldsymbol{I}_{nt} - \boldsymbol{F}_{1n} \otimes \omega \boldsymbol{I}_t + \boldsymbol{E}_{1n} \otimes \omega^2 \boldsymbol{I}_t] \\ &= [\boldsymbol{I}_n \otimes (1 - \omega^2)^{-1} \boldsymbol{\Sigma}^{-1}][(\boldsymbol{I}_n - \omega \boldsymbol{F}_{1n} + \omega^2 \boldsymbol{E}_{1n}) \otimes \boldsymbol{I}_t] \\ &= [(1 - \omega^2)^{-1}(\boldsymbol{I}_n - \omega \boldsymbol{F}_{1n} + \omega^2 \boldsymbol{E}_{1n})] \otimes \boldsymbol{\Sigma}^{-1} \\ &= (\text{ar1}(\omega) \otimes \boldsymbol{\Sigma})^{-1} \end{aligned}$$

  This model is similar to models Y3 and Y4 in Table 6.1 but with spatial correlation in only one direction.

- If $\boldsymbol{\Sigma} = \mathsf{corv}$, $\boldsymbol{\Omega} = \omega \boldsymbol{J}_t$, where $\boldsymbol{J}_t$ is a $t \times t$ matrix of 1's

- If $\boldsymbol{\Sigma} = \mathsf{ar1v}$, $\boldsymbol{\Omega} = \omega_1 \boldsymbol{I}_t + \omega_2 \boldsymbol{E}_{1t} - \omega_2 \boldsymbol{F}_{1t}$

where $\boldsymbol{E}_{1t}$ and $\boldsymbol{F}_{1t}$ are $t \times t$ matrices as defined above.

Other more general forms of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ will need constraints applied to the parameters to ensure $\boldsymbol{\Omega\Sigma}$ is symmetric.

8.2.4 Constraints required to impose symmetry condition for $\boldsymbol{\Omega\Sigma}$.

Firstly consider the 2×2 case. If $\boldsymbol{\Omega\Sigma} = \boldsymbol{\Sigma\Omega}^T$ with $\boldsymbol{\Sigma}$ a symmetric 2×2 covariance matrix,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

and

$$\boldsymbol{\Omega} = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}$$

Then

$$\boldsymbol{\Omega\Sigma} = \begin{bmatrix} \omega_{11}\sigma_{11} + \omega_{12}\sigma_{12} & \omega_{11}\sigma_{12} + \omega_{12}\sigma_{22} \\ \omega_{21}\sigma_{11} + \omega_{22}\sigma_{12} & \omega_{21}\sigma_{12} + \omega_{22}\sigma_{22} \end{bmatrix}$$

and

$$\boldsymbol{\Sigma\Omega}^T = \begin{bmatrix} \omega_{11}\sigma_{11} + \omega_{12}\sigma_{12} & \omega_{21}\sigma_{11} + \omega_{22}\sigma_{12} \\ \omega_{11}\sigma_{12} + \omega_{12}\sigma_{22} & \omega_{21}\sigma_{12} + \omega_{22}\sigma_{22} \end{bmatrix}$$

Hence for these two matrices to be equal we require,

$$\omega_{11}\sigma_{12} + \omega_{12}\sigma_{22} = \omega_{21}\sigma_{11} + \omega_{22}\sigma_{12}$$

Note that if $\omega_{12} = \omega_{21} = 0$ then $\omega_{11} = \omega_{22} = \omega$, say, and $\boldsymbol{\Omega} = \omega \boldsymbol{I}$. If $\omega_{12} = \omega_{21}$ but not equal to 0, then the constraint reduces to $\omega_{12}(\sigma_{22} - \sigma_{11}) = (\omega_{22} - \omega_{11})\sigma_{12}$.

For the 3×3 case there are 3 constraints, namely

$$\begin{aligned} \omega_{11}\sigma_{12} + \omega_{12}\sigma_{22} + \omega_{13}\sigma_{32} &= \sigma_{11}\omega_{21} + \sigma_{12}\omega_{22} + \sigma_{13}\omega_{23} \\ \omega_{11}\sigma_{13} + \omega_{12}\sigma_{23} + \omega_{13}\sigma_{33} &= \sigma_{11}\omega_{31} + \sigma_{12}\omega_{32} + \sigma_{13}\omega_{33} \\ \omega_{21}\sigma_{13} + \omega_{22}\sigma_{23} + \omega_{23}\sigma_{33} &= \sigma_{12}\omega_{31} + \sigma_{22}\omega_{32} + \sigma_{23}\omega_{33} \end{aligned}$$

For the 4×4 case there are 6 constraints.

The general form for the constraints for each element (i, j) of the matrix $\mathbf{\Omega\Sigma}$, is given by

$$\sum_{k=1}^{t} \omega_{ik}\sigma_{kj} = \sum_{k=1}^{t} \sigma_{ik}\omega_{jk}$$

with the constraints for the diagonal elements automatically holding, and the constraints for the upper off diagonals being the same as those for the lower off diagonals. Hence in general there are $t(t-1)/2$ constraints where t is the number of measurements on each plot.

8.2.5 Impact of symmetry constraint on cross correlations

Consider the case of 2 measurements made on each plot in a row (take the measurements to be at times 1 and 2 but note that the same principles hold for multi-trait data). The matrix of covariances for plots 1 apart (e.g. for plots 1 and 2) is given by $\mathbf{\Omega\Sigma}$. Let its elements be referred to as $\{c_{ij}\}$ so that the diagonal elements c_{ii} represent the covariance between plots 1 and 2 at time i. The element c_{12} represents the covariance between plot 1 at time 1 and plot 2 at time 2 while c_{21} represents the covariance between plot 2 at time 1 and plot 1 at time 2. This matrix is symmetric (due to the symmetry constraint) so that implies the covariance between plot 1 at time 1 and plot 2 at time 2 is equal to the covariance between plot 2 at time 1 and plot 1 at time 2. As the variance for each plot is equal, this implies that the correlation between these measurements is also equal. These are similar to the cross correlations that Sain & Cressie (2007) refer to in a Multivariate conditional autoregressive (MCAR) model, in which they allow for non-equal cross correlations. In their example of zinc and cadmium measurements upstream and downstream in a river this difference in cross correlations makes sense as the plots (sites) are ordered and cannot be reversed (the upstream site is always upstream of the downstream site). However in the example here of data on plots in a row the results are required to hold if the order of the plots is reversed and ordered from left to right along the row or right to left. Therefore intuitively it makes sense for the cross correlations in this case to be the same.

8.2.6 Alternative forms for Ω and Σ requiring constraints

Σ

The matrix $\mathbf{\Sigma}$ represents the covariance structure within each plot, hence needs to be a suitable covariance matrix for the situation involved. For multi-trait data (different traits measured on each plot at a single time), this covariance structure may be best represented using an unstructured (`us`) model in which different variances are assumed for each trait and different covariances between all pairs of traits are also assumed. The `us` model may require a large number of parameters to be estimated.

For multi-harvest data (where measurements are made for the same trait on each plot

at multiple times) more parsimonious models may be suitable that make use of the ordering of the measurement times, and the fact that the correlation between measurements in time usually decays as the time between the measurements increases. For example, exponential or autoregressive or antedependence structures may be used to model the covariance matrix.

$\boldsymbol{\Omega}$

The matrix $\boldsymbol{\Omega}$ is a spatial dependence matrix which connects the plots in a suitable way. The form of $\boldsymbol{\Omega}$ will also depend on the situation involved and may also depend on the choice of $\boldsymbol{\Sigma}$ due to the symmetry constraint.

The separable residual models of the previous chapter are equivalent to the MVAR1 model defined by (8.2.1) and (8.2.2), with $\boldsymbol{\Omega}$ being represented by a diagonal matrix i.e. $\boldsymbol{\Omega} = \omega \boldsymbol{I}$. Hence

$$e_{(i+1)k} \quad = \quad \omega e_{ik} + \epsilon_{(i+1)k}$$

where e_{ik} represents the residual on plot i at time (or trait) k, for $k = 1, \ldots, t$. Therefore, the same spatial dependency parameter between neighbouring plots (ω) is assumed for each time (or trait). Ideally a model that allows for different spatial dependency at each time (or trait) would be more suitable and it may also be desirable to specify spatial dependencies between residuals on a plot and it's neighbouring plots at other times (or traits).

Due to the symmetry constraint a purely diagonal form for $\boldsymbol{\Omega}$ with different spatial dependency parameters for each time (or trait) cannot be fitted, unless $\boldsymbol{\Sigma}$ is also a diagonal matrix (this is equivalent to a separate analysis of each trait or each time). That is, if

$$\boldsymbol{\Omega} = \begin{bmatrix} \omega_{11} & 0 & \ldots & 0 \\ 0 & \omega_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \omega_{tt} \end{bmatrix}$$

and $\boldsymbol{\Sigma}$ is not a diagonal matrix, then the symmetry constraints force this model to revert to the separable case with $\omega_{11} = \omega_{22} = \ldots \omega_{tt}$. To see this consider

$$\begin{aligned} \boldsymbol{\Omega\Sigma} \quad &= \quad \begin{bmatrix} \omega_{11} & 0 & \ldots & 0 \\ 0 & \omega_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \omega_{tt} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1t} \\ \sigma_{12} & \sigma_{22} & \ldots & \sigma_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1t} & \sigma_{2t} & \ldots & \sigma_{tt} \end{bmatrix} \\ &= \quad \begin{bmatrix} \omega_{11}\sigma_{11} & \omega_{11}\sigma_{12} & \ldots & \omega_{11}\sigma_{1t} \\ \omega_{22}\sigma_{12} & \omega_{22}\sigma_{22} & \ldots & \omega_{22}\sigma_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{tt}\sigma_{1t} & \omega_{tt}\sigma_{2t} & \ldots & \omega_{tt}\sigma_{tt} \end{bmatrix} \end{aligned}$$

For this matrix to be symmetric each of the off-diagonals above and below the diagonal must be equal. Thus for example,

$$\omega_{11}\sigma_{12} = \omega_{22}\sigma_{12}$$

which implies

$$\omega_{11} = \omega_{22}$$

Similarly it can be shown that

$$\omega_{11} = \omega_{22} = \omega_{33} = \ldots = \omega_{tt}$$

Hence to allow for different spatial dependency parameters at each time (or trait) the model for $\mathbf{\Omega}$ will need to have elements in the off-diagonals and therefore the model must allow for spatial dependency between measurements on neighbouring plots at other times (or traits).

This means $\mathbf{\Omega}$ contains many parameters to be estimated. It also makes interpretation (in a practical sense) of $\mathbf{\Omega}$ slightly difficult. While it is reasonable to consider the diagonals of Omega as the spatial dependency parameters between neighbouring plots measured at the same time or on the same trait, the off-diagonals (which reflect the spatial dependency between measurements on neighbouring plots at different times (or between different traits)) may be more difficult to interpret. The first off diagonal gives a measure of how variable 1 at plot i impacts spatially on variable 2 on neighbouring plot (i+1).

General omega (gen$\mathbf{\Omega}$)

The most general form of $\mathbf{\Omega}$ is the fully parameterized non-symmetric matrix (hereafter referred to as gen$\mathbf{\Omega}$), given by

$$\mathbf{\Omega} = \begin{bmatrix} \omega_{11} & \omega_{12} & \ldots & \omega_{1t} \\ \omega_{21} & \omega_{22} & \ldots & \omega_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{t1} & \omega_{2t} & \ldots & \omega_{tt} \end{bmatrix}$$

This form for $\mathbf{\Omega}$ may be most suitable for multi-trait data. While this model may also be applicable to the multi-harvest situation when there are small numbers of harvests, for large numbers of harvests the number of parameters requiring estimation is likely to make fitting this model difficult. Alternative more parsimonious forms of $\mathbf{\Omega}$ may be required. Possible alternative models for $\mathbf{\Omega}$ that allow for different spatial dependencies between neighbouring plots at each time, but have less parameters than a fully parameterised non-symmetric general $\mathbf{\Omega}$ are :

Symmetric omega (symΩ)

$$\mathbf{\Omega} = \begin{bmatrix} \omega_{11} & \omega_{12} & \dots & \omega_{1t} \\ \omega_{12} & \omega_{22} & \dots & \omega_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{1t} & \omega_{2t} & \dots & \omega_{tt} \end{bmatrix}$$

Hence the model can be written as

$$\begin{aligned} e_{(i+1)1} &= \omega_{11}e_{i1} + \omega_{12}e_{i2} + \ldots + \omega_{1t}e_{it} + \epsilon_{(i+1)1} \\ e_{(i+1)2} &= \omega_{12}e_{i1} + \omega_{22}e_{i2} + \ldots + \omega_{2t}e_{it} + \epsilon_{(i+1)2} \\ &\vdots \\ e_{(i+1)t} &= \omega_{1t}e_{i1} + \omega_{2t}e_{i2} + \ldots + \omega_{tt}e_{it} + \epsilon_{(i+1)t} \end{aligned}$$

This form of $\mathbf{\Omega}$ is suitable for all valid covariance matrices $\mathbf{\Sigma}$ (corh, ar1h, ante, us). This symmetric model for $\mathbf{\Omega}$ assumes the spatial dependency between plot 1 at time 1 and plot 2 at time 2 is equal to the spatial dependency between plot 2 at time 1 and plot 1 at time 2 (and similarly for all neighbouring plots). It has been shown (in section 8.2.5) that the symmetry constraint forces the cross correlations to be equal. That is, the correlation between plot 1 time 1 and plot 2 time 2 is constrained to be the same as the correlation between plot 2 time 1 and plot 1 time 2. Hence while this model may not be as flexible as the fully general $\mathbf{\Omega}$ it may enable a reduction in the number of parameters without forcing too many extra restrictions on the covariance structure. It may be a useful model for providing starting values for the full general $\mathbf{\Omega}$ (gen$\mathbf{\Omega}$).

Banded omega (bandΩ)

$$\mathbf{\Omega} = \begin{bmatrix} \omega_{11} & \omega_{12} & 0 & \dots & 0 \\ \omega_{21} & \omega_{22} & \omega_{23} & \dots & 0 \\ 0 & \omega_{32} & \omega_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_{t(t-1)} & \omega_{tt} \end{bmatrix}$$

Hence the model can be written as

$$\begin{aligned} e_{(i+1)1} &= \omega_{11}e_{i1} + \omega_{12}e_{i2} + \epsilon_{(i+1)1} \\ e_{(i+1)2} &= \omega_{21}e_{i1} + \omega_{22}e_{i2} + \omega_{23}e_{i3} + \epsilon_{(i+1)2} \\ &\vdots \\ e_{(i+1)t} &= \omega_{t(t-1)}e_{i(t-1)} + \omega_{tt}e_{it} + +\epsilon_{(i+1)t} \end{aligned}$$

This model provides spatial dependency parameters between observations one time apart but no further apart. While this form of $\mathbf{\Omega}$ would seem appealing due to the reduced number of parameters, the symmetry constraint may make this model very restrictive.

If we consider the general form of the inverse covariance matrix for the ar1v, ar1h and ante forms for $\mathbf{\Sigma}$ we can investigate the impact of the symmetry constraint on the parameters in $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ when the above banded form for $\mathbf{\Omega}$ is assumed. The general form of the inverse covariance matrix in all three cases can be written as a banded tri-diagonal matrix as follows:

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} \phi_{11} & \phi_{12} & 0 & \dots & 0 \\ \phi_{21} & \phi_{22} & \phi_{23} & \dots & 0 \\ 0 & \phi_{32} & \phi_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_{t(t-1)} & \phi_{tt} \end{bmatrix}$$

Hence assuming the banded form of $\mathbf{\Omega}$ above, $\mathbf{\Sigma}^{-1}\mathbf{\Omega}$ is given by

$$\mathbf{\Sigma}^{-1}\mathbf{\Omega} = \begin{bmatrix} \phi_{11}\omega_{11} + \phi_{12} + \omega_{21} & \phi_{11}\omega_{12} + \phi_{12}\omega_{22} & \phi_{12}\omega_{23} & \dots \\ \phi_{21}\omega_{11} + \phi_{22}\omega_{21} & \phi_{21}\omega_{12} + \phi_{22}\omega_{22} + \phi_{23}\omega_{32} & \phi_{22}\omega_{23} + \phi_{23}\omega_{33} & \dots \\ \phi_{32}\omega_{21} & \phi_{32}\omega_{22} + \phi_{33}\omega_{32} & \dots & \dots \\ \vdots & \vdots & \vdots & \dots \end{bmatrix}$$

Now under the symmetry condition this matrix is symmetric and hence there are constraints required on these parameters. Noting that in the antedependence case (also ar1v and ar1h) $\phi_{ij} = \phi_{ji}$ we can write these constraints in the general form

$$\phi_{i(i-1)}/\phi_{i(i+1)} = \omega_{i(i-1)}/\omega_{i(i+1)}$$

for $i = 2, \dots, t-1$ and

$$\phi_{i(i+1)}(\omega_{ii} - \omega_{(i+1)(i+1)}) = \phi_{ii}\omega_{i(i+1)} - \phi_{(i+1)(i+1)}\omega_{(i+1)i}$$

for $i = 1, \dots, t-1$

In the ar1h case where $\phi_{ij} = -\phi\sigma_i\sigma_j$ for all $i \neq j$, $\phi_{11} = \sigma_1^2$, $\phi_{tt} = \sigma_t^2$, and $\phi_{ii} = (1+\phi^2)\sigma_i^2$ for $i = 2, \dots, t-1$, then these conditions reduce further to

$$\sigma_{i-1}/\sigma_{i+1} = \omega_{i(i-1)}/\omega_{i(i+1)}$$

for $i = 2, \dots, t-1$ and

$$\phi\sigma_i\sigma_{i+1}(\omega_{ii} - \omega_{(i+1)(i+1)}) = (1+\phi^2)[\sigma_i^2\omega_{i(i+1)} - \sigma_{i+1}^2\omega_{(i+1)i}]$$

for $i = 1, \dots, t-1$

In the ar1v case where $\phi_{ij} = -\phi$ for all $i \neq j$, $\phi_{11} = \phi_{tt} = 1$, and $\phi_{ii} = (1 + \phi^2)$ for $i = 2, \ldots, t-1$, then these conditions reduce further to

$$\omega_{i(i-1)} = \omega_{i(i+1)}$$

for $i = 2, \ldots, t-1$ and

$$\phi(\omega_{ii} - \omega_{(i+1)(i+1)}) = \omega_{i(i+1)} - (1 + \phi^2)\omega_{(i+1)i}$$

for $i = 1, \ldots, t-1$.

8.2.7 Multi-harvest, multi-trait models

The above models for the residual structure for multi-trait data incorporating an unstructured (us) variance matrix for $\boldsymbol{\Sigma}$ and a fully parameterized matrix for $\boldsymbol{\Omega}$ may also be suitable in the case of multi-harvest, multi-trait data where observations have been made at a small number of harvest times on a small number of traits. The number of parameters to be estimated becomes prohibitive as the number of harvest times and/or number of traits increases. An alternative, more parsimonious model may be written based on the model of Jaffrezic et al. (2003) for bivariate data collected at multiple times.

Jaffrezic et al. (2003) present a bivariate Structured Antedependence Model (SAD) model for repeated measures bivariate data that has inverse covariance matrix given by $\boldsymbol{L}^T\boldsymbol{D}^{-1}\boldsymbol{L}$ where $\boldsymbol{L}$ is a lower triangular block matrix containing the antedependence parameters for the traits and $\boldsymbol{D}$ is a block diagonal matrix containing the variance parameters.

In a multi-harvest, multi-trait case of 4 harvest times by 2 traits (with data ordered as plots within traits within harvests), a non-separable MVAR1 model, with $\boldsymbol{\Sigma}$ based on the bivariate SAD model, could be proposed for the residuals from each plot. The vector of residuals (for plot i) may be written as

$$\boldsymbol{e}_i = \begin{bmatrix} e_{i11} \\ e_{i21} \\ e_{i12} \\ e_{i22} \\ e_{i13} \\ e_{i23} \\ e_{i14} \\ e_{i24} \end{bmatrix}$$

where e_{izh} is the residual value for trait z at harvest h on plot i. If a bivariate SAD model is assumed for the residuals within each plot we may write

$$\begin{aligned} e_{i1h} &= \phi_1 e_{i1(h-1)} + \psi_1 e_{i2(h-1)} + \varepsilon_{i1h} \\ e_{i2h} &= \phi_2 e_{i2(h-1)} + \psi_2 e_{i1(h-1)} + \varepsilon_{i2h} \end{aligned}$$

with the initial conditions $e_{i11} = \varepsilon_{i11}$ and $e_{i21} = \varepsilon_{i21}$, In this model ϕ_z denotes the antedependence parameter connecting trait z at harvest h, and trait z at the previous harvest time, $h-1$, and ψ_z denotes the antedependence parameter connecting trait z at harvest time h, and trait w at the previous harvest time, $h-1$. The ε_{izh}'s are assumed to be bivariate normally distributed with mean 0 and variance for trait z at harvest time h given by σ^2_{zh}, and covariance between trait z time h and trait w at time h given by σ_{zhwh}. In this model ϕ_z and ψ_z are assumed constant for all harvests.

The inverse covariance matrix $\boldsymbol{\Sigma}^{-1}$ containing the variances and covariances between traits and times for each plot for this bivariate SAD model is given by $\boldsymbol{\Sigma}^{-1} = \boldsymbol{L}^T\boldsymbol{D}^{-1}\boldsymbol{L}$

where

$$\boldsymbol{L} = \left[\begin{array}{cc|cc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline -\phi_1 & -\psi_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -\psi_2 & -\phi_2 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & -\phi_1 & -\psi_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\psi_2 & -\phi_2 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & -\phi_1 & -\psi_1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -\psi_2 & -\phi_2 & 0 & 1 \end{array}\right]$$

and

$$\boldsymbol{D} = \left[\begin{array}{cc|cc|cc|cc} \sigma^2_{11} & \sigma_{1121} & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_{1121} & \sigma^2_{21} & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \sigma^2_{12} & \sigma_{1222} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{1222} & \sigma^2_{22} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \sigma^2_{13} & \sigma_{1323} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{1323} & \sigma^2_{23} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{14} & \sigma_{1424} \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{1424} & \sigma^2_{24} \end{array}\right]$$

In this MVAR1 model $\boldsymbol{\Omega}$ may be specified as a general fully parameterized matrix (gen$\boldsymbol{\Omega}$).

8.2.8 Estimation of parameters: Derivatives for specific cases of $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$

REML (Chapter 2) is used to estimate the parameters in the MVAR1 model. This requires maximizing the REML log-likelihood subject to the symmetry constraints. This can be achieved using Lagrange multipliers ψ_{ij}. Let

$$C = \sum_{i<j}\sum \psi_{ij} \sum_{k=1}^{t} (\omega_{ik}\sigma_{kj} - \omega_{jk}\sigma_{ik})$$

for $i < j$.

The constrained log-likelihood l_r^* needs to be maximized, where

$$\begin{aligned} l_r^* &= l_r + C \\ &= l_r + \sum_{i<j}\sum \psi_{ij} c_{ij} \\ &= l_r + \sum_{i=1}^{t-1}\sum_{j=i+1}^{t} \psi_{ij} \sum_{k=1}^{t} (\omega_{ik}\sigma_{kj} - \omega_{jk}\sigma_{ik}) \end{aligned}$$

where l_r is the REML log-likelihood (2.3.10).

Note that C can be written as

$$C = [\mathsf{vechl}(\boldsymbol{\psi})]^T [\mathsf{vechl}(\boldsymbol{\Omega\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma\Omega^T})]$$

where $\mathsf{vechl}(\boldsymbol{A})$ is the lower half vectorization of the $n \times n$ symmetric matrix $\boldsymbol{A}$. That is, the $n(n-1)/2$ column vector obtained by vectorizing the lower triangular part (without diagonals) of $\boldsymbol{A}$. $\boldsymbol{\psi}$ is the $t \times t$ symmetric matrix of Lagrange multipliers ψ_{ij} with $\psi_{ii} = 0$.

To maximize the log-likelihood, the derivatives of l_r^* are taken with respect to the variance parameters and are equated to zero. Therefore for parameters in $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ (for example σ_{rs}) this will mean calculating

$$\frac{\partial l_r^*}{\partial \sigma_{rs}} = \boldsymbol{U}(\sigma_{rs}) + \frac{\partial C}{\partial \sigma_{rs}}$$

where $\boldsymbol{U}(\sigma_{rs})$ is the score equation for σ_{rs} described in Chapter 2.

This requires the derivatives of the full MVAR1 covariance matrix $\boldsymbol{R}$, where

$$\begin{aligned} \boldsymbol{R}^{-1} &= [\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1}][\boldsymbol{I}_{nt} - \boldsymbol{F}_{1n} \otimes \boldsymbol{\Omega} + \boldsymbol{E}_{1n} \otimes \boldsymbol{\Omega}^2] \\ &= \boldsymbol{P}_1 \boldsymbol{P}_2 \end{aligned}$$

where

$$\boldsymbol{P}_1 = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1} \tag{8.2.12}$$

and

$$\boldsymbol{P}_2 = \boldsymbol{I}_{nt} - \boldsymbol{F}_{1n} \otimes \boldsymbol{\Omega} + \boldsymbol{E}_{1n} \otimes \boldsymbol{\Omega}^2 \tag{8.2.13}$$

The derivatives for the constraints C are also required. These derivatives are presented below.

Derivatives for $\boldsymbol{\Sigma}$ = us and $\boldsymbol{\Omega}$ =genΩ (fully parameterized): model for multi-trait data

The MVAR1 model referred to as USgen (in the code in Appendix D) with $\boldsymbol{\Sigma}$ an unstructured symmetric covariance matrix, and $\boldsymbol{\Omega}$ being a fully parameterized non-symmetric matrix (gen$\boldsymbol{\Omega}$) is likely to be suitable for modelling the residual covariance structure for multivariate data collected from spatially correlated plots. This model may be most suit-

able for modelling the covariance for multi-trait data where all traits may be correlated. The $t \times t$ matrices $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ may be written as follows (assuming there are n plots and t traits (or times) measured on each plot).

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1t} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1t} & \sigma_{2t} & \dots & \sigma_{tt} \end{bmatrix} = \{\sigma_{rs}\}$$

and

$$\mathbf{\Omega} = \begin{bmatrix} \omega_{11} & \omega_{12} & \dots & \omega_{1t} \\ \omega_{21} & \omega_{22} & \dots & \omega_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{t1} & \omega_{2t} & \dots & \omega_{tt} \end{bmatrix} = \{\omega_{ij}\}$$

The derivatives of $\boldsymbol{R}^{-1}$ with respect to the variance parameters $\{\sigma_{rs}\}$ and $\{\omega_{ij}\}$ (combined in vector $\kappa_k, k = 1, \dots, t(t+1)/2 + t^2$) are given by

$$\frac{\partial \boldsymbol{R}^{-1}}{\partial \kappa_k} = \frac{\partial \boldsymbol{P}_1}{\partial \kappa_k}\boldsymbol{P}_2 + \boldsymbol{P}_1\frac{\partial \boldsymbol{P}_2}{\partial \kappa_k}$$

where $\boldsymbol{P}_1$ and $\boldsymbol{P_2}$ are given in (8.2.12) and (8.2.13).

For the parameters in κ_k, with $k = 1, \dots, t(t+1)/2$ (parameters σ_{rs} from $\mathbf{\Sigma}$)

$$\begin{aligned} \frac{\partial \boldsymbol{P}_1}{\partial \sigma_{rs}} &= \frac{\partial \boldsymbol{I}_n}{\partial \sigma_{rs}} \otimes \mathbf{\Sigma}^{-1}(\boldsymbol{I}_t - \mathbf{\Omega}^2)^{-1} + \boldsymbol{I}_n \otimes \frac{\partial}{\partial \sigma_{rs}}(\mathbf{\Sigma}^{-1}(\boldsymbol{I}_t - \mathbf{\Omega}^2)^{-1}) \\ &= \boldsymbol{I}_n \otimes [\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_{rs}}(\boldsymbol{I}_t - \mathbf{\Omega}^2)^{-1} + \mathbf{\Sigma}^{-1}\frac{\partial}{\partial \sigma_{rs}}(\boldsymbol{I}_t - \mathbf{\Omega}^2)^{-1}] \\ &= \boldsymbol{I}_n \otimes \frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_{rs}}(\boldsymbol{I}_t - \mathbf{\Omega}^2)^{-1} \\ &= \boldsymbol{I}_n \otimes [-\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \sigma_{rs}}\mathbf{\Sigma}^{-1}(\boldsymbol{I}_t - \mathbf{\Omega}^2)^{-1}] \end{aligned} \tag{8.2.14}$$

where

$$\frac{\partial \mathbf{\Sigma}}{\partial \sigma_{rs}} = (\boldsymbol{z}_r\boldsymbol{z}_s^T) + (\boldsymbol{z}_r\boldsymbol{z}_s^T)^T - \boldsymbol{I}_t \odot (\boldsymbol{z}_r\boldsymbol{z}_s^T)$$

where $\boldsymbol{z}_i$, is a $t \times 1$ vector containing all zeros except for the i^{th} element (row) which equals 1 and $\boldsymbol{A} \odot \boldsymbol{B}$ gives the Hadamard (elementwise) product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

The derivative of $\boldsymbol{P}_2$ wrt σ_{rs} is given by

$$\frac{\partial \boldsymbol{P}_2}{\partial \sigma_{rs}} = 0 \times \boldsymbol{I}_{nt} = \mathbf{0}$$

For the parameters from $\boldsymbol{\Omega}$, that is in κ_k for $k = (1 + t(t+1)/2, \ldots, t^2)$

$$\begin{aligned}
\frac{\partial \boldsymbol{P}_1}{\partial \omega_{ij}} &= \boldsymbol{I}_n \otimes \frac{\partial}{\partial \omega_{ij}} (\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1}) \\
&= \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{-1} \frac{\partial}{\partial \omega_{ij}} (\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1} \\
&= \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{-1} ((\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1} \frac{\partial \boldsymbol{\Omega}^2}{\partial \omega_{ij}} (\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1}) \\
&= \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{-1} ((\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1} (\boldsymbol{\Omega} \frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}} + \frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}} \boldsymbol{\Omega}) (\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1}) \qquad (8.2.15)
\end{aligned}$$

where

$$\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}} = \boldsymbol{z}_i \boldsymbol{z}_j^T$$

where $\boldsymbol{z}_i$ is defined as above.

The derivative of $\boldsymbol{P}_2$ wrt ω_{ij} is given by

$$\begin{aligned}
\frac{\partial \boldsymbol{P}_2}{\partial \omega_{ij}} &= -\boldsymbol{F}_{1n} \otimes \frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}} + \boldsymbol{E}_{1n} \otimes \frac{\partial \boldsymbol{\Omega}^2}{\partial \omega_{ij}} \\
&= -\boldsymbol{F}_{1n} \otimes \frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}} + \boldsymbol{E}_{1n} \otimes (\boldsymbol{\Omega} \frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}} + \frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}} \boldsymbol{\Omega}) \qquad (8.2.16)
\end{aligned}$$

where $\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}}$ is defined as above.

Derivatives of constraints

The derivative of the constraints C with respect to σ_{rs} is given by

$$\begin{aligned}
\frac{\partial \boldsymbol{C}}{\partial \sigma_{rs}} &= \frac{\partial}{\partial \sigma_{rs}} ([\mathsf{vechl}(\boldsymbol{\psi})]^T [\mathsf{vechl}(\boldsymbol{\Omega}\boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma}\boldsymbol{\Omega}^{\boldsymbol{T}})]) \\
&= [\mathsf{vechl}(\boldsymbol{\psi})]^T [\mathsf{vechl}(\boldsymbol{\Omega} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{rs}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{rs}} \boldsymbol{\Omega}^T)]
\end{aligned}$$

where $\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{rs}}$ is defined as above and $\mathsf{vechl(A)}$ is the lower half vectorization of the $n \times n$ symmetric matrix $\boldsymbol{A}$. That is, the $n(n-1)/2$ column vector obtained by vectorizing the lower triangular part (without diagonals) of $\boldsymbol{A}$.

The derivative of the constraints C with respect to ω_{rs} is given by

$$\frac{\partial \boldsymbol{C}}{\partial \omega_{rs}} = [\mathsf{vechl}(\boldsymbol{\psi})]^T [\mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{rs}} \boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma} \frac{\partial \boldsymbol{\Omega}^T}{\partial \omega_{rs}})]$$

where $\frac{\partial \boldsymbol{\Omega}^T}{\partial \omega_{rs}} = \left(\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{rs}}\right)^T$ and $\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{rs}}$ is defined as above.

The derivative of the constraints C with respect to ψ_{rs} is given by

$$\begin{aligned}
\frac{\partial C}{\partial \psi_{rs}} &= [\mathsf{vechl}(\frac{\partial \boldsymbol{\psi}}{\partial \psi_{rs}})]^T [\mathsf{vechl}(\boldsymbol{\Omega}\boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma}\boldsymbol{\Omega}^T)] \\
&= [\mathsf{vechl}(\boldsymbol{z}_r \boldsymbol{z}_s^T)]^T [\mathsf{vechl}(\boldsymbol{\Omega}\boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma}\boldsymbol{\Omega}^T)]
\end{aligned}$$

The second differentials of the constraints are also required to update the Average Information matrix. Many of these are zero but the non zero differentials are as follows. The second derivative of the constraints C with respect to ψ_{rs} and ω_{uv} is given by

$$\begin{aligned}\frac{\partial C}{\partial \psi_{rs}\partial \omega_{uv}} &= \frac{\partial}{\partial \omega_{uv}}[\mathsf{vechl}(\boldsymbol{z}_r\boldsymbol{z}_s^T)]^T[\mathsf{vechl}(\boldsymbol{\Omega}\boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma}\boldsymbol{\Omega}^T)] \\ &= [\mathsf{vechl}(\boldsymbol{z}_r\boldsymbol{z}_s^T)]^T[\mathsf{vechl}(\frac{\partial \Omega}{\partial \omega_{uv}}\boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma}\frac{\partial \Omega^T}{\partial \omega_{uv}})]\end{aligned}$$

The second derivative of the constraints C with respect to ψ_{rs} and σ_{uv} is given by

$$\begin{aligned}\frac{\partial C}{\partial \psi_{rs}\partial \sigma_{uv}} &= \frac{\partial}{\partial \sigma_{uv}}[\mathsf{vechl}(\boldsymbol{z}_r\boldsymbol{z}_s^T)]^T[\mathsf{vechl}(\boldsymbol{\Omega}\boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma}\boldsymbol{\Omega}^T)] \\ &= [\mathsf{vechl}(\boldsymbol{z}_r\boldsymbol{z}_s^T)]^T[\mathsf{vechl}(\boldsymbol{\Omega}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}}\boldsymbol{\Omega}^T)]\end{aligned}$$

The second derivative of the constraints C with respect to ω_{rs} and σ_{uv} is given by

$$\begin{aligned}\frac{\partial C}{\partial \omega_{rs}\partial \sigma_{uv}} &= \frac{\partial}{\partial \sigma_{uv}}[\mathsf{vechl}(\boldsymbol{\psi})]^T[\mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{rs}}\boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma}(\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{rs}})^T)] \\ &= [\mathsf{vechl}(\boldsymbol{\psi})]^T[\mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{rs}}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}}(\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{rs}})^T)]\end{aligned}$$

where $\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{rs}}$ and $\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{rs}}$ are defined as above.

Derivatives for $\boldsymbol{\Sigma} = \mathsf{ar1h}$ and $\boldsymbol{\Omega} = \mathsf{sym}\boldsymbol{\Omega}$ (fully parameterized symmetric matrix): model for multi-harvest data

As discussed above, the MVAR1 model (referred to as ar1hsym in the code in the appendix) with $\boldsymbol{\Sigma} = \mathsf{ar1h}$ and $\boldsymbol{\Omega} = \mathsf{sym}\boldsymbol{\Omega}$ (a symmetric fully parameterized matrix) may be suitable for multi-harvest data collected at evenly spaced times. This model will not be suitable for multi-trait data. The $t \times t$ matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ may be written as follows (assuming there are n plots and t times where measurements have been made on each plot).

$$\boldsymbol{\Sigma} = \boldsymbol{D}(\boldsymbol{I}_t + \sum_{j=1}^{t-1}\phi^j\boldsymbol{F}_{jt})\boldsymbol{D}$$

where $\boldsymbol{D}$ is a diagonal matrix of standard deviations for each harvest time ($\mathsf{diag}(\sigma_{ii}^{0.5})$) for $i = 1, \ldots t$ and $\boldsymbol{F}_{jt}$ is a $t \times t$ matrix which has 'ones' on the j^{th} sub and super-diagonals and zeroes elsewhere.

$$\boldsymbol{\Omega} = \begin{bmatrix} \omega_{11} & \omega_{12} & \ldots & \omega_{1t} \\ \omega_{12} & \omega_{22} & \ldots & \omega_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{1t} & \omega_{2t} & \ldots & \omega_{tt} \end{bmatrix}$$

As the inverse variance matrix $\boldsymbol{\Sigma}^{-1}$ is a sparse banded matrix it will be used to form

the derivatives. The inverse variance matrix is given as follows

$$\begin{aligned}\boldsymbol{\Sigma}^{-1} &= \boldsymbol{D}^{-1}(\frac{1}{1-\phi^2})(\boldsymbol{I}_t + \phi^2\boldsymbol{E}_{1t} - \phi\boldsymbol{F}_{1t})\boldsymbol{D}^{-1} \\ &= \boldsymbol{D}^{-1}(\boldsymbol{\Sigma}_{ar1v}^{-1})\boldsymbol{D}^{-1}\text{say}\end{aligned}$$

As above, the derivatives of $\boldsymbol{R}^{-1}$ with respect to (wrt) the variance parameters $\{\sigma_{rr}\}$, ϕ, and $\{\omega_{ij}\}$ (combined in vector $\kappa_k, k = 1, \ldots, t(t+1)/2 + t + 1$) are given by

$$\frac{\partial \boldsymbol{R}^{-1}}{\partial \kappa_k} = \frac{\partial \boldsymbol{P}_1}{\partial \kappa_k}\boldsymbol{P}_2 + \boldsymbol{P}_1\frac{\partial \boldsymbol{P}_2}{\partial \kappa_k}$$

where

$$\boldsymbol{P}_1 = [\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1}]$$

and

$$\boldsymbol{P}_2 = [\boldsymbol{I}_{nt} - \boldsymbol{F}_{1n} \otimes \boldsymbol{\Omega} + \boldsymbol{E}_{1n} \otimes \boldsymbol{\Omega}^2]$$

For the parameters in κ_k, with $k = 1, \ldots, t$ (parameters σ_{rr} from $\boldsymbol{\Sigma}^{-1}$ in $\boldsymbol{D}$)

$$\frac{\partial \boldsymbol{P}_1}{\partial \kappa_k} = \frac{\partial \boldsymbol{P}_1}{\partial \sigma_{rr}} \quad = \quad \boldsymbol{I}_n \otimes \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_{rr}}(\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1}$$

as in (8.2.14), where

$$\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_{rr}} = \frac{\partial \boldsymbol{D}^{-1}}{\partial \sigma_{rr}}(\boldsymbol{\Sigma}_{ar1v}^{-1}\boldsymbol{D}^{-1}) + \boldsymbol{D}^{-1}\boldsymbol{\Sigma}_{ar1v}^{-1}\frac{\partial \boldsymbol{D}^{-1}}{\partial \sigma_{rr}}$$

and

$$\frac{\partial \boldsymbol{D}^{-1}}{\partial \sigma_{rr}} = -0.5(\sigma_{rr}^{-1.5})(\boldsymbol{z}_r\boldsymbol{z}_r^T)$$

The derivative of $\boldsymbol{P}_2$ wrt σ_{rr} is given by

$$\frac{\partial \boldsymbol{P}_2}{\partial \sigma_{rr}} = 0 \times \boldsymbol{I}_{nt} = \boldsymbol{0}$$

For the parameter ϕ in $\boldsymbol{\Sigma}$ (that is κ_k with $k = t + 1$) the derivatives are as follows

$$\frac{\partial \boldsymbol{P}_1}{\partial \phi} \quad = \quad \boldsymbol{I}_n \otimes \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi}(\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1}$$

where

$$\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi} = \boldsymbol{D}^{-1}\frac{\partial \boldsymbol{\Sigma}_{ar1v}^{-1}}{\partial \phi}\boldsymbol{D}^{-1}$$

and

$$\frac{\partial \boldsymbol{\Sigma}_{ar1v}^{-1}}{\partial \phi} = (1-\phi^2)^{-1}(2\phi\boldsymbol{E}_{1t} - \boldsymbol{F}_{1t}) + 2\phi(1-\phi^2)^{-2}(\boldsymbol{I}_t + \phi^2\boldsymbol{E}_{1t} - \phi\boldsymbol{F}_{1t})$$

The derivative of $\boldsymbol{P}_2$ wrt ϕ is given by

$$\frac{\partial \boldsymbol{P}_2}{\partial \phi} = 0 \times \boldsymbol{I}_{rt} = \boldsymbol{0}$$

For the parameters ω_{ij} from $\boldsymbol{\Omega}$, that is in κ_k for $k = (t+2, \ldots, t(t+1)/2 + t + 1)$

$$\frac{\partial \boldsymbol{P}_1}{\partial \kappa_k} = \frac{\partial \boldsymbol{P}_1}{\partial \omega_{ij}} \quad = \quad \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{-1}((\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1}(\boldsymbol{\Omega}\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}} + \frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}}\boldsymbol{\Omega})(\boldsymbol{I}_t - \boldsymbol{\Omega}^2)^{-1})$$

using (8.2.15) where

$$\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}} = (\boldsymbol{z}_i \boldsymbol{z}_j^T) + (\boldsymbol{z}_i \boldsymbol{z}_j^T)^T - \boldsymbol{I}_t \odot (\boldsymbol{z}_i \boldsymbol{z}_j^T)$$

where $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ are defined above.

The derivative of $\boldsymbol{P}_2$ wrt ω_{ij} is given by

$$\frac{\partial \boldsymbol{P}_2}{\partial \omega_{ij}} \quad = \quad -\boldsymbol{F}_{1n} \otimes \frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}} + \boldsymbol{E}_{1n} \otimes (\boldsymbol{\Omega}\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}} + \frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}}\boldsymbol{\Omega})$$

using (8.2.16), where $\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}}$ is given above.

Derivatives of constraints

As $\boldsymbol{\Sigma}^{-1}$ is a sparse matrix, it is simpler computationally to express the constraints $\boldsymbol{C}$ in terms of $\boldsymbol{\Sigma}^{-1}$. That is,

$$\boldsymbol{C} = [\mathsf{vechl}(\boldsymbol{\psi})]^T[\mathsf{vechl}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}) - \mathsf{vechl}(\boldsymbol{\Omega}^T\boldsymbol{\Sigma}^{-1})]$$

Therefore the derivatives of the constraints wrt the parameters $\{\sigma_{rr}\}$, ϕ and $\{\omega_{ij}\}$ can be written as

$$\frac{\partial \boldsymbol{C}}{\partial \sigma_{rr}} \quad = \quad [\mathsf{vechl}(\boldsymbol{\psi})]^T[\mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_{rr}}\boldsymbol{\Omega}) - \mathsf{vechl}(\boldsymbol{\Omega}^T\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_{rr}})]$$

$$\frac{\partial \boldsymbol{C}}{\partial \phi} \quad = \quad [\mathsf{vechl}(\boldsymbol{\psi})]^T[\mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi}\boldsymbol{\Omega}) - \mathsf{vechl}(\boldsymbol{\Omega}^T\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi})]$$

$$\frac{\partial \boldsymbol{C}}{\partial \omega_{ij}} \quad = \quad [\mathsf{vechl}(\boldsymbol{\psi})]^T[\mathsf{vechl}(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}^T}{\partial \omega_{ij}}\boldsymbol{\Sigma}^{-1})]$$

$$\frac{\partial \boldsymbol{C}}{\partial \psi_{rs}} \quad = \quad [\mathsf{vechl}(\boldsymbol{z}_r \boldsymbol{z}_s^T)]^T[\mathsf{vechl}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}) - \mathsf{vechl}(\boldsymbol{\Omega}^T\boldsymbol{\Sigma}^{-1})]$$

The second derivatives of the constraints are

$$\frac{\partial^2 \boldsymbol{C}}{\partial \sigma_{rr}^2} = [\mathsf{vechl}(\boldsymbol{\psi})]^T[\mathsf{vechl}(\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \sigma_{rr}^2}\boldsymbol{\Omega}) - \mathsf{vechl}(\boldsymbol{\Omega}^T\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \sigma_{rr}^2})]$$

where

$$\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \sigma_{rr}^2} = \frac{\partial^2 \boldsymbol{D}^{-1}}{\partial \sigma_{rr}^2}(\boldsymbol{\Sigma}_{ar1v}^{-1}\boldsymbol{D}^{-1}) + \frac{\partial \boldsymbol{D}^{-1}}{\partial \sigma_{rr}}(\boldsymbol{\Sigma}_{ar1v}^{-1}\frac{\partial \boldsymbol{D}^{-1}}{\partial \sigma_{rr}}) + \boldsymbol{D}^{-1}\boldsymbol{\Sigma}_{ar1v}^{-1}\frac{\partial^2 \boldsymbol{D}^{-1}}{\partial \sigma_{rr}^2} + \frac{\partial \boldsymbol{D}^{-1}}{\partial \sigma_{rr}}\boldsymbol{\Sigma}_{ar1v}^{-1}\frac{\partial \boldsymbol{D}^{-1}}{\partial \sigma_{rr}}$$

If $r \neq s$

$$\frac{\partial^2 \boldsymbol{C}}{\partial \sigma_{rr}\partial \sigma_{ss}} = 0$$

$$\frac{\partial^2 \boldsymbol{C}}{\partial \phi^2} = [\mathsf{vechl}(\boldsymbol{\psi})]^T[\mathsf{vechl}(\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \phi^2}\boldsymbol{\Omega}) - \mathsf{vechl}(\boldsymbol{\Omega}^T\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \phi^2})]$$

where

$$\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \phi^2} = \boldsymbol{D}^{-1}\frac{\partial^2 \boldsymbol{\Sigma}_{ar1v}^{-1}}{\partial \phi^2}\boldsymbol{D}^{-1}$$

and where

$$\begin{aligned}\frac{\partial^2 \boldsymbol{\Sigma}_{ar1v}^{-1}}{\partial \phi^2} &= 2(1-\phi^2)^{-1}\boldsymbol{E}_{1t} + 4\phi(1-\phi^2)^{-2}(2\phi\boldsymbol{E}_{1t} - \boldsymbol{F}_{1t}) + \\ &\quad (2(1-\phi^2)^{-2} + 8\phi^2(1-\phi^2)^{-3})(\boldsymbol{I}_t + \phi^2\boldsymbol{E}_{1t} - \phi\boldsymbol{F}_{1t})\end{aligned}$$

$$\frac{\partial^2 \boldsymbol{C}}{\partial \sigma_{rr}\partial \omega_{ij}} = [\mathsf{vechl}(\boldsymbol{\psi})]^T[\mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_{rr}}\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}^T}{\partial \omega_{ij}}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_{rr}})]$$

$$\frac{\partial^2 \boldsymbol{C}}{\partial \phi\partial \omega_{ij}} = [\mathsf{vechl}(\boldsymbol{\psi})]^T[\mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi}\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}^T}{\partial \omega_{ij}}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi})]$$

$$\begin{aligned}\frac{\partial^2 \boldsymbol{C}}{\partial \psi_{uv}\partial \omega_{ij}} &= [\mathsf{vechl}(\frac{\partial \boldsymbol{\psi}}{\partial \psi_{uv}})]^T[\mathsf{vechl}(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}^T}{\partial \omega_{ii}}\boldsymbol{\Sigma}^{-1})] \\ &= [\mathsf{vechl}(\boldsymbol{z}_u\boldsymbol{z}_v^T)]^T[\mathsf{vechl}(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}^T}{\partial \omega_{ii}}\boldsymbol{\Sigma}^{-1})]\end{aligned}$$

$$\frac{\partial^2 \boldsymbol{C}}{\partial \psi_{uv}\partial \sigma_{rr}} = [\mathsf{vechl}(\boldsymbol{z}_u\boldsymbol{z}_v^T)]^T[\mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_{rr}}\boldsymbol{\Omega}) - \mathsf{vechl}(\boldsymbol{\Omega}^T\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_{rr}})]$$

$$\frac{\partial^2 \boldsymbol{C}}{\partial \psi_{uv} \partial \phi} = [\mathsf{vechl}(\boldsymbol{z}_u \boldsymbol{z}_v^T)]^T [\mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi} \boldsymbol{\Omega}) - \mathsf{vechl}(\boldsymbol{\Omega}^T \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \phi})]$$

$$\frac{\partial^2 \boldsymbol{C}}{\partial \phi \partial \sigma_{rr}} = [\mathsf{vechl}(\boldsymbol{\psi})]^T [\mathsf{vechl}(\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \phi \partial \sigma_{rr}} \boldsymbol{\Omega}) - \mathsf{vechl}(\boldsymbol{\Omega}^T \frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \phi \partial \sigma_{rr}})]$$

where

$$\frac{\partial^2 \boldsymbol{\Sigma}^{-1}}{\partial \phi \partial \sigma_{rr}} = \frac{\partial \boldsymbol{D}^{-1}}{\partial \sigma_{rr}} \frac{\partial \boldsymbol{\Sigma}_{ar1v}^{-1}}{\partial \phi} \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1} \frac{\partial \boldsymbol{\Sigma}_{ar1v}^{-1}}{\partial \phi} \frac{\partial \boldsymbol{D}^{-1}}{\partial \sigma_{rr}}$$

8.3 Application of MVAR1 model

In this section the non-separable multivariate autoregressive (MVAR1) residual variance model is applied in analyses of the multi-harvest lucerne yield data from Terry Hie Hie, and also in two separate bivariate analyses of multi-trait data (the yield and persistence at a common time) from two sites (Terry Hie Hie and Leadville). The results from each analysis are compared to an analysis using a separable residual model fitted in ASReml (as described in the previous chapter). The non-separable MVAR1 model has not yet been implemented into ASReml so code for fitting these models (using REML) has been written in R (R Development Core Team, 2012). The code is presented in Appendix D.

To investigate the different residual models a similar modelling approach to that applied in the previous chapter has been taken, in that a simple genetic model has been assumed while the residual models are investigated. Once the best residual model has been decided upon more complex genetic models may be investigated. In this section focus has only been on the residual models under a simple genetic model.

The convergence of these models is very dependent on providing good starting values for the parameters. In some cases the number of starting values required is very large. In the following analyses starting values have been found for $\boldsymbol{\Omega}$ by fitting a separate (id(Column).ar1(Row)) spatial residual model for each harvest (or trait) in ASReml to get a separate spatial correlation parameter estimate for each time (or trait). These values have been taken as the starting values for the diagonal elements of the spatial dependency matrix $\boldsymbol{\Omega}$, while the off diagonal elements have been set to a small value (0.001). The starting values for the covariance matrix $\boldsymbol{\Sigma}$ have been taken from the temporal (or multi-trait) covariance structure estimated from fitting a comparable separable residual model in ASReml. For example, for the multi-trait examples the starting values have been taken from an analysis fitting a separable (us(Trait).id(Column).ar1(Row)) model.

8.3.1 Multi-trait examples

Yield and persistence at Terry Hie Hie

To illustrate the MVAR1 models presented above, and their suitability for modelling the residual structure for multi-trait data, a bivariate analysis of the lucerne yield (at harvest time 22/6/2004) and persistence (at harvest time 30/6/2004) from Terry Hie Hie was performed.

Table 8.1 presents the log-likelihoods and parameter estimates for the models fitted and presents a comparison between the separable us(Trait).id(Column).ar1(Row) model fitted in ASReml, the MVAR1 code (referred to as USI in Table 8.1 and in the code in Appendix D) with $\boldsymbol{\Sigma}$ an unstructured matrix and $\boldsymbol{\Omega} = \phi_r \boldsymbol{I}$ (which is equivalent to the separable model), and the non-separable MVAR1 model with an unstructured (us) 2×2 matrix for $\boldsymbol{\Sigma}$ and a non-symmetric fully parameterized matrix (2×2) for $\boldsymbol{\Omega}$ (referred to as USgen in Table 8.1 and in the code in Appendix D). For further comparison purposes the results from fitting a non-separable MVAR1 model with an unstructured (us) form for $\boldsymbol{\Sigma}$ and a symmetric matrix (sym$\boldsymbol{\Omega}$) for $\boldsymbol{\Omega}$ (referred to as USsym in Table 8.1 and in the code in Appendix D) are also presented.

The spatial correlation parameters from the individual analyses of the two traits at the individual harvest times performed in Chapter 3 (including a random column effect for yield), were 0.37 for yield and 0 for persistence (see Tables 3.2 and 3.4).

The modelling process commenced with a very simple genetic model (Trait:ID + Trait:Rep) and fitted a separate residual structure for each trait, assuming a first order autoregressive spatial correlation structure in the row direction and independence in the column direction. The estimated spatial correlation parameter in the row direction for persistence was 0.021 and for yield was 0.628. These parameter estimates provided starting values for the diagonal elements of $\boldsymbol{\Omega}$ for the non-separable models. Note the large difference between spatial correlation parameters for the two traits (this is larger than the difference between the spatial parameters obtained from the analyses in Chapter 3), as a random column effect for yield (which was fitted in Chapter 3) has not yet been included. Comparing the log-likelihood value for the non-separable residual model with fully parameterized $\boldsymbol{\Omega}$ matrix (USgen) ($LL = 104.719$) to that of the separable model ($LL = 90.097$), it can be seen that the non-separable model provides a significantly better fit (REMLRT = 29.244 on 3 df, $P < 0.001$). The values estimated for $\boldsymbol{\Omega}$ are quite different for the diagonal elements ($\omega_{11} = 0.059$, the spatial dependency between neighbouring plots for persistence and $\omega_{22} = 0.646$, the spatial dependency between neighbouring plots for yield). It is also clear that the off diagonal elements are quite different from each other with ω_{21} (the spatial dependency parameter between yield on the first plot and persistence on the neighbouring plot) being larger than ω_{12} (the spatial dependency between persistence on the first plot and yield on the neighbouring plot). This provision for different off diagonal elements in $\boldsymbol{\Omega}$ provides a better model fit than the assumption of a symmetric $\boldsymbol{\Omega}$ as shown by the REMLRT comparison of the USsym and USgen models (REMLRT = 9.334 on 1df, $P = 0.002$).

The next models fitted included a separate genetic variance for each trait and a separate Rep effect for each Trait (section 2 in Table 8.1). Once again, the comparison between non-separable and separable residual models demonstrate that the non-separable model is significantly better (log-likelihood for USgen = 138.622 and log-likelihood for separable model = 121.981, with REMLRT= 33.282 on 3 df, $P < 0.001$).

The final set of models included the random column effect for yield that was significant in the previous analyses in Chapter 3. The inclusion of this spatial effect had the effect of reducing the spatial correlation estimate in the row direction for yield and hence the spatial correlation between the two traits to be more similar. It is interesting to note that the difference between log-likelihoods for the non-separable model USgen ($LL = 141.608$) and the separable model ($LL = 138.955$) is now less and no longer significant (REMLRT= 5.306 on 3 df, $P = 0.15$).

The parameter estimates from the final non-separable model USgen for $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ are given in Table 8.1. The values estimated for the diagonal elements of $\mathbf{\Omega}$ are quite different, with $\omega_{11} = 0.141$ (the spatial dependency between neighbouring plots for persistence) and $\omega_{22} = 0.427$ (the spatial dependency between neighbouring plots for yield). The off diagonal elements also differ from each other with $\omega_{21} = -0.192$ (the spatial dependency between yield on the first plot and persistence on the neighbouring plot) being larger in magnitude than $\omega_{12} = -0.066$ (the spatial dependency between persistence on the first plot and yield on the neighbouring plot). The symmetric ω model USsym did not fit as well as the USgen model (REMLRT = 4.732 on 1 df, $P = 0.03$).

Yield and persistence at Leadville

To further investigate the suitability of the MVAR1 model for modelling the residual covariance structure in a bivariate situation, an analysis was also performed on the lucerne yield and persistence data at harvest time 25/11/04, from Leadville. Table 8.2 gives an overview of the different models fitted to the Leadville data.

The initial set of models (1) in Table 8.2 fitted a simple genetic plus randomisation model with a variance component estimated for Trait:ID and Trait:Rep. The non-separable models (USgen and USsym) provided a better fit to the data than the separable model (REMLRT for USgen vs separable = 15.702 on 3 df, P=0.001, and REMLRT for USsym vs separable = 14.584 on 2 df, P<0.001). The second set of models (2) also included additional random and fixed effect spatial terms identified as being important in the initial spatial analyses of each separate trait in earlier chapters. The non-separable residual models are a significant improvement on the separable model (REMLRT for USgen vs separable = 15.164 on 3 df, P=0.002, and REMLRT for USsym vs separable = 12.388 on 2 df, P=0.002).

The separable residual model in model 2 estimates the common spatial correlation parameter as 0.5 while the non-separable residual model estimates the spatial dependency parameter for persistence (ω_{11}) as 0.240 and the spatial dependency parameter for yield (ω_{22}) as 0.668, which are quite different. It can be seen that the change in residual model has had an effect on the genetic variance estimates with the persistence genetic variance

Table 8.1: Summary of models fitted for bivariate analysis of yield and persistence at Terry Hie Hie using separable and non-separable MVAR1 residual models. Residual log-likelihoods (l) are presented for each model, as are variance components for the random terms in the models and parameter estimates for $\mathbf{\Sigma}(\sigma_{ij})$, ϕ_r and $\mathbf{\Omega}(\omega_{ij})$. The traits are ordered persistence then yield.

	Separable residual model us(Harvest):id(Col):ar1(Row) ASReml	Non separable MVAR1 residual model			
			USI	USsym	USgen
1.	$l = 90.097$		$l = 90.097$	$l = 100.052$	$l = 104.719$
Trait:ID	0.132		0.132	0.164	0.166
Trait:Rep	0.087		0.087	0.065	0.058
$\hat{\sigma}_{11} = \hat{\sigma}_{pp}$	0.087		0.087	0.072	0.072
$\hat{\sigma}_{12} = \hat{\sigma}_{yp}$	0.017		0.017	-0.013	-0.002
$\hat{\sigma}_{22} = \hat{\sigma}_{yy}$	0.249		0.249	0.275	0.282
Spatial param	$\hat{\phi}_r = 0.369$	$\hat{\omega}_{11}$	0.369	0.004	0.059
		$\hat{\omega}_{12}$	0	-0.041	-0.091
		$\hat{\omega}_{21}$	0	-0.041	-0.342
		$\hat{\omega}_{22}$	0.369	0.641	0.646
2.	$l = 121.981$		$l = 121.981$	$l = 133.448$	$l = 138.622$
Trait1(p):ID	0.014		0.014	0.011	0.012
Trait2(y):ID	0.326		0.326	0.341	0.345
Trait1:Rep	0.001		0.001	0.004	0.003
Trait2:Rep	0.177		0.177	0.172	0.172
$\hat{\sigma}_{11}$	0.097		0.097	0.081	0.077
$\hat{\sigma}_{12}$	0.018		0.018	-0.023	-0.006
$\hat{\sigma}_{22}$	0.215		0.215	0.265	0.268
Spatial param	$\hat{\phi}_r = 0.421$	$\hat{\omega}_{11}$	0.421	0.091	0.128
		$\hat{\omega}_{12}$	0	-0.073	-0.104
		$\hat{\omega}_{21}$	0	-0.073	-0.317
		$\hat{\omega}_{22}$	0.421	0.673	0.655
3.	$l = 138.955$		$l = 138.955$	$l = 139.242$	$l = 141.608$
Trait1(p):ID	0.013		0.013	0.014	0.013
Trait2(y):ID	0.328		0.328	0.326	0.335
Trait1:Rep	0.003		0.003	0.003	0.003
Trait2:Rep	0.092		0.092	0.088	0.112
Trait2:Col	0.192		0.192	0.195	0.143
$\hat{\sigma}_{11}$	0.079		0.079	0.076	0.076
$\hat{\sigma}_{12}$	0.021		0.021	0.018	0.012
$\hat{\sigma}_{22}$	0.151		0.151	0.157	0.168
Spatial param	$\hat{\phi}_r = 0.209$	$\hat{\omega}_{11}$	0.209	0.167	0.141
		$\hat{\omega}_{12}$	0	0.025	-0.066
		$\hat{\omega}_{21}$	0	0.025	-0.192
		$\hat{\omega}_{22}$	0.209	0.277	0.427

changing from 0.014 to 0.006 and the genetic variance estimate for yield changing from 0.026 to 0.044.

Table 8.2: Summary of models fitted for bivariate analysis of yield and persistence at Leadville using separable and non-separable MVAR1 residual models. Residual log-likelihoods (l) are presented for each model, as are variance components for the random terms, P-values for the fixed effect terms, and parameter estimates for $\mathbf{\Sigma}(\sigma_{ij})$, ϕ_r and $\mathbf{\Omega}(\omega_{ij})$. The traits are ordered persistence then yield.

	Separable residual model us(Harvest):id(Col):ar1(Row) ASReml	Non separable MVAR1 residual model			
			USI	USsym	USgen
1.	$l = 101.880$		$l = 101.880$	$l = 109.172$	$l = 109.731$
Trait:ID	0.018		0.018	0.015	0.015
Trait:Rep	0.231		0.231	0.176	0.174
$\hat{\sigma}_{11} = \hat{\sigma}_{pp}$	0.127		0.127	0.094	0.094
$\hat{\sigma}_{12} = \hat{\sigma}_{yp}$	0.033		0.033	0.034	0.027
$\hat{\sigma}_{22} = \hat{\sigma}_{yy}$	0.533		0.533	0.717	0.723
Spatial param	$\hat{\phi}_r = 0.581$	$\hat{\omega}_{11}$	0.581	0.325	0.320
		$\hat{\omega}_{12}$	0	0.021	0.003
		$\hat{\omega}_{21}$	0	0.021	-0.095
		$\hat{\omega}_{22}$	0.581	0.721	0.731
2.	$l = 107.774$		$l = 107.774$	$l = 113.968$	$l = 115.356$
Trait1(p):ID	0.014		0.014	0.006	0.006
Trait2(y):ID	0.026		0.026	0.042	0.044
Trait1:Rep	0.038		0.038	0.037	0.038
Trait2:Rep	0.476		0.476	0.453	0.458
Trait1:Range	0.008		0.008	0.013	0.012
Trait2:lin(Row)	P<0.001		P<0.001	P<0.001	P<0.001
$\hat{\sigma}_{11}$	0.111		0.111	0.091	0.091
$\hat{\sigma}_{12}$	0.023		0.023	0.011	0.003
$\hat{\sigma}_{22}$	0.427		0.427	0.520	0.523
Spatial param	$\hat{\phi}_r = 0.500$	$\hat{\omega}_{11}$	0.500	0.238	0.240
		$\hat{\omega}_{12}$	0	0.010	-0.028
		$\hat{\omega}_{21}$	0	0.010	-0.174
		$\hat{\omega}_{22}$	0.500	0.661	0.668

8.3.2 Multi-harvest example

An analysis of the complete set of multi-harvest yield data from Terry Hie Hie (10 harvests) was performed. As there were so many parameters required in both $\mathbf{\Omega}$ and $\mathbf{\Sigma}$, more parsimonious structures than the USgen and USsym (used above) were required. A summary of the models fitted to this multi-harvest data is presented in Table 8.3.

The first non-separable MVAR1 model to be fitted was a model with $\mathbf{\Sigma}$ as a heterogeneous autoregressive structure (ar1h) and $\mathbf{\Omega}$ as a symmetric fully parameterized matrix (sym$\mathbf{\Omega}$), referred to as ar1hsym (in Table 8.3 and in the code in Appendix D). In this

case $\boldsymbol{\Sigma}$ required 11 parameters to be estimated and $\boldsymbol{\Omega}$ required 55 parameters. In order to obtain suitable starting values for these structures, initial models were fitted to a subset of the harvests (first 4 harvests, then first 6 harvests, then first 8 harvests etc.) with the initial starting values for the 4 harvest problem obtained from fitting a separable ar1h(Harvest).id(Column).ar1(Row) residual model for the starting values for Σ and a separate residual structure for each harvest to get the starting values for the diagonal elements of $\boldsymbol{\Omega}$. This illustrates the difficulty in fitting these models for extensive multi-harvest data.

The ar1hsym model for 10 harvests (with simple genetic/ blocking model of Trait:ID + Trait:Rep) converged with a log-likelihood of 747.2687. The comparable separable model fitted in ASReml (with residual model ar1h(Harvest).id(Column).ar1(Row)) and genetic / blocking model as above, converged with a log-likelihood of 706.855). There is a difference of 54 parameters between these two models. A REMLRT on this result shows the non-separable (MVAR1) model to be a significant improvement in fit than the separable model (REMLRT = 80.829 on 54 df, $P = 0.011$).

The resulting parameter estimates for $\boldsymbol{\Sigma}$ in the ar1hsym model were as follows: diagonal elements (variances) for harvests 1 to 10 equal to 0.330, 0.316, 0.168, 0.214, 0.196, 0.671, 0.132, 0.177, 0.140, 0.606 and the autoregressive parameter is 0.579.

The REML parameter estimates for symmetric $\boldsymbol{\Omega}$ in the ar1hsym model were as follows (for brevity only the upper half of the symmetric matrix is presented):

$$\hat{\boldsymbol{\Omega}} = \begin{bmatrix} 0.484 & 0.124 & 0.013 & -0.074 & -0.108 & 0.098 & 0.029 & 0.030 & -0.002 & -0.067 \\ & 0.419 & 0.012 & -0.015 & 0.125 & -0.054 & -0.033 & -0.018 & -0.007 & -0.054 \\ & & 0.374 & 0.220 & -0.103 & -0.006 & -0.012 & -0.058 & -0.053 & -0.012 \\ & & & 0.415 & 0.059 & -0.030 & -0.040 & -0.072 & -0.044 & -0.016 \\ & & & & 0.497 & -0.032 & -0.048 & -0.015 & 0.004 & -0.043 \\ & & & & & 0.395 & 0.097 & 0.028 & -0.018 & -0.107 \\ & & & & & & 0.084 & 0.087 & 0.011 & -0.002 \\ & & & & & & & 0.140 & 0.095 & 0.057 \\ & & & & & & & & 0.108 & 0.144 \\ & & & & & & & & & 0.573 \end{bmatrix}$$

The second non-separable MVAR1 model to be fitted also assumed an ar1h form for $\boldsymbol{\Sigma}$, but assumed a general fully parameterised form for $\boldsymbol{\Omega}$. This model is referred to as ar1hgen (in Table 8.3 and in the code in the appendix). This model required a large number of parameters (111) to be estimated. The parameter estimates from the previous model were used as starting values for this model. The model converged with a log-likelihood of 813.386. Despite there being a difference of 45 parameters between this model and the ar1hsym model the improvement in log-likelihood was highly significant (REMLRT = 132.234 on 45 df, P$<$ 0.001).

The parameter estimates for $\boldsymbol{\Sigma}$ in the ar1gen model were as follows: diagonal elements (variances) for harvests 1 to 10 = (0.285, 0.345, 0.186, 0.250, 0.311, 0.779, 0.170, 0.199, 0.179, 0.961) and the autoregressive parameter = 0.610.

The REML parameter estimates for the non-symmetric $\boldsymbol{\Omega}$ in the ar1gen model were as follows:

$$\hat{\boldsymbol{\Omega}} = \begin{bmatrix} 0.337 & 0.123 & -0.034 & 0.031 & -0.006 & 0.002 & -0.180 & -0.044 & 0.119 & -0.040 \\ 0.046 & 0.483 & -0.148 & 0.139 & 0.101 & -0.078 & -0.099 & -0.018 & -0.066 & 0.006 \\ 0.016 & 0.017 & 0.224 & 0.261 & 0.045 & -0.076 & -0.058 & 0.011 & -0.115 & 0.022 \\ -0.059 & 0.010 & 0.217 & 0.330 & 0.260 & -0.127 & -0.055 & -0.030 & -0.188 & 0.014 \\ -0.094 & 0.064 & -0.047 & -0.144 & 0.921 & -0.038 & 0.000 & -0.090 & -0.192 & 0.042 \\ -0.054 & -0.033 & -0.032 & -0.515 & 0.558 & 0.291 & 0.470 & -0.095 & -0.131 & -0.008 \\ -0.081 & -0.004 & 0.008 & -0.196 & 0.054 & 0.119 & 0.323 & 0.044 & -0.108 & 0.027 \\ -0.039 & -0.023 & 0.047 & -0.174 & -0.011 & 0.057 & 0.010 & 0.360 & 0.017 & 0.042 \\ 0.037 & -0.040 & 0.026 & -0.171 & -0.014 & 0.038 & -0.098 & 0.145 & 0.057 & 0.203 \\ -0.091 & -0.013 & 0.143 & -0.315 & 0.138 & -0.052 & -0.053 & -0.236 & 0.429 & 0.721 \end{bmatrix}$$

Table 8.3: Summary of models fitted for multi-harvest analysis of yield at Terry Hie Hie (10 harvests) using separable and non-separable MVAR1 residual models. Number of parameters (npar and totpar) and residual log-likelihoods (l) are presented for each model. REMLRT results comparing models with the previous model are also presented

Model	$\boldsymbol{\Sigma}$	npar	$\boldsymbol{\Omega}$	npar	totpar	l	REMLRT
Separable							
ar1h(Harvest):id(Col):ar1(Row)					12	706.855	
Non Separable MVAR1							
ar1hsym	ar1h	11	sym	55	66	747.269	P=0.011
ar1hgen	ar1h	11	gen	100	111	813.386	P<0.001

In an attempt to reduce the number of parameters required to be estimated, a model with $\boldsymbol{\Sigma}$ = ante and $\boldsymbol{\Omega}$ = band$\boldsymbol{\Omega}$ (referred to as anteband in the code in the appendix) was fitted to the multi-harvest data. Despite numerous attempts with different starting values and fitting to subsets of smaller number of harvests this model would not achieve convergence.

8.4 Summary

In this chapter models have been developed for modelling the residual covariance structure in multi-harvest and multi-trait data, that appear to be a significant first step extension to the separable residual models of the previous chapter. The MVAR1 models introduced here are suitable for modelling the spatio-temporal (or multi-trait) residual covariance when there is spatial correlation present in one direction. Extensions are required to accommodate spatial correlation in two directions. The MVAR1 models developed in this chapter differ from other previously published applications of multivariate autoregressive models in that they are suitable for modelling multi-variate data on spatially correlated,

evenly spaced plots (due to a symmetry constraint). They are also implemented in a linear mixed model with estimation using REML which differs from other published examples. Prior to this thesis MVAR1 models had not been used to model the spatial and temporal covariance structure in multi-harvest variety selection data in perennial crops.

In the investigations into structures for $\mathbf{\Omega}$ it is not obvious whether more parsimonious structures for multi-harvest data are available or suitable. Forms such as banded $\mathbf{\Omega}$ were investigated and while they appeared to be suitable the models would not converge when applied to the example data sets. It is possible that the restrictions imposed by the symmetry constraint make these models too restrictive and unable to be fitted in general. This could be an area of further research. The proposed general and symmetric forms for $\mathbf{\Omega}$ do require a large number of parameters for a large number of harvests however they were able to be fitted to a 10 harvest example. With careful selection of starting values, these models may be suitable for even larger numbers of harvests. Models for $\mathbf{\Sigma}$ have been proposed for multi-harvest, multi-trait examples but there may be difficulties fitting these models to large numbers of traits and times if more parsimonious structures are not available for $\mathbf{\Omega}$. In the analysis of the multi-harvest data from Terry Hie Hie (with 10 harvests) the non-separable MVAR1 models provided a significantly better fit to the data (based on REMLRT) than the comparable separable model even despite the large number of parameters that were required to be estimated. The MVAR1 model is desirable from a practical viewpoint as it allows the provision for different spatial correlation parameters for each of the harvest times rather than assuming constant spatial correlation parameters for all times. Similar improvements in model fit were seen for the MVAR1 model over the separable model in the bivariate analysis examples. However it may be the case that if the spatial correlation is not greatly different between the two traits the separable model may be similar in goodness of fit to the non-separable MVAR1 model. It is clear that there are instances where the MVAR1 model will outperform the separable model for bivariate data especially if the spatial correlation is substantially different between the two traits. The model also makes more sense from a practical viewpoint by allowing for different spatial correlation parameters for the different traits.

The MVAR1 models provide a significant improvement in model fit than the separable residual models in both multi-harvest and multi-trait examples where there is spatial correlation in one direction. Extensions are required to extend these non-separable models to the two dimensional spatial case. This will be the focus of the next chapter.

Chapter 9

Non-separable residual models: Multivariate AR1 in 2 spatial dimensions

9.1 Introduction

In the previous chapter it was shown that the MVAR1 model may be used to model the residual spatio-temporal correlation for multi-harvest or multi-trait data on plots that are spatially correlated in one direction (e.g. row direction). To extend this to two directions (row and column) in a lattice is not as straightforward as the univariate case, where the spatial correlation between plots in 2 directions may be modelled using a separable residual model $\Sigma_c \otimes \Sigma_r$ (where Σ_c and Σ_r are spatial correlation matrices (typically ar1) in the column and row directions respectively). Simply taking the Kronecker product of two multivariate autoregressive processes ($\mathsf{MVAR1}_c \otimes \mathsf{MVAR1}_r$) will not suffice. This can be seen by considering the individual MVAR1 covariance matrices which are of dimension $tc \times tc$ and $tr \times tr$ respectively (where t is the number of harvest times, r is the number of rows, c is the number of columns). The full covariance matrix of the Kronecker product of these two matrices would be of dimension $t^2rc \times t^2rc$, when what is required is a covariance matrix of dimension $trc \times trc$. A full non-separable model for time by row by column covariance is required. In this chapter the Multivariate Conditional Autoregressive (MCAR) model is used as a vehicle to define a suitable non-separable model.

While the MCAR model has been used in the literature to model spatio-temporal data there is no known application or theory developed for the model to be suitable to fit a MVAR1 structure to data in both row and column directions. Firstly the univariate case is re-visited and the links between the univariate Conditional Autoregressive (CAR) model in 2 spatial directions and the separable ar1(Column).ar1(Row) model are established. The models are subsequently extended to the multivariate case which may be suitable for multi-harvest and/or multi-trait data.

9.2 Link between CAR model and ar1(Column).ar1(Row) model

In Chapter 1 it was shown that a CAR model can be defined to give the same covariance structure as an ar1 process in a row of r plots. This result was obtained using the ideas of Cressie (1993) who shows that any Gaussian distribution on a finite set of sites ($\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$) can be expressed as a CAR model. That is, by taking

$$\boldsymbol{M} = \text{diag}\,(\boldsymbol{\Sigma}) = \text{diag}\left(\boldsymbol{\Sigma}^{-1}\right)^{-1}$$

where $\boldsymbol{\Sigma}$ is the variance covariance matrix for the autoregressive process of order 1 ar1, and

$$\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{M}\boldsymbol{\Sigma}^{-1}$$

results in

$$\boldsymbol{\Sigma} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{M}$$

and

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{M}^{-1}(\boldsymbol{I} - \boldsymbol{B})$$

Hence, taking

$$\begin{aligned} \boldsymbol{M} &= \left[\frac{1}{\sigma^2}\text{diag}\left(\frac{1}{1-\phi^2}, \frac{1+\phi^2}{1-\phi^2}, \frac{1+\phi^2}{1-\phi^2}, \ldots, \frac{1}{1-\phi^2}\right)\right]^{-1} \\ &= \sigma^2\text{diag}\left(1-\phi^2, \frac{1-\phi^2}{1+\phi^2}, \frac{1-\phi^2}{1+\phi^2}, \ldots, 1-\phi^2\right) \end{aligned}$$

and

$$\boldsymbol{B} = \begin{bmatrix} 0 & \phi & 0 & \ldots & 0 \\ \frac{\phi}{1+\phi^2} & 0 & \frac{\phi}{1+\phi^2} & \ldots & 0 \\ 0 & \frac{\phi}{1+\phi^2} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \phi & 0 \end{bmatrix}$$

a CAR model, which has the same covariance structure as the ar1 process, can be defined in terms of the following conditional distributions

$$\begin{aligned} \text{E}\,(y_i|y_{-i}) &= \mu_i + \sum_{j=1}^{r} b_{ij}(y_j - \mu_j) \\ \text{var}\,(y_i|y_{-i}) &= \sigma_i^2 \end{aligned}$$

for $i, j = 1, \ldots, r$, where y_{-i} indicates all y_j such that $j \neq i$, and where the conditional variances (σ_i^2), are the diagonal elements of $\boldsymbol{M}$, and the spatial dependency parameters (b_{ij}) between neighbouring plots i and j, are given by the elements of $\boldsymbol{B}$.

It can be seen in this example of r plots in a row, the end plots ($i = 1$ and $i = r$) have different conditional variances (first and last diagonal elements of $\boldsymbol{M}$) and different spatial dependency parameters (b_{12} and $b_{r(r-1)}$) than the internal ($i = 2, \ldots, r-1$) plots. These end plots have only 1 neighbour while the internal plots have 2 neighbours. This is an important point that will carry over to the 2 dimensional lattice case.

Now consider a 2 dimensional lattice of r rows and c columns. It is desirable to define a 2 dimensional CAR model on this lattice that has an equivalent covariance structure to the separable ar1(Column).ar1(Row) structure. The inverse covariance matrix for the ar1(Column).ar1(Row) process (where data is ordered as rows within columns) is given by

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2(1-\phi_r^2)(1-\phi_c^2)} \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{A}_2 & 0 & \ldots & 0 \\ \boldsymbol{A}_2 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \ldots & 0 \\ 0 & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \boldsymbol{A}_2 & \boldsymbol{A}_0 \end{bmatrix}$$

where $\boldsymbol{A}_0$, $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are $r \times r$ matrices given by

$$\begin{aligned} \boldsymbol{A}_0 &= (\boldsymbol{I}_r + \phi_r^2 \boldsymbol{E}_{1r} - \phi_r \boldsymbol{F}_{1r}) \\ \boldsymbol{A}_1 &= (1+\phi_c^2)(\boldsymbol{I}_r + \phi_r^2 \boldsymbol{E}_{1r} - \phi_r \boldsymbol{F}_{1r}) \\ \boldsymbol{A}_2 &= -\phi_c(\boldsymbol{I}_r + \phi_r^2 \boldsymbol{E}_{1r} - \phi_r \boldsymbol{F}_{1r}) \end{aligned}$$

where $\boldsymbol{F}_{1r}$ and $\boldsymbol{E}_{1r}$ are defined previously.

Following the approach above, taking $\boldsymbol{M} = \text{diag}\left(\boldsymbol{\Sigma}^{-1}\right)^{-1}$

$$\boldsymbol{M} = \sigma^2(1-\phi_r^2)(1-\phi_c^2) \begin{bmatrix} \boldsymbol{M}_1 & 0 & 0 & \ldots & 0 \\ 0 & \boldsymbol{M}_2 & 0 & \ldots & 0 \\ 0 & 0 & \boldsymbol{M}_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & \boldsymbol{M}_1 \end{bmatrix}$$

where

$$\boldsymbol{M}_1 = \text{diag}\left(1, \frac{1}{1+\phi_r^2}, \ldots, \frac{1}{1+\phi_r^2}, 1\right)$$

and

$$\boldsymbol{M}_2 = \text{diag}\left(\frac{1}{1+\phi_c^2}, \frac{1}{(1+\phi_c^2)(1+\phi_r^2)}, \ldots, \frac{1}{(1+\phi_c^2)(1+\phi_r^2)}, \frac{1}{1+\phi_c^2}\right)$$

and also taking $\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{M}\boldsymbol{\Sigma}^{-1}$ results in

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 & 0 & \dots & 0 \\ \boldsymbol{B}_3 & \boldsymbol{B}_1 & \boldsymbol{B}_3 & \dots & 0 \\ 0 & \boldsymbol{B}_3 & \boldsymbol{B}_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{B}_2 & \boldsymbol{B}_1 \end{bmatrix}$$

where $\boldsymbol{B}$ is a $rc \times rc$ matrix of c blocks of $\boldsymbol{B}_1$ down the diagonal, with

$$\boldsymbol{B}_1 = \begin{bmatrix} 0 & \phi_r & 0 & \dots & 0 \\ \frac{\phi_r}{1+\phi_r^2} & 0 & \frac{\phi_r}{1+\phi_r^2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_r & 0 \end{bmatrix}$$

and

$$\boldsymbol{B}_2 = \begin{bmatrix} \phi_c & -\phi_r\phi_c & 0 & \dots & 0 \\ \frac{-\phi_r\phi_c}{1+\phi_r^2} & \phi_c & \frac{-\phi_r\phi_c}{1+\phi_r^2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -\phi_r\phi_c & \phi_c \end{bmatrix}$$

and

$$\boldsymbol{B}_3 = \begin{bmatrix} \frac{\phi_c}{1+\phi_c^2} & \frac{-\phi_r\phi_c}{1+\phi_c^2} & 0 & \dots & 0 \\ \frac{-\phi_r\phi_c}{(1+\phi_c^2)(1+\phi_r^2)} & \frac{\phi_c}{1+\phi_c^2} & \frac{-\phi_r\phi_c}{(1+\phi_c^2)(1+\phi_r^2)} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{-\phi_r\phi_c}{1+\phi_c^2} & \frac{\phi_c}{1+\phi_c^2} \end{bmatrix}$$

Then

$$\Sigma_{2dCAR}^{-1} = \boldsymbol{M}^{-1}(\boldsymbol{I}_{rc} - \boldsymbol{B}) \tag{9.2.1}$$

has the same covariance structure as the ar1(Column).ar1(Row) process.

The conditional variances and spatial dependency parameters differ between corner plots of the lattice (those with only 3 neighbours), plots on the edge of the lattice (first and last rows and columns, with 5 neighbours) and internal plots (those with 8 neighbours).

The conditional variances (from $\boldsymbol{M}$) and the spatial dependency parameters (from $\boldsymbol{B}$) can be used to write the 2dCAR model (with same covariance structure as an ar1(Column).ar1(Row)) in terms of conditional expectations and variances as follows (assuming $\mu_{ij} = 0$):

For corner plots (3 neighbours) the conditional expectations are given by

$$\begin{aligned}
\mathrm{E}\left(y_{11}|y_{-11}\right) &= \phi_r y_{21} + \phi_c y_{12} - \phi_r\phi_c y_{22} \\
\mathrm{E}\left(y_{rc}|y_{-rc}\right) &= \phi_r y_{r,c-1} + \phi_c y_{r-1,c-1} - \phi_r\phi_c y_{r-1,c} \\
\mathrm{E}\left(y_{r1}|y_{-r1}\right) &= \phi_r y_{r-1,1} + \phi_c y_{r,2} - \phi_r\phi_c y_{r-1,2} \\
\mathrm{E}\left(y_{1c}|y_{-1c}\right) &= \phi_r y_{2,c} + \phi_c y_{1,c-1} - \phi_r\phi_c y_{2,c-1}
\end{aligned}$$

(Note that in this model definition, to have the same variance structure as the ar1(Column).ar1(Row) the spatial dependency parameter for diagonal neighbours of the corner plots is defined to be $\phi_{rc} = -\phi_r\phi_c$.)

For edge plots (in first and similarly for last column) with 5 neighbours

$$\mathrm{E}\left(y_{i,1}|y_{-i1}\right) = \frac{\phi_r}{1+\phi_r^2}y_{i-1,1} + \frac{\phi_r}{1+\phi_r^2}y_{i+1,1} + \frac{-\phi_r\phi_c}{1+\phi_r^2}y_{i-1,2} + \phi_c y_{i,2} + \frac{-\phi_r\phi_c}{1+\phi_r^2}y_{i+1,2}$$

For edge plots (in first and similarly for last row) with 5 neighbours

$$\mathrm{E}\left(y_{1,j}|y_{-1j}\right) = \frac{\phi_c}{1+\phi_c^2}y_{1,j-1} + \frac{\phi_c}{1+\phi_c^2}y_{1,j+1} + \frac{-\phi_r\phi_c}{1+\phi_c^2}y_{2,j-1} + \phi_r y_{2,j} + \frac{-\phi_r\phi_c}{1+\phi_c^2}y_{2,j+1}$$

and for internal plots (with 8 neighbours)

$$\begin{aligned}
\mathrm{E}\left(y_{i,j}|y_{-ij}\right) = {} & \frac{-\phi_r\phi_c}{(1+\phi_r^2)(1+\phi_c^2)}y_{i-1,j-1} + \frac{\phi_c}{1+\phi_c^2}y_{i,j-1} + \frac{-\phi_r\phi_c}{(1+\phi_r^2)(1+\phi_c^2)}y_{i+1,j-1} + \\
& \frac{\phi_r}{1+\phi_r^2}y_{i-1,j} + \frac{\phi_r}{1+\phi_r^2}y_{i+1,j} + \frac{-\phi_r\phi_c}{(1+\phi_r^2)(1+\phi_c^2)}y_{i-1,j+1} + \\
& \frac{\phi_c}{1+\phi_c^2}y_{i,j+1} + \frac{-\phi_r\phi_c}{(1+\phi_r^2)(1+\phi_c^2)}y_{i+1,j+1}
\end{aligned}$$

The conditional variances are defined as follows: For corner plots (3 neighbours)

$$\begin{aligned}
\mathrm{var}\left(y_{11}|y_{-11}\right) &= (1-\phi_r^2)(1-\phi_c^2)\sigma^2 \\
\mathrm{var}\left(y_{rc}|y_{-rc}\right) &= (1-\phi_r^2)(1-\phi_c^2)\sigma^2 \\
\mathrm{var}\left(y_{r1}|y_{-r1}\right) &= (1-\phi_r^2)(1-\phi_c^2)\sigma^2 \\
\mathrm{var}\left(y_{1c}|y_{-1c}\right) &= (1-\phi_r^2)(1-\phi_c^2)\sigma^2
\end{aligned}$$

For edge plots (in first and similarly for last column) with 5 neighbours

$$\mathrm{var}\left(y_{i,1}|y_{-i1}\right) = \frac{1}{1+\phi_r^2}(1-\phi_r^2)(1-\phi_c^2)\sigma^2$$

For edge plots (in first and similarly for last row) with 5 neighbours

$$\mathrm{var}\left(y_{1,j}|y_{-1j}\right) = \frac{1}{1+\phi_c^2}(1-\phi_r^2)(1-\phi_c^2)\sigma^2$$

and for internal plots (with 8 neighbours)

$$\text{var}\,(y_{i,j}|y_{-ij}) = \frac{1}{(1+\phi_c^2)(1+\phi_r^2)}(1-\phi_r^2)(1-\phi_c^2)\sigma^2$$

9.3 Two dimensional lattice MCAR model

The 2dCAR model of (9.2.1) may be extended to the multivariate MCAR case, with the aim of modelling the residual covariance structure from data collected at multiple times (or traits) on each plot in a two dimensional (2d) array of r rows and c columns ($n = r \times c$).

If $\boldsymbol{e}_{ij}$ is a t (t being the number of measurements on each plot) dimensional vector of residuals (assumed to have zero mean) from plot (i,j) (in the 2d lattice) then a MCAR model for the residuals can be written in terms of conditional expectations and conditional variances as

$$E(\boldsymbol{e}_{ij}|\boldsymbol{e}_{-ij}) = \sum_{(uv)\in N(ij)} \boldsymbol{\Omega}_{(ij)(uv)}\boldsymbol{e}_{(uv)}$$

and

$$\text{var}\,(\boldsymbol{e}_{ij}|\boldsymbol{e}_{-ij}) = \boldsymbol{\Gamma}_{ij}$$

where $\boldsymbol{\Omega}_{(i,j),(u,v)}$ represents the (possibly asymmetric) $t \times t$ spatial dependence matrix (between neighbouring plots (i,j) and (u,v)), $\boldsymbol{\Gamma}_{ij}$ represents the $t \times t$ conditional covariance matrix for plot (i,j). Let $\boldsymbol{\Omega}$ be a $nt \times nt$ block matrix with blocks of $t \times t$ $\boldsymbol{\Omega}_{(ij)(uv)}$ matrices, and $\boldsymbol{\Gamma}$ a block diagonal matrix with n blocks of $t \times t$ $\boldsymbol{\Gamma}_{ij}$ covariance matrices. Assuming the following conditions also hold (Mardia, 1988)

- $\boldsymbol{\Omega}_{(ij)(uv)}\boldsymbol{\Gamma}_{uv} = \boldsymbol{\Gamma}_{ij}\boldsymbol{\Omega}^T_{(uv)(ij)}$
- $\boldsymbol{\Omega}_{(ij)(ij)} = \boldsymbol{0}$ and $\boldsymbol{\Omega}_{(ij)(uv)} = \boldsymbol{0}$ if plot (u,v) is not a neighbour of (i,j)
- $(\boldsymbol{I}_{rct} - \boldsymbol{\Omega})^{-1}\boldsymbol{\Gamma}$ is positive definite

then the full distribution for $\boldsymbol{e}$ can be written as

$$\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{R})$$

where

$$\boldsymbol{R} = (\boldsymbol{I}_{rct} - \boldsymbol{\Omega})^{-1}\boldsymbol{\Gamma}$$

This implies the inverse covariance matrix is given by

$$\boldsymbol{R}^{-1} = \boldsymbol{\Gamma}^{-1}(\boldsymbol{I}_{rct} - \boldsymbol{\Omega})$$

9.4 Two dimensional lattice multivariate autoregressive (2dMVAR1) model

The MCAR model presented in the previous section may be used to define a model suitable for the residual covariance structure for multivariate data collected on n spatially

correlated plots in a 2 dimensional lattice of r rows and c columns (where $n = r \times c$), with correlation between plots assumed to follow an autoregressive ar1(Column).ar1(Row) process. For this to be the case some additional conditions (to those above) need to be satisfied. Generalizing from the univariate ar1(Column).ar1(Row) case requires the model to be second order stationary and hence:

- The variance matrix is constant for all plots. That is $\text{var}(\boldsymbol{e}_{ij}) = \boldsymbol{\Sigma}$.
- The matrix of covariances between plots at the same displacement is constant. Hence the matrices of covariances between plots may be referred to in terms of the row and column lags between the plots. Let $\boldsymbol{\Sigma}_{g_1 g_2}$ define the matrix of covariances between plots at row and column lags g_1 and g_2 respectively.

The model will also need to be invariant to direction in both the row and column directions. That is starting at first row or last row or first or last column will produce the same model. Hence symmetry conditions such as those discussed in the previous chapter for the MVAR1 model will be assumed to hold in both the row and column directions.

The form of the MCAR model to satisfy these conditions to be a multivariate AR1 model in 2 spatial dimensions (hereafter referred to as a 2dMVAR1 model) will be derived below.

As in the univariate case the plots may be separated into three groups, namely those that are corner plots (with only 3 neighbours), edge plots (with 5 neighbours) and internal plots (with 8 neighbours). Each of these sets of plots may require different conditional variances and different spatial dependency parameters (as was the case for the univariate 2dCAR model).

Plots with 3 neighbours (corner plots) e.g. plot$(1,1)$

The spatial dependency parameters for the corner plots are defined as

- $\boldsymbol{\Omega}_r$ for adjacent plots in the the row direction
- $\boldsymbol{\Omega}_c$ for adjacent plots in the column direction
- $\boldsymbol{\Omega}_{rc}$ for adjacent plots in the diagonal direction

and the conditional variance of each of the corner plots is defined to be

$$\text{var}(\boldsymbol{e}_{ij}|\boldsymbol{e}_{-ij}) = \boldsymbol{\Gamma}_3$$

where $(ij) = (11), (1c), (r1), (rc)$.

As for the MVAR1 model of the previous chapter we must consider the symmetry constraints introduced to make the model invariant to direction. In the case of a single row the symmetry constraint is

$$\boldsymbol{\Omega}_r \boldsymbol{\Sigma} = \boldsymbol{\Sigma} \boldsymbol{\Omega}_r^T \tag{9.4.2}$$

In the case of a single column the symmetry constraint is given by

$$\boldsymbol{\Omega}_c\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{\Omega}_c^T \tag{9.4.3}$$

Hence for the model to hold in general and hence both of these cases of a single row or column it is intuitive to assume the above constraints both hold.

Consider the four corner plots (without loss of generality consider the corner plot (in row 1 column 1) with residual $\boldsymbol{e}_{11}$). The conditional expectation of this corner plot is given by

$$\mathrm{E}\left(\boldsymbol{e}_{11}|\boldsymbol{e}_{-11}\right) = \boldsymbol{\Omega}_r\boldsymbol{e}_{21} + \boldsymbol{\Omega}_c\boldsymbol{e}_{12} + \boldsymbol{\Omega}_{rc}\boldsymbol{e}_{22}$$

We wish to find the form of the covariances of the corner plots in terms of $\boldsymbol{\Omega}_r$, $\boldsymbol{\Omega}_c$ and $\boldsymbol{\Sigma}$.

The covariance between neighbouring plots in the row direction (for corner plot $\boldsymbol{e}_{11}$) is

$$\begin{aligned}
\mathrm{cov}\left(\boldsymbol{e}_{11},\ \boldsymbol{e}_{21}\right) = \boldsymbol{\Sigma}_{10} &= \mathrm{E}\left(\boldsymbol{e}_{11}\boldsymbol{e}_{21}^T\right) \\
&= \mathrm{E}\left(\mathrm{E}\left(\boldsymbol{e}_{11}\boldsymbol{e}_{21}^T|\boldsymbol{e}_{-11}\right)\right) \\
&= \mathrm{E}\left(\mathrm{E}\left(\boldsymbol{e}_{11}|\boldsymbol{e}_{-11}\right)\boldsymbol{e}_{21}^T\right) \\
&= \mathrm{E}\left((\boldsymbol{\Omega}_r\boldsymbol{e}_{21} + \boldsymbol{\Omega}_c\boldsymbol{e}_{12} + \boldsymbol{\Omega}_{rc}\boldsymbol{e}_{22})\boldsymbol{e}_{21}^T\right) \\
&= \boldsymbol{\Omega}_r\boldsymbol{\Sigma} + \boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{11} + \boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma}_{01}
\end{aligned}$$

The covariance between neighbouring plots in the column direction may similarly be expressed as

$$\mathrm{cov}\left(\boldsymbol{e}_{11},\ \boldsymbol{e}_{12}\right) = \boldsymbol{\Sigma}_{01} = \boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{11} + \boldsymbol{\Omega}_c\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma}_{10}$$

and the covariance between first neighbour plots in the diagonal direction may be written as

$$\mathrm{cov}\left(\boldsymbol{e}_{11},\ \boldsymbol{e}_{22}\right) = \boldsymbol{\Sigma}_{11} = \boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{01} + \boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{10} + \boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma}$$

These relationships can be rearranged to give

$$\begin{aligned}
\boldsymbol{\Sigma}_{10} - \boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma}_{01} - \boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{11} &= \boldsymbol{\Omega}_r\boldsymbol{\Sigma} && (9.4.4) \\
-\boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma}_{10} + \boldsymbol{\Sigma}_{01} - \boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{11} &= \boldsymbol{\Omega}_c\boldsymbol{\Sigma} && (9.4.5) \\
-\boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{10} - \boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{01} + \boldsymbol{\Sigma}_{11} &= \boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma} && (9.4.6)
\end{aligned}$$

and now solve for $\boldsymbol{\Sigma}_{10}$, $\boldsymbol{\Sigma}_{01}$ and $\boldsymbol{\Sigma}_{11}$.

Taking (9.4.4) and adding $\boldsymbol{\Omega_c}$ multiplied by (9.4.6) results in

$$(\boldsymbol{I} - \boldsymbol{\Omega}_c^2)\boldsymbol{\Sigma}_{10} - (\boldsymbol{\Omega}_{rc} + \boldsymbol{\Omega}_c\boldsymbol{\Omega}_r)\boldsymbol{\Sigma}_{01} = (\boldsymbol{\Omega}_r + \boldsymbol{\Omega}_c\boldsymbol{\Omega}_{rc})\boldsymbol{\Sigma}$$

and taking (9.4.5) and adding $\boldsymbol{\Omega}_r$ times (9.4.6) gives

$$-(\boldsymbol{\Omega}_{rc} + \boldsymbol{\Omega}_r\boldsymbol{\Omega}_c)\boldsymbol{\Sigma}_{10} + (\boldsymbol{I} - \boldsymbol{\Omega}_r^2)\boldsymbol{\Sigma}_{01} = (\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_r\boldsymbol{\Omega}_{rc})\boldsymbol{\Sigma} \tag{9.4.7}$$

Hence

$$\boldsymbol{\Sigma}_{10} = (\boldsymbol{I} - \boldsymbol{\Omega}_c^2)^{-1}[(\boldsymbol{\Omega}_r + \boldsymbol{\Omega}_c\boldsymbol{\Omega}_{rc})\boldsymbol{\Sigma} + (\boldsymbol{\Omega}_{rc} + \boldsymbol{\Omega}_c\boldsymbol{\Omega}_r)\boldsymbol{\Sigma}_{01}] \tag{9.4.8}$$

and substituting (9.4.8) into (9.4.7) results in

$$\begin{aligned} -(\boldsymbol{\Omega}_{rc} + \boldsymbol{\Omega}_r\boldsymbol{\Omega}_c)(\boldsymbol{I} - \boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{\Omega}_r + \boldsymbol{\Omega}_c\boldsymbol{\Omega}_{rc})\boldsymbol{\Sigma} \\ -(\boldsymbol{\Omega}_{rc} + \boldsymbol{\Omega}_r\boldsymbol{\Omega}_c)(\boldsymbol{I} - \boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{\Omega}_{rc} + \boldsymbol{\Omega}_c\boldsymbol{\Omega}_r)\boldsymbol{\Sigma}_{01} \\ +(\boldsymbol{I} - \boldsymbol{\Omega}_r^2)\boldsymbol{\Sigma}_{01} = (\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_r\boldsymbol{\Omega}_r\boldsymbol{\Omega}_{rc})\boldsymbol{\Sigma} \end{aligned}$$

Therefore

$$\begin{aligned} &[(\boldsymbol{I} - \boldsymbol{\Omega}_r^2) - (\boldsymbol{\Omega}_{rc} + \boldsymbol{\Omega}_r\boldsymbol{\Omega}_c)(\boldsymbol{I} - \boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{\Omega}_{rc} + \boldsymbol{\Omega}_c\boldsymbol{\Omega}_r)]\boldsymbol{\Sigma}_{01} \\ &= [(\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_r\boldsymbol{\Omega}_{rc}) + (\boldsymbol{\Omega}_{rc} + \boldsymbol{\Omega}_r\boldsymbol{\Omega}_c)(\boldsymbol{I} - \boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{\Omega}_r + \boldsymbol{\Omega}_c\boldsymbol{\Omega}_{rc})]\boldsymbol{\Sigma} \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{\Sigma}_{01} = [(\boldsymbol{I} - \boldsymbol{\Omega}_r^2) - (\boldsymbol{\Omega}_{rc} + \boldsymbol{\Omega}_r\boldsymbol{\Omega}_c)(\boldsymbol{I} - \boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{\Omega}_{rc} + \boldsymbol{\Omega}_c\boldsymbol{\Omega}_r)]^{-1}[(\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_r\boldsymbol{\Omega}_{rc}) + \\ (\boldsymbol{\Omega}_{rc} + \boldsymbol{\Omega}_r\boldsymbol{\Omega}_c)(\boldsymbol{I} - \boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{\Omega}_r + \boldsymbol{\Omega}_c\boldsymbol{\Omega}_{rc})]\boldsymbol{\Sigma} \end{aligned}$$

If the following is assumed

$$\boldsymbol{\Omega}_{rc} = -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c = -\boldsymbol{\Omega}_c\boldsymbol{\Omega}_r \tag{9.4.9}$$

which is similar to the univariate case, then the equations simplify greatly. It follows that

$$(\boldsymbol{I} - \boldsymbol{\Omega}_r^2)\boldsymbol{\Sigma}_{01} = (\boldsymbol{I} - \boldsymbol{\Omega}_r^2)\boldsymbol{\Omega}_c\boldsymbol{\Sigma}$$

and hence

$$\boldsymbol{\Sigma}_{01} = \boldsymbol{\Omega}_c\boldsymbol{\Sigma}$$

The simplifying constraint $\boldsymbol{\Omega}_{rc} = -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c = -\boldsymbol{\Omega}_c\boldsymbol{\Omega}_r$ will be assumed for the remaining derivations.

Substituting (9.4.9) into (9.4.4), (9.4.5) and (9.4.6) it can be shown in a similar fashion that

$$\boldsymbol{\Sigma}_{10} = \boldsymbol{\Omega}_r \boldsymbol{\Sigma}$$

and

$$\boldsymbol{\Sigma}_{11} = \boldsymbol{\Omega}_c \boldsymbol{\Omega}_r \boldsymbol{\Sigma} = \boldsymbol{\Omega}_r \boldsymbol{\Omega}_c \boldsymbol{\Sigma}$$

using (9.4.9).

Hence for all plots the covariance between plots

- 1 apart in row direction $= \boldsymbol{\Sigma}_{10} = \boldsymbol{\Omega}_r \boldsymbol{\Sigma}$
- 1 apart in column direction $= \boldsymbol{\Sigma}_{01} = \boldsymbol{\Omega}_c \boldsymbol{\Sigma}$
- 1 apart in diagonal direction $= \boldsymbol{\Sigma}_{11} = \boldsymbol{\Omega}_c \boldsymbol{\Omega}_r \boldsymbol{\Sigma} = \boldsymbol{\Omega}_r \boldsymbol{\Omega}_c \boldsymbol{\Sigma}$

If the covariance between plots separated by 2 rows and 1 column is defined as $\boldsymbol{\Sigma}_{21}$ and similarly for plots separated by 1 row and 2 columns the covariance is defined as $\boldsymbol{\Sigma}_{12}$ then for plots two apart in the row direction the covariance may be written as

$$\begin{aligned} \operatorname{cov}(\boldsymbol{e}_{11},\ \boldsymbol{e}_{31}) = \boldsymbol{\Sigma}_{20} &= \mathrm{E}\left(\boldsymbol{e}_{11}\boldsymbol{e}_{31}^T\right) \\ &= \mathrm{E}\left(\mathrm{E}\left(\boldsymbol{e}_{11}|\boldsymbol{e}_{-11}\right)\boldsymbol{e}_{31}^T\right) \\ &= \mathrm{E}\left((\boldsymbol{\Omega}_r\boldsymbol{e}_{21} + \boldsymbol{\Omega}_c\boldsymbol{e}_{12} + \boldsymbol{\Omega}_{rc}\boldsymbol{e}_{22})\boldsymbol{e}_{31}^T\right) \\ &= \boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{10} + \boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{21} + \boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma}_{11} \\ &= \boldsymbol{\Omega}_r^2\boldsymbol{\Sigma} + \boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{21} - \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2\boldsymbol{\Sigma} \end{aligned} \tag{9.4.10}$$

In the column direction the covariance may be written as

$$\begin{aligned} \operatorname{cov}(\boldsymbol{e}_{11},\ \boldsymbol{e}_{13}) = \boldsymbol{\Sigma}_{02} &= \mathrm{E}\left(\boldsymbol{e}_{11}\boldsymbol{e}_{13}^T\right) \\ &= \boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{12} + \boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{01} + \boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma}_{11} \\ &= \boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{12} + \boldsymbol{\Omega}_c^2\boldsymbol{\Sigma} - \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2\boldsymbol{\Sigma} \end{aligned}$$

while the covariance between plots in the direction one column across and two rows down is

$$\begin{aligned} \operatorname{cov}(\boldsymbol{e}_{11},\ \boldsymbol{e}_{32}) = \boldsymbol{\Sigma}_{21} &= \mathrm{E}\left(\boldsymbol{e}_{11}\boldsymbol{e}_{32}^T\right) \\ &= \boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{11} + \boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{20} + \boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma}_{10} \\ &= \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c\boldsymbol{\Sigma} + \boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{20} - \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c\boldsymbol{\Sigma} \\ &= \boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{20} \end{aligned}$$

Similarly

$$\boldsymbol{\Sigma}_{12} = \boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{02}$$

Hence substituting into (9.4.10) gives

$$\begin{aligned}\boldsymbol{\Sigma}_{20} &= \boldsymbol{\Omega}_r^2\boldsymbol{\Sigma} + \boldsymbol{\Omega}_c^2\boldsymbol{\Sigma}_{20} - \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2\boldsymbol{\Sigma} \\ (\boldsymbol{I} - \boldsymbol{\Omega}_c^2)\boldsymbol{\Sigma}_{20} &= (\boldsymbol{I} - \boldsymbol{\Omega}_c^2)\boldsymbol{\Omega}_r^2\boldsymbol{\Sigma} \\ \boldsymbol{\Sigma}_{20} &= \boldsymbol{\Omega}_r^2\boldsymbol{\Sigma}\end{aligned}$$

It also follows that $\boldsymbol{\Sigma}_{02} = \boldsymbol{\Omega}_c^2\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{22} = \boldsymbol{\Omega}_r^2\boldsymbol{\Sigma}$.

Furthermore, it can be shown that

$$\begin{aligned}\boldsymbol{\Sigma}_{j0} &= \boldsymbol{\Omega}_r^j\boldsymbol{\Sigma} \\ \boldsymbol{\Sigma}_{0j} &= \boldsymbol{\Omega}_c^j\boldsymbol{\Sigma} \\ \boldsymbol{\Sigma}_{jj} &= \boldsymbol{\Omega}_r^j\boldsymbol{\Omega}_c^j\boldsymbol{\Sigma}\end{aligned}$$

for $j = 1, 2, \ldots, r-1$.

Conditional variances for corner plots

Now consider the variances of the corner plots (those with 3 neighbours) in order to work out the form of the conditional variances for the corner plots.

$$\text{var}\left(\boldsymbol{e}_{11}\right) = \text{E}\left(\text{var}\left(\boldsymbol{e}_{11}|\boldsymbol{e}_{-11}\right)\right) + \text{var}\left(\text{E}\left(\boldsymbol{e}_{11}|\boldsymbol{e}_{-11}\right)\right)$$

Therefore

$$\begin{aligned}\boldsymbol{\Sigma} &= \text{var}\left(\boldsymbol{e}_{11}|\boldsymbol{e}_{-11}\right) + \text{var}\left(\boldsymbol{\Omega}_r\boldsymbol{e}_{21} + \boldsymbol{\Omega}_c\boldsymbol{e}_{12} + \boldsymbol{\Omega}_{rc}\boldsymbol{e}_{22}\right) \\ &= \boldsymbol{\Gamma}_3 + \boldsymbol{\Omega}_r\boldsymbol{\Sigma}\boldsymbol{\Omega}_r^T + \boldsymbol{\Omega}_c\boldsymbol{\Sigma}\boldsymbol{\Omega}_c^T + \boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma}\boldsymbol{\Omega}_{rc}^T \\ &\quad + \boldsymbol{\Omega}_r\text{cov}\left(\boldsymbol{e}_{21},\ \boldsymbol{e}_{12}\right)\boldsymbol{\Omega}_c^T + \boldsymbol{\Omega}_c\text{cov}\left(\boldsymbol{e}_{12},\ \boldsymbol{e}_{21}\right)\boldsymbol{\Omega}_r^T \\ &\quad + \boldsymbol{\Omega}_c\text{cov}\left(\boldsymbol{e}_{12},\ \boldsymbol{e}_{22}\right)\boldsymbol{\Omega}_{rc}^T + \boldsymbol{\Omega}_{rc}\text{cov}\left(\boldsymbol{e}_{22},\ \boldsymbol{e}_{12}\right)\boldsymbol{\Omega}_c^T \\ &\quad + \boldsymbol{\Omega}_r\text{cov}\left(\boldsymbol{e}_{21},\ \boldsymbol{e}_{22}\right)\boldsymbol{\Omega}_{rc}^T + \boldsymbol{\Omega}_{rc}\text{cov}\left(\boldsymbol{e}_{22},\ \boldsymbol{e}_{21}\right)\boldsymbol{\Omega}_r^T\end{aligned}$$

Using (9.4.2), (9.4.3) and (9.4.9) it can be shown that

$$\begin{aligned}\boldsymbol{\Sigma} &= \boldsymbol{\Gamma}_3 + \boldsymbol{\Omega}_r^2\boldsymbol{\Sigma} + \boldsymbol{\Omega}_c^2\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{rc}^2\boldsymbol{\Sigma} + \boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{11}\boldsymbol{\Omega}_c^T + \boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{11}\boldsymbol{\Omega}_r^T + \boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{10}\boldsymbol{\Omega}_{rc}^T + \boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma}_{10}\boldsymbol{\Omega}_c^T + \\ &\quad \boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{01}\boldsymbol{\Omega}_{rc}^T + \boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma}_{01}\boldsymbol{\Omega}_r^T \\ &= \boldsymbol{\Gamma}_3 + \boldsymbol{\Omega}_r^2\boldsymbol{\Sigma} + \boldsymbol{\Omega}_c^2\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{rc}^2\boldsymbol{\Sigma} + \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c\boldsymbol{\Sigma}\boldsymbol{\Omega}_c^T + \boldsymbol{\Omega}_c^2\boldsymbol{\Omega}_r\boldsymbol{\Sigma}\boldsymbol{\Omega}_r^T + \boldsymbol{\Omega}_c\boldsymbol{\Omega}_{\boldsymbol{r}}\boldsymbol{\Sigma}\boldsymbol{\Omega}_{rc}^T + \\ &\quad \boldsymbol{\Omega}_{rc}\boldsymbol{\Omega}_r\boldsymbol{\Sigma}\boldsymbol{\Omega}_c^T + \boldsymbol{\Omega}_r\boldsymbol{\Omega}_c\boldsymbol{\Sigma}\boldsymbol{\Omega}_{rc}^T + \boldsymbol{\Omega}_{rc}\boldsymbol{\Omega}_c\boldsymbol{\Sigma}\boldsymbol{\Omega}_r^T \\ &= \boldsymbol{\Gamma}_3 + (\boldsymbol{\Omega}_r^2 + \boldsymbol{\Omega}_c^2 + \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2 + \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2 + \boldsymbol{\Omega}_c^2\boldsymbol{\Omega}_r^2 - \boldsymbol{\Omega}_c^2\boldsymbol{\Omega}_r^2 - \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2 - \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2 - \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2)\boldsymbol{\Sigma}\end{aligned}$$

Therefore

$$\begin{aligned}\boldsymbol{\Gamma}_3 &= (\boldsymbol{I} - \boldsymbol{\Omega}_r^2 - \boldsymbol{\Omega}_c^2 + \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2)\boldsymbol{\Sigma} \\ &= (\boldsymbol{I} - \boldsymbol{\Omega}_r^2)(\boldsymbol{I} - \boldsymbol{\Omega}_c^2)\boldsymbol{\Sigma} \\ &= (\boldsymbol{I} - \boldsymbol{\Omega}_c^2)(\boldsymbol{I} - \boldsymbol{\Omega}_r^2)\boldsymbol{\Sigma}\end{aligned}$$

Plots with 5 neighbours (edge plots) e.g. plot$(2,1)$

Now consider the edge plots (those that have 5 neighbours) for example Plot$(2,1)$. Different spatial dependency parameters are defined for these plots and the aim is to express these spatial dependency parameters in terms of $\boldsymbol{\Omega}_r$, $\boldsymbol{\Omega}_c$ and $\boldsymbol{\Omega}_{rc}$. The spatial dependency parameters for the edge plots are defined as

- $\boldsymbol{\Omega}_{r5}$ for adjacent plots in the the row direction
- $\boldsymbol{\Omega}_{c5}$ for adjacent plots in the column direction
- $\boldsymbol{\Omega}_{rc5}$ for adjacent plots in the diagonal direction

The constraint $\boldsymbol{\Omega}_{rc5} = -\boldsymbol{\Omega}_{c5}\boldsymbol{\Omega}_{r5} = -\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_{c5}$, is assumed as for the corner plots.

The covariances are defined to be the same for all plots (corner, edge and internal), therefore the covariance 1 apart across the rows is given by

$$\begin{aligned}\text{cov}\,(\boldsymbol{e}_{21},\ \boldsymbol{e}_{11}) &= \boldsymbol{\Sigma}_{10} = \boldsymbol{\Omega}_r\boldsymbol{\Sigma} \\ &= \text{E}\,(\boldsymbol{e}_{21}\boldsymbol{e}_{11}) \\ &= \text{E}\left(\text{E}\,(\boldsymbol{e}_{21}|\boldsymbol{e}_{-21})\,\boldsymbol{e}_{11}^T\right) \\ &= \text{E}\,((\boldsymbol{\Omega}_{r5}\boldsymbol{e}_{11} + \boldsymbol{\Omega}_{r5}\boldsymbol{e}_{31} + \boldsymbol{\Omega}_{c5}\boldsymbol{e}_{22} + \boldsymbol{\Omega}_{rc5}\boldsymbol{e}_{12} + \boldsymbol{\Omega}_{rc5}\boldsymbol{e}_{32})\boldsymbol{e}_{11}) \\ &= \boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{r5}\text{cov}\,(\boldsymbol{e}_{31},\ \boldsymbol{e}_{11}) + \boldsymbol{\Omega}_{c5}\text{cov}\,(\boldsymbol{e}_{22},\ \boldsymbol{e}_{11}) + \boldsymbol{\Omega}_{rc5}\text{cov}\,(\boldsymbol{e}_{12},\ \boldsymbol{e}_{11}) \\ &= \boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma}_{20} + \boldsymbol{\Omega}_{c5}\boldsymbol{\Sigma}_{11} + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma}_{01} + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma}_{21} \qquad (9.4.11)\end{aligned}$$

Using the results derived previously

$$\begin{aligned}\boldsymbol{\Sigma}_{20} &= \boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{10} = \boldsymbol{\Omega}_r^2\boldsymbol{\Sigma} \\ \boldsymbol{\Sigma}_{02} &= \boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{01} = \boldsymbol{\Omega}_c^2\boldsymbol{\Sigma} \\ \boldsymbol{\Sigma}_{21} &= \boldsymbol{\Omega}_c\boldsymbol{\Sigma}_{20} = \boldsymbol{\Omega}_c\boldsymbol{\Omega}_r^2\boldsymbol{\Sigma}\end{aligned}$$

and substituting these into (9.4.11) gives

$$\begin{aligned}\boldsymbol{\Omega}_r &= \boldsymbol{\Omega}_{r5}(\boldsymbol{I} + \boldsymbol{\Omega}_r^2) + \boldsymbol{\Omega}_{c5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_{rc5}(\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c) \\ &= \boldsymbol{\Omega}_{r5}(\boldsymbol{I} + \boldsymbol{\Omega}_r^2) + \boldsymbol{\Omega}_{c5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_{rc5}(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)\boldsymbol{\Omega}_c\end{aligned}$$

The covariance for plots 1 apart across columns is given by

$$\begin{aligned}
\text{cov}\left(\boldsymbol{e}_{21},\ \boldsymbol{e}_{22}\right) &= \boldsymbol{\Sigma}_{01} = \boldsymbol{\Omega}_c\boldsymbol{\Sigma} \\
&= \text{E}\left(\boldsymbol{e}_{21}\boldsymbol{e}_{22}\right) \\
&= \text{E}\left(\text{E}\left(\boldsymbol{e}_{21}|\boldsymbol{e}_{-21}\right)\boldsymbol{e}_{22}^T\right) \\
&= \text{E}\left((\boldsymbol{\Omega}_{r5}\boldsymbol{e}_{11} + \boldsymbol{\Omega}_{r5}\boldsymbol{e}_{31} + \boldsymbol{\Omega}_{c5}\boldsymbol{e}_{22} + \boldsymbol{\Omega}_{rc5}\boldsymbol{e}_{12} + \boldsymbol{\Omega}_{rc5}\boldsymbol{e}_{32})\boldsymbol{e}_{22}\right) \\
&= \boldsymbol{\Omega}_{r5}\text{cov}\left(\boldsymbol{e}_{11},\ \boldsymbol{e}_{22}\right) + \boldsymbol{\Omega}_{r5}\text{cov}\left(\boldsymbol{e}_{31},\ \boldsymbol{e}_{22}\right) + \\
&\quad \boldsymbol{\Omega}_{c5}\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{rc5}\text{cov}\left(\boldsymbol{e}_{12},\ \boldsymbol{e}_{22}\right) + \boldsymbol{\Omega}_{rc5}\text{cov}\left(\boldsymbol{e}_{32},\ \boldsymbol{e}_{22}\right) \\
&= \boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma}_{11} + \boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma}_{11} + \boldsymbol{\Omega}_{c5}\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma}_{10} + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma}_{10} \\
&= \boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{c5}\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r\boldsymbol{\Sigma} + \boldsymbol{\Omega}\boldsymbol{\Omega}_r\boldsymbol{\Sigma}
\end{aligned}$$

Therefore

$$\begin{aligned}
\boldsymbol{\Omega}_c &= \boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_{c5} + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r \\
&= 2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_{c5} + 2\boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r
\end{aligned}$$

The covariance for plots 1 apart diagonally e.g. $\text{cov}\left(\boldsymbol{e}_{21},\ \boldsymbol{e}_{12}\right)$ is given by

$$\begin{aligned}
\text{cov}\left(\boldsymbol{e}_{21},\ \boldsymbol{e}_{12}\right) &= \boldsymbol{\Sigma}_{11} = \boldsymbol{\Omega}_r\boldsymbol{\Omega}_c\boldsymbol{\Sigma} \\
&= \text{E}\left(\boldsymbol{e}_{21}\boldsymbol{e}_{12}\right) = \text{E}\left(\text{E}\left(\boldsymbol{e}_{21}|\boldsymbol{e}_{-21}\right)\boldsymbol{e}_{12}\right) \\
&= \boldsymbol{\Omega}_{r5}\text{cov}\left(\boldsymbol{e}_{11},\ \boldsymbol{e}_{12}\right) + \boldsymbol{\Omega}_{r5}\text{cov}\left(\boldsymbol{e}_{31},\ \boldsymbol{e}_{12}\right) + \\
&\quad \boldsymbol{\Omega}_{c5}\text{cov}\left(\boldsymbol{e}_{22},\ \boldsymbol{e}_{12}\right) + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{rc5}\text{cov}\left(\boldsymbol{e}_{32},\ \boldsymbol{e}_{12}\right) \\
&= \boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma}_{01} + \boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma}_{21} + \boldsymbol{\Omega}_{c5}\boldsymbol{\Sigma}_{10} + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma}_{20} \\
&= \boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_c\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_c\boldsymbol{\Omega}_r^2\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{c5}\boldsymbol{\Omega}_r\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma} + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r^2\boldsymbol{\Sigma}
\end{aligned}$$

Therefore

$$\begin{aligned}
\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c &= \boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_{c5}\boldsymbol{\Omega}_r + \boldsymbol{\Omega}_{rc5} + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r^2 \\
&= \boldsymbol{\Omega}_{r5}(\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c) + \boldsymbol{\Omega}_{c5}\boldsymbol{\Omega}_r + \boldsymbol{\Omega}_{rc5}(\boldsymbol{I} + \boldsymbol{\Omega}_r^2) \\
&= \boldsymbol{\Omega}_{r5}(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_{c5}\boldsymbol{\Omega}_r + \boldsymbol{\Omega}_{rc5}(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)
\end{aligned}$$

Therefore

$$\boldsymbol{\Omega}_r = \boldsymbol{\Omega}_{r5}(\boldsymbol{I} + \boldsymbol{\Omega}_r^2) + \boldsymbol{\Omega}_{c5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_{rc5}(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)\boldsymbol{\Omega}_c \tag{9.4.12}$$

$$\boldsymbol{\Omega}_c = 2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_{c5} + 2\boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r \tag{9.4.13}$$

$$\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c = \boldsymbol{\Omega}_{r5}(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_{c5}\boldsymbol{\Omega}_r + \boldsymbol{\Omega}_{rc5}(\boldsymbol{I} + \boldsymbol{\Omega}_r^2) \tag{9.4.14}$$

Taking (9.4.13) multiplied by $\boldsymbol{\Omega}_r$ minus (9.4.14) gives

$$\begin{aligned}
2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c\boldsymbol{\Omega}_r - \boldsymbol{\Omega}_{r5}(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)\boldsymbol{\Omega}_c + 2\boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r^2 - \boldsymbol{\Omega}_{rc5}(\boldsymbol{I}+\boldsymbol{\Omega}_r^2) &= \boldsymbol{\Omega}_c\boldsymbol{\Omega}_r - \boldsymbol{\Omega}_r\boldsymbol{\Omega}_c \\
2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c - \boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_c - \boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c + 2\boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r^2 - \boldsymbol{\Omega}_{rc5} - \boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r^2 &= 0 \\
\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c - \boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r^2 - \boldsymbol{\Omega}_{rc5} &= 0 \\
\boldsymbol{\Omega}_{r5}(\boldsymbol{\Omega}_r^2 - \boldsymbol{I})\boldsymbol{\Omega}_c + \boldsymbol{\Omega}_{rc5}(\boldsymbol{\Omega}_r^2 - \boldsymbol{I}) &= 0 \qquad (9.4.15)
\end{aligned}$$

Taking (9.4.12) minus (9.4.13) multiplied by $\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c$ results in

$$\begin{aligned}
\boldsymbol{\Omega}_{r5}(\boldsymbol{I}+\boldsymbol{\Omega}_r^2) - 2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2 + \boldsymbol{\Omega}_{rc5}(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)\boldsymbol{\Omega}_c - 2\boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c &= \boldsymbol{\Omega}_r - \boldsymbol{\Omega}_c\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c \\
\boldsymbol{\Omega}_{r5}(\boldsymbol{I}+\boldsymbol{\Omega}_r^2 - 2\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2) + \boldsymbol{\Omega}_{rc5}(\boldsymbol{I}-\boldsymbol{\Omega}_r^2)\boldsymbol{\Omega}_c &= \boldsymbol{\Omega}_r(\boldsymbol{I}-\boldsymbol{\Omega}_c^2) \qquad (9.4.16)
\end{aligned}$$

Now taking (9.4.15) multiplied by $\boldsymbol{\Omega}_c$ and adding (9.4.16) gives

$$\begin{aligned}
\boldsymbol{\Omega}_{r5}(\boldsymbol{\Omega}_r^2 - \boldsymbol{I})\boldsymbol{\Omega}_c^2 + \boldsymbol{\Omega}_{r5}(\boldsymbol{I}+\boldsymbol{\Omega}_r^2 - 2\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2) &= \boldsymbol{\Omega}_r(\boldsymbol{I}-\boldsymbol{\Omega}_c^2) \\
\boldsymbol{\Omega}_{r5}(\boldsymbol{I}+\boldsymbol{\Omega}_r^2 - \boldsymbol{\Omega}_c^2 - \boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2) &= \boldsymbol{\Omega}_r(\boldsymbol{I}-\boldsymbol{\Omega}_c^2) \\
\boldsymbol{\Omega}_{r5}(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)(\boldsymbol{I}-\boldsymbol{\Omega}_c^2) &= \boldsymbol{\Omega}_r(\boldsymbol{I}-\boldsymbol{\Omega}_c^2)
\end{aligned}$$

Hence

$$\boldsymbol{\Omega}_{r5} = \boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}$$

Substituting $\boldsymbol{\Omega}_{r5}$ into (9.4.15) gives

$$\boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{\Omega}_r^2 - \boldsymbol{I})\boldsymbol{\Omega}_c = \boldsymbol{\Omega}_{rc5}(\boldsymbol{I}-\boldsymbol{\Omega}_r^2)$$

and hence

$$\begin{aligned}
\boldsymbol{\Omega}_{rc5} &= -\boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}\boldsymbol{\Omega}_c \\
&= -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}
\end{aligned}$$

Substituting into (9.4.13) gives

$$\begin{aligned}
\boldsymbol{\Omega}_{c5} &= \boldsymbol{\Omega}_c - 2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c - 2\boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_c \\
&= \boldsymbol{\Omega}_c - 2\boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c + 2\boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c \\
&= \boldsymbol{\Omega}_c
\end{aligned}$$

Therefore the three spatial dependency parameters for edge plots (in the first and last column of the lattice) are given by

$$\begin{aligned}
\boldsymbol{\Omega}_{r5:c} &= \boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1} \\
\boldsymbol{\Omega}_{c5:c} &= \boldsymbol{\Omega}_c \\
\boldsymbol{\Omega}_{rc5:c} &= -\boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}\boldsymbol{\Omega}_c = -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}
\end{aligned}$$

Similarly it can be shown that the spatial dependency parameters for edge plots in the first and last row are given by

$$\begin{aligned}
\boldsymbol{\Omega}_{r5:r} &= \boldsymbol{\Omega}_r \\
\boldsymbol{\Omega}_{c5:r} &= \boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1} \\
\boldsymbol{\Omega}_{rc5:r} &= -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1}
\end{aligned}$$

Now consider the variances of edge plots in the first and last column. The conditional variance for edge plots in the first and last column is denoted by $\boldsymbol{\Gamma}_{5col}$.

$$\begin{aligned}
\text{var}\,(\boldsymbol{e}_{21}) = \boldsymbol{\Sigma} &= \text{E}\,(\text{var}\,(\boldsymbol{e}_{21}|\boldsymbol{e}_{-21})) + \text{var}\,(\text{E}\,(\boldsymbol{e}_{21}|\boldsymbol{e}_{-21})) \\
&= \text{var}\,(\boldsymbol{e}_{21}|\boldsymbol{e}_{-21}) + \text{var}\,(\boldsymbol{\Omega}_{r5}\boldsymbol{e}_{11} + \boldsymbol{\Omega}_{r5}\boldsymbol{e}_{31} + \boldsymbol{\Omega}_{c5}\boldsymbol{e}_{22} + \boldsymbol{\Omega}_{rc5}(\boldsymbol{e}_{12}+\boldsymbol{e}_{32})) \\
&= \boldsymbol{\Gamma}_{5col} + 2\boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1})^T + \boldsymbol{\Omega}_c\boldsymbol{\Sigma}\boldsymbol{\Omega}_c^T + \\
&\quad 2\boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma}\boldsymbol{\Omega}_{rc5}^T + 2\boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma}_{20}\boldsymbol{\Omega}_{r5}^T + \boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma}_{11}\boldsymbol{\Omega}_{c5}^T + \\
&\quad \boldsymbol{\Omega}_{c5}\boldsymbol{\Sigma}_{11}\boldsymbol{\Omega}_{r5}^T + \boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma}_{01}\boldsymbol{\Omega}_{rc5}^T + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma}_{01}\boldsymbol{\Omega}_{r5}^T + \boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{11}\boldsymbol{\Omega}_{rc5}^T + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{11}\boldsymbol{\Omega}_{r5}^T + \\
&\quad \boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma}_{11}\boldsymbol{\Omega}_{c5}^T + \boldsymbol{\Omega}_{c5}\boldsymbol{\Sigma}_{11}\boldsymbol{\Omega}_{r5}^T + \boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{11}\boldsymbol{\Omega}_{rc5}^T + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{11}\boldsymbol{\Omega}_{r5}^T + \\
&\quad \boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma}_{01}\boldsymbol{\Omega}_{rc5}^T + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma}_{01}\boldsymbol{\Omega}_{r5}^T + \boldsymbol{\Omega}_{c5}\boldsymbol{\Sigma}_{10}\boldsymbol{\Omega}_{rc5}^T + \boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma}_{10}\boldsymbol{\Omega}_{c5}^T + \boldsymbol{\Omega}_{c5}\boldsymbol{\Sigma}_{10}\boldsymbol{\Omega}_{rc5}^T + \\
&\quad \boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma}_{10}\boldsymbol{\Omega}_{c5}^T + 2\boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r\boldsymbol{\Sigma}_{10}\boldsymbol{\Omega}_{rc5}^T
\end{aligned}$$

Hence

$$\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{\Gamma}_{5col} + 2\boldsymbol{\Omega}_{r5}^2\boldsymbol{\Sigma} + \boldsymbol{\Omega}_c^2\boldsymbol{\Sigma} + 2\boldsymbol{\Omega}_{rc5}^2\boldsymbol{\Sigma} + \\
&\quad 2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma} + 2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c\boldsymbol{\Omega}_{c5}\boldsymbol{\Sigma} + \\
&\quad 2\boldsymbol{\Omega}_{c5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c\boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma} + 2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_c\boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma} + \\
&\quad 2\boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_c\boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma} + 2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c\boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma} + \\
&\quad 2\boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c\boldsymbol{\Omega}_{r5}\boldsymbol{\Sigma} + 2\boldsymbol{\Omega}_{c5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma} + \\
&\quad 2\boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_{c5}\boldsymbol{\Sigma} + 2\boldsymbol{\Omega}_{rc5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_{rc5}\boldsymbol{\Sigma}
\end{aligned}$$

Now substituting $\boldsymbol{\Omega}_{c5} = \boldsymbol{\Omega}_c$ and $\boldsymbol{\Omega}_{rc5} = -\boldsymbol{\Omega}_c\boldsymbol{\Omega}_{r5}$ gives

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{\Gamma}_{5col} + [2\boldsymbol{\Omega}_{r5}^2 + \boldsymbol{\Omega}_c^2 + 2\boldsymbol{\Omega}_{rc5}^2 + 2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_{r5} + 2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c^2 + 2\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c^2\boldsymbol{\Omega}_{r5} - 2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_c^2\boldsymbol{\Omega}_r \\
&\quad -2\boldsymbol{\Omega}_c^2\boldsymbol{\Omega}_{r5}^2 - 2\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2\boldsymbol{\Omega}_{r5} - 2\boldsymbol{\Omega}_c\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c\boldsymbol{\Omega}_{r5} - 2\boldsymbol{\Omega}_c\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c\boldsymbol{\Omega}_{r5} \\
&\quad -2\boldsymbol{\Omega}_c\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c + 2\boldsymbol{\Omega}_c\boldsymbol{\Omega}_{r5}\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c\boldsymbol{\Omega}_{r5}]\boldsymbol{\Sigma} \\
&= \boldsymbol{\Gamma}_{5col} + [2\boldsymbol{\Omega}_{r5}^2 + \boldsymbol{\Omega}_c^2 + 2\boldsymbol{\Omega}_{r5}^2\boldsymbol{\Omega}_r^2 - 2\boldsymbol{\Omega}_{r5}^2\boldsymbol{\Omega}_c^2 - 2\boldsymbol{\Omega}_{r5}^2\boldsymbol{\Omega}_r^2\boldsymbol{\Omega}_c^2]\boldsymbol{\Sigma}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\boldsymbol{\Gamma}_{5col} &= [\boldsymbol{I} - 2\boldsymbol{\Omega}_{r5}^2 - 2\boldsymbol{\Omega}_{r5}^2\boldsymbol{\Omega}_r^2 - \boldsymbol{\Omega}_c^2(\boldsymbol{I} - 2\boldsymbol{\Omega}_{r5}^2 - 2\boldsymbol{\Omega}_{r5}^2\boldsymbol{\Omega}_r^2)]\boldsymbol{\Sigma} \\
&= (\boldsymbol{I} - \boldsymbol{\Omega}_c^2)(\boldsymbol{I} - 2\boldsymbol{\Omega}_{r5}^2 - 2\boldsymbol{\Omega}_{r5}^2\boldsymbol{\Omega}_r^2)\boldsymbol{\Sigma} \\
&= (\boldsymbol{I} - \boldsymbol{\Omega}_c^2)(\boldsymbol{I} - 2\boldsymbol{\Omega}_r^2(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-2} - 2\boldsymbol{\Omega}_r^4(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-2})\boldsymbol{\Sigma} \\
&= (\boldsymbol{I} - \boldsymbol{\Omega}_c^2)(\boldsymbol{I} - 2\boldsymbol{\Omega}_r^2(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-2}(\boldsymbol{I} + \boldsymbol{\Omega}_r^2))\boldsymbol{\Sigma} \\
&= (\boldsymbol{I} - \boldsymbol{\Omega}_c^2)(\boldsymbol{I} - 2\boldsymbol{\Omega}_r^2(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-1})\boldsymbol{\Sigma} \\
&= (\boldsymbol{I} - \boldsymbol{\Omega}_c^2)((\boldsymbol{I} + \boldsymbol{\Omega}_r^2) - 2\boldsymbol{\Omega}_r^2)(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-1}\boldsymbol{\Sigma} \\
&= (\boldsymbol{I} - \boldsymbol{\Omega}_c^2)(\boldsymbol{I} - \boldsymbol{\Omega}_r^2)(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-1}\boldsymbol{\Sigma} \\
&= (\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I} - \boldsymbol{\Omega}_c^2)(\boldsymbol{I} - \boldsymbol{\Omega}_r^2)\boldsymbol{\Sigma}
\end{aligned}
$$

Using a similar approach it can be shown that the conditional variance for edge plots with 5 neighbours in the first and last row is given by

$$\boldsymbol{\Gamma}_{5row} = (\boldsymbol{I} + \boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{I} - \boldsymbol{\Omega}_c^2)(\boldsymbol{I} - \boldsymbol{\Omega}_r^2)\boldsymbol{\Sigma}$$

Internal plots with 8 neighbours (e.g. plot $(2, 2)$)

Different spatial dependency parameters are defined for the internal plots, namely

- $\boldsymbol{\Omega}_{r8}$ for adjacent plots in the the row direction
- $\boldsymbol{\Omega}_{c8}$ for adjacent plots in the column direction
- $\boldsymbol{\Omega}_{rc8}$ for adjacent plots in the diagonal direction

with $\boldsymbol{\Omega}_{rc8} = -\boldsymbol{\Omega}_{r8}\boldsymbol{\Omega}_{c8} = -\boldsymbol{\Omega}_{c8}\boldsymbol{\Omega}_{r8}$.

Once again it is required to find the form of these spatial dependencies in terms of $\boldsymbol{\Omega}_r$, $\boldsymbol{\Omega}_c$ and $\boldsymbol{\Omega}_{rc}$ so that the conditions of equal variances and covariances hold for all plots.

The same approach as for the edge plots is followed, namely taking the covariance between plots 1 apart in the column direction ($\text{cov}\,(\boldsymbol{e}_{ij},\ \boldsymbol{e}_{i,j+1}) = \boldsymbol{\Sigma}_{01} = \boldsymbol{\Omega}_c\boldsymbol{\Sigma}$), the covariance between plots 1 apart in the row direction ($\text{cov}\,(\boldsymbol{e}_{ij},\ \boldsymbol{e}_{i+1,j}) = \boldsymbol{\Sigma}_{10} = \boldsymbol{\Omega}_r\boldsymbol{\Sigma}$) and the covariance between plots in the diagonal direction ($\text{cov}\,(\boldsymbol{e}_{ij},\ \boldsymbol{e}_{i+1,j+1}) = \boldsymbol{\Sigma}_{11} = \boldsymbol{\Omega}_r\boldsymbol{\Omega}_c\boldsymbol{\Sigma}$) and solving the resulting three equations for $\boldsymbol{\Omega}_{r8}$, $\boldsymbol{\Omega}_{c8}$ and $\boldsymbol{\Omega}_{rc8}$.

The algebra has been omitted here but it can be shown that

$$\begin{aligned}
\boldsymbol{\Omega}_{c8} &= \boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1} \\
\boldsymbol{\Omega}_{r8} &= \boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1} \\
\boldsymbol{\Omega}_{rc8} &= -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1}
\end{aligned}$$

In a similar way the conditional covariance matrices for internal plots (with 8 neighbours) can be shown to be

$$\boldsymbol{\Gamma}_8 = (\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I}-\boldsymbol{\Omega}_c^2)(\boldsymbol{I}-\boldsymbol{\Omega}_r^2)\boldsymbol{\Sigma}$$

Therefore, in summary the 2dMVAR1 model is defined with conditional covariance matrices as follows,
for corner plots:

$$\begin{aligned}
\boldsymbol{\Gamma}_3 &= (\boldsymbol{I}-\boldsymbol{\Omega}_r^2)\boldsymbol{\Sigma}(\boldsymbol{I}-\boldsymbol{\Omega}_c^2) \\
&= (\boldsymbol{I}-\boldsymbol{\Omega}_c^2)(\boldsymbol{I}-\boldsymbol{\Omega}_r^2)\boldsymbol{\Sigma}
\end{aligned}$$

for edge plots, first and last row:

$$\begin{aligned}
\boldsymbol{\Gamma}_{5row} &= (\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{I}-\boldsymbol{\Omega}_c^2)(\boldsymbol{I}-\boldsymbol{\Omega}_r^2)\boldsymbol{\Sigma} \\
&= (\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1}\boldsymbol{\Gamma}_3
\end{aligned}$$

for edge plots, first and last column:

$$\begin{aligned}
\boldsymbol{\Gamma}_{5col} &= (\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I}-\boldsymbol{\Omega}_c^2)(\boldsymbol{I}-\boldsymbol{\Omega}_r^2)\boldsymbol{\Sigma} \\
&= (\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}\boldsymbol{\Gamma}_3
\end{aligned}$$

for internal plots:

$$\begin{aligned}
\boldsymbol{\Gamma}_8 &= (\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I}-\boldsymbol{\Omega}_c^2)(\boldsymbol{I}-\boldsymbol{\Omega}_r^2)\boldsymbol{\Sigma} \\
&= (\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}\boldsymbol{\Gamma}_3
\end{aligned}$$

The spatial dependency matrices are defined as follows,
for corner plots (3 neighbours):

$$\begin{aligned}
Row &: \boldsymbol{\Omega}_r \\
Col &: \boldsymbol{\Omega}_c \\
Diagonal &: \boldsymbol{\Omega}_{rc} = -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c = -\boldsymbol{\Omega}_c\boldsymbol{\Omega}_r
\end{aligned}$$

for edge plots (5 neighbours), first and last column

$$\begin{aligned} Row &: \boldsymbol{\Omega}_{r5:c} = \boldsymbol{\Omega}_r(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-1} \\ Col &: \boldsymbol{\Omega}_{c5:c} = \boldsymbol{\Omega}_c \\ Diagonal &: \boldsymbol{\Omega}_{rc5:c} = -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-1} \end{aligned}$$

for edge plots, first and last row

$$\begin{aligned} Row &: \boldsymbol{\Omega}_{r5:r} = \boldsymbol{\Omega}_r \\ Col &: \boldsymbol{\Omega}_{c5:r} = \boldsymbol{\Omega}_c(\boldsymbol{I} + \boldsymbol{\Omega}_c^2)^{-1} \\ Diagonal &: \boldsymbol{\Omega}_{rc5:r} = -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I} + \boldsymbol{\Omega}_c^2)^{-1} \end{aligned}$$

for internal plots (8 neighbours)

$$\begin{aligned} Row &: \boldsymbol{\Omega}_{r8} = \boldsymbol{\Omega}_r(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-1} \\ Col &: \boldsymbol{\Omega}_{c8} = \boldsymbol{\Omega}_c(\boldsymbol{I} + \boldsymbol{\Omega}_c^2)^{-1} \\ Diagonal &: \boldsymbol{\Omega}_{rc8} = -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I} + \boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-1} \end{aligned}$$

Using these results the form for $\boldsymbol{\Gamma}$, (the block diagonal matrix with blocks $\boldsymbol{\Gamma}_{ij}$, representing the conditional covariance matrix for each plot (i,j)) can be written. The plots are ordered rows within columns ($i = 1, \ldots, r$ and $j = 1, \ldots, c$). Therefore

$$\begin{aligned} \boldsymbol{\Gamma} &= \text{diag}\left(\boldsymbol{\Gamma}_{11}, \boldsymbol{\Gamma}_{21}, \boldsymbol{\Gamma}_{31}, \ldots, \boldsymbol{\Gamma}_{r1}, \boldsymbol{\Gamma}_{12}, \boldsymbol{\Gamma}_{22}, \boldsymbol{\Gamma}_{32}, \ldots, \boldsymbol{\Gamma}_{r2}, \ldots, \boldsymbol{\Gamma}_{1c}, \boldsymbol{\Gamma}_{2c}, \boldsymbol{\Gamma}_{3c}, \ldots, \boldsymbol{\Gamma}_{rc}\right) \\ &= \text{diag}\left(\boldsymbol{\Gamma}_3, \boldsymbol{\Gamma}_{5col}, \boldsymbol{\Gamma}_{5col}, \ldots, \boldsymbol{\Gamma}_3, \boldsymbol{\Gamma}_{5row}, \boldsymbol{\Gamma}_8, \boldsymbol{\Gamma}_8, \ldots, \boldsymbol{\Gamma}_{5row}, \ldots, \boldsymbol{\Gamma}_3, \boldsymbol{\Gamma}_{5col}, \boldsymbol{\Gamma}_{5col}, \ldots, \boldsymbol{\Gamma}_3\right) \end{aligned}$$

Now using this form of $\boldsymbol{\Gamma}$,

$$\boldsymbol{\Gamma}^{-1} = \text{diag}\left(\boldsymbol{M}_1, \boldsymbol{M}_2, \ldots, \boldsymbol{M}_2, \boldsymbol{M}_1\right)$$

where

$$\begin{aligned} \boldsymbol{M}_1 &= (\boldsymbol{I}_r \otimes \boldsymbol{\Gamma}_3^{-1})(\text{diag}\left(\boldsymbol{I}, (\boldsymbol{I} + \boldsymbol{\Omega}_r^2), \ldots, \boldsymbol{I}\right)) \\ &= (\boldsymbol{I}_r \otimes \boldsymbol{\Gamma}_3^{-1})(\boldsymbol{I}_{rt} + \boldsymbol{E}_{1r} \otimes \boldsymbol{\Omega}_r^2) \\ &= (\boldsymbol{I}_r \otimes (\boldsymbol{\Sigma}^{-1}(\boldsymbol{I} - \boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I} - \boldsymbol{\Omega}_c^2)^{-1}))(\boldsymbol{I}_{rt} + \boldsymbol{E}_{1r} \otimes \boldsymbol{\Omega}_r^2) \end{aligned}$$

and

$$\begin{aligned}
\boldsymbol{M}_2 &= (\boldsymbol{I}_r \otimes \boldsymbol{\Gamma}_3^{-1})(\text{diag}\left((\boldsymbol{I}+\boldsymbol{\Omega}_c^2), (\boldsymbol{I}+\boldsymbol{\Omega}_c^2)(\boldsymbol{I}+\boldsymbol{\Omega}_r^2), \ldots, (\boldsymbol{I}+\boldsymbol{\Omega}_c^2)\right)) \\
&= (\boldsymbol{I}_r \otimes \boldsymbol{\Gamma}_3^{-1})(\boldsymbol{I}_{rt} + \boldsymbol{E}_{1r} \otimes \boldsymbol{\Omega}_r^2)[\boldsymbol{I}_r \otimes (\boldsymbol{I}_t + \boldsymbol{\Omega}_c^2)] \\
&= \boldsymbol{M}_1[\boldsymbol{I}_r \otimes (\boldsymbol{I}_t + \boldsymbol{\Omega}_c^2)]
\end{aligned}$$

This can be simplified to

$$\begin{aligned}
\boldsymbol{\Gamma}^{-1} &= (\boldsymbol{I}_c \otimes \boldsymbol{M}_1)[\text{diag}\left(\boldsymbol{I}_{rt}, \boldsymbol{I}_r \otimes (\boldsymbol{I}_t + \boldsymbol{\Omega}_c^2), \ldots, \boldsymbol{I}_{rt}\right)] \\
&= (\boldsymbol{I}_c \otimes \boldsymbol{M}_1)(\boldsymbol{I}_{rct} + \boldsymbol{E}_{1c} \otimes (\boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c^2)) \\
&= [\boldsymbol{I}_c \otimes ((\boldsymbol{I}_r \otimes (\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}-\boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I}-\boldsymbol{\Omega}_c^2)^{-1}))(\boldsymbol{I}_{rt} + \boldsymbol{E}_{1r} \otimes \boldsymbol{\Omega}_r^2))][\boldsymbol{I}_{rct} + \boldsymbol{E}_{1c} \otimes \boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c^2]
\end{aligned}$$

The form for $\boldsymbol{\Omega}$, the $rct \times rct$ block matrix of spatial dependency matrices for neighbouring plots can also be written as (Note the plots are ordered rows within columns)

$$\boldsymbol{\Omega} = \begin{bmatrix}
\boldsymbol{B}_0 & \boldsymbol{B}_2 & 0 & \ldots & 0 \\
\boldsymbol{B}_3 & \boldsymbol{B}_1 & \boldsymbol{B}_3 & \ldots & 0 \\
0 & \boldsymbol{B}_3 & \boldsymbol{B}_1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & \boldsymbol{B}_2 & \boldsymbol{B}_0
\end{bmatrix} \tag{9.4.17}$$

where where $\boldsymbol{B}_0$ is a $rt \times rt$ matrix with

$$\begin{aligned}
\boldsymbol{B}_0 &= \begin{bmatrix}
\boldsymbol{0} & \boldsymbol{\Omega}_r & \boldsymbol{0} & \ldots & \boldsymbol{0} \\
\boldsymbol{\Omega}_{r5:c} & \boldsymbol{0} & \boldsymbol{\Omega}_{r5:c} & \ldots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{\Omega}_{r5:c} & \boldsymbol{0} & \ldots & \boldsymbol{0} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{\Omega}_r & \boldsymbol{0}
\end{bmatrix} \\
&= \begin{bmatrix}
\boldsymbol{0} & \boldsymbol{\Omega}_r & \boldsymbol{0} & \ldots & \boldsymbol{0} \\
(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}\boldsymbol{\Omega}_r & \boldsymbol{0} & (\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}\boldsymbol{\Omega}_r & \ldots & \boldsymbol{0} \\
\boldsymbol{0} & (\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}\boldsymbol{\Omega}_r & \boldsymbol{0} & \ldots & \boldsymbol{0} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{\Omega}_r & \boldsymbol{0}
\end{bmatrix}
\end{aligned}$$

and $\boldsymbol{B}_1$ is a $rt \times rt$ matrix with

$$\begin{aligned}
\boldsymbol{B}_1 &= \begin{bmatrix} \mathbf{0} & \boldsymbol{\Omega}_{r5:r} & \mathbf{0} & \dots & \mathbf{0} \\ \boldsymbol{\Omega}_{r8} & \mathbf{0} & \boldsymbol{\Omega}_{r8} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{r8} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Omega}_{r5:r} & \mathbf{0} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{0} & \boldsymbol{\Omega}_r & \mathbf{0} & \dots & \mathbf{0} \\ \boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1} & \mathbf{0} & \boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Omega}_r & \mathbf{0} \end{bmatrix} \\
&= \boldsymbol{B}_0
\end{aligned}$$

These matrices can be written as

$$\begin{aligned}
\boldsymbol{B}_1 &= \boldsymbol{F}_{1r} \otimes (\boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}) + \boldsymbol{D}_r \otimes (\boldsymbol{\Omega}_r - (\boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1})) \\
&= (\boldsymbol{F}_{1r} - \boldsymbol{D}_r) \otimes (\boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}) + \boldsymbol{D}_r \otimes \boldsymbol{\Omega}_r
\end{aligned}$$

where $\boldsymbol{D}_r$ is a $r \times r$ matrix of all zeros except elements $D_r[1,2]$ and $D_r[r,(r-1)]$ which equal 1 and $\boldsymbol{F}_{1r}$ is a $r \times r$ matrix previously defined.

The matrix $\boldsymbol{B}_2$ in (9.4.17) is given by

$$\begin{aligned}
\boldsymbol{B}_2 &= \begin{bmatrix} \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{rc} & \mathbf{0} & \dots & \mathbf{0} \\ \boldsymbol{\Omega}_{rc5:c} & \boldsymbol{\Omega}_c & \boldsymbol{\Omega}_{rc5:c} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{rc5:c} & \boldsymbol{\Omega}_c & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Omega}_{rc} & \boldsymbol{\Omega}_c \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\Omega}_c & -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c & \mathbf{0} & \dots & \mathbf{0} \\ -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1} & \boldsymbol{\Omega}_c & -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1} & \dots & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1} & \boldsymbol{\Omega}_c & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c & \boldsymbol{\Omega}_c \end{bmatrix} \\
&= \boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c + (\boldsymbol{F}_{1r} - \boldsymbol{D}_r) \otimes \boldsymbol{\Omega}_{rc5:c} + \boldsymbol{D}_r \otimes \boldsymbol{\Omega}_{rc} \\
&= \boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c - (\boldsymbol{F}_{1r} - \boldsymbol{D}_r) \otimes (\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}) - \boldsymbol{D}_r \otimes (\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c) \\
&= (\boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c)(\boldsymbol{I}_{rt} - \boldsymbol{B}_1)
\end{aligned}$$

The matrix $\boldsymbol{B}_3$ in (9.4.17) is given by

$$\begin{aligned}
\boldsymbol{B}_3 &= \begin{bmatrix} \boldsymbol{\Omega}_{c5:r} & \boldsymbol{\Omega}_{rc5:r} & \mathbf{0} & \dots & \mathbf{0} \\ \boldsymbol{\Omega}_{rc8} & \boldsymbol{\Omega}_{c8} & \boldsymbol{\Omega}_{rc8} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{rc8} & \boldsymbol{\Omega}_{c8} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Omega}_{rc5:r} & \boldsymbol{\Omega}_{c5:r} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1} & -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1} & \dots & \mathbf{0} \\ -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1} & \boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & -\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1} & \boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1} \end{bmatrix} \\
&= \boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1} - (\boldsymbol{F}_{1r}-\boldsymbol{D}_r) \otimes (\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1}) - \\
&\quad \boldsymbol{D}_r \otimes (\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1}) \\
&= \boldsymbol{B}_2(\boldsymbol{I}_r \otimes (\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1})
\end{aligned}$$

Hence,

$$\begin{aligned}
\boldsymbol{\Omega} &= \boldsymbol{I}_c \otimes \boldsymbol{B}_1 + (\boldsymbol{F}_{1c}-\boldsymbol{D}_c) \otimes \boldsymbol{B}_3 + \boldsymbol{D}_c \otimes \boldsymbol{B}_2 \\
&= \boldsymbol{I}_c \otimes [(\boldsymbol{F}_{1r}-\boldsymbol{D}_r) \otimes (\boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}) + \boldsymbol{D}_r \otimes \boldsymbol{\Omega}_r] \\
&\quad +(\boldsymbol{F}_{1c}-\boldsymbol{D}_c) \otimes [\boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1} - (\boldsymbol{F}_{1r}-\boldsymbol{D}_r) \otimes (\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1}) \\
&\quad -\boldsymbol{D}_r \otimes (\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1})] \\
&\quad +\boldsymbol{D}_c \otimes [\boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c - (\boldsymbol{F}_{1r}-\boldsymbol{D}_r) \otimes (\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}) - \boldsymbol{D}_r \otimes (\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c)]
\end{aligned}$$

where $\boldsymbol{D}_c$ is a $c \times c$ matrix of all zeros except elements $D_c[1,2]$ and $D_c[c,(c-1)]$ which equal 1.

The full inverse covariance matrix is given by

$$\begin{aligned}
\boldsymbol{R}^{-1} &= \boldsymbol{\Gamma}^{-1}(\boldsymbol{I}_{rct}-\boldsymbol{\Omega}) \\
&= \{[\boldsymbol{I}_c \otimes ((\boldsymbol{I}_r \otimes (\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}-\boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I}-\boldsymbol{\Omega}_c^2)^{-1}))(\boldsymbol{I}_{rt}+\boldsymbol{E}_{1r} \otimes \boldsymbol{\Omega}_r^2))] \qquad (9.4.18) \\
&\quad [\boldsymbol{I}_{rct}+\boldsymbol{E}_{1c} \otimes \boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c^2]\}\{\boldsymbol{I}_{rct} - \\
&\quad (\boldsymbol{I}_c \otimes [(\boldsymbol{F}_{1r}-\boldsymbol{D}_r) \otimes (\boldsymbol{\Omega}_r(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}) + \boldsymbol{D}_r \otimes \boldsymbol{\Omega}_r] \\
&\quad +(\boldsymbol{F}_{1c}-\boldsymbol{D}_c) \otimes [\boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1} - (\boldsymbol{F}_{1r}-\boldsymbol{D}_r) \otimes (\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1}) \\
&\quad -\boldsymbol{D}_r \otimes (\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)^{-1})] \\
&\quad +\boldsymbol{D}_c \otimes [\boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c - (\boldsymbol{F}_{1r}-\boldsymbol{D}_r) \otimes (\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)^{-1}) - \boldsymbol{D}_r \otimes (\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c)])\} \qquad (9.4.19)
\end{aligned}$$

For this result to be valid, there are 3 sets of constraints that need to hold, namely

$$\begin{aligned}
\boldsymbol{\Omega}_r\boldsymbol{\Sigma} &= \boldsymbol{\Sigma}\boldsymbol{\Omega}_r^T \\
\boldsymbol{\Omega}_c\boldsymbol{\Sigma} &= \boldsymbol{\Sigma}\boldsymbol{\Omega}_c^T \\
\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c &= \boldsymbol{\Omega}_c\boldsymbol{\Omega}_r
\end{aligned}$$

Note that the third constraint holds if the first two constraints hold and $\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c\boldsymbol{\Sigma}$ is also symmetric.

That is, if

$$\begin{aligned}\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c\boldsymbol{\Sigma} &= \boldsymbol{\Sigma}\boldsymbol{\Omega}_c^T\boldsymbol{\Omega}_r^T \\ &= \boldsymbol{\Omega}_c\boldsymbol{\Sigma}\boldsymbol{\Omega}_r^T \\ &= \boldsymbol{\Omega}_c\boldsymbol{\Omega}_r\boldsymbol{\Sigma}\end{aligned}$$

then

$$\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c = \boldsymbol{\Omega}_c\boldsymbol{\Omega}_r$$

If the above form for $\boldsymbol{R}^{-1}$ (9.4.19) is multiplied out it can be seen that the explicit form of the inverse covariance matrix (written here for a lattice of 3 rows and 3 columns, for brevity) is as follows:

$$\boldsymbol{R}^{-1} = [\boldsymbol{I}_{rc} \otimes (\boldsymbol{\Sigma}^{-1}(\boldsymbol{I} - \boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I} - \boldsymbol{\Omega}_c^2)^{-1})]\boldsymbol{\Delta}$$

where $\boldsymbol{\Delta}=$

$$\begin{bmatrix}
\boldsymbol{I} & -\boldsymbol{\Omega}_r & \mathbf{0} & -\boldsymbol{\Omega}_c & -\boldsymbol{\Omega}_{rc} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
-\boldsymbol{\Omega}_r & \boldsymbol{I}+\boldsymbol{\Omega}_r^2 & -\boldsymbol{\Omega}_r & -\boldsymbol{\Omega}_{rc} & -(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)\boldsymbol{\Omega}_c & -\boldsymbol{\Omega}_{rc} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & -\boldsymbol{\Omega}_r & \boldsymbol{I} & \mathbf{0} & -\boldsymbol{\Omega}_{rc} & -\boldsymbol{\Omega}_c & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
-\boldsymbol{\Omega}_c & -\boldsymbol{\Omega}_{rc} & \mathbf{0} & \boldsymbol{I}+\boldsymbol{\Omega}_c^2 & -(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)\boldsymbol{\Omega}_r & \mathbf{0} & -\boldsymbol{\Omega}_c & -\boldsymbol{\Omega}_{rc} & \mathbf{0} \\
-\boldsymbol{\Omega}_{rc} & -(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)\boldsymbol{\Omega}_c & -\boldsymbol{\Omega}_{rc} & -(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)\boldsymbol{\Omega}_r & (\boldsymbol{I}+\boldsymbol{\Omega}_c^2)(\boldsymbol{I}+\boldsymbol{\Omega}_r^2) & -(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)\boldsymbol{\Omega}_r & -\boldsymbol{\Omega}_{rc} & -(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)\boldsymbol{\Omega}_c & -\boldsymbol{\Omega}_{rc} \\
\mathbf{0} & -\boldsymbol{\Omega}_{rc} & -\boldsymbol{\Omega}_c & \mathbf{0} & -(\boldsymbol{I}+\boldsymbol{\Omega}_c^2)\boldsymbol{\Omega}_r & \boldsymbol{I}+\boldsymbol{\Omega}_c^2 & \mathbf{0} & -\boldsymbol{\Omega}_{rc} & -\boldsymbol{\Omega}_c \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & -\boldsymbol{\Omega}_c & -\boldsymbol{\Omega}_{rc} & \mathbf{0} & \boldsymbol{I} & -\boldsymbol{\Omega}_r & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & -\boldsymbol{\Omega}_{rc} & -(\boldsymbol{I}+\boldsymbol{\Omega}_r^2)\boldsymbol{\Omega}_c & -\boldsymbol{\Omega}_{rc} & -\boldsymbol{\Omega}_r & \boldsymbol{I}+\boldsymbol{\Omega}_r^2 & -\boldsymbol{\Omega}_r \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\boldsymbol{\Omega}_{rc} & -\boldsymbol{\Omega}_c & \mathbf{0} & -\boldsymbol{\Omega}_r & \boldsymbol{I}
\end{bmatrix}$$

which can be seen to be the direct multivariate extension of the inverse covariance matrix for a univariate ar1(Column).ar1(Row) process (with multivariate structures $\boldsymbol{\Omega}_r$ and $\boldsymbol{\Omega}_c$ replacing ϕ_r and ϕ_c and $\boldsymbol{\Sigma}$ replacing σ. However, it is clear that this multivariate inverse covariance matrix cannot be simply written as the Kronecker product of two one directional (row) and (column) multivariate MVAR1 processes, as in the univariate case.

It can also be seen that these results follow from simply replacing the univariate spatial dependency parameters and variances with their multivariate counterparts in the MCAR model of (9.2.1).

9.4.1 Fitting the 2dMVAR1 model

To estimate the parameters in the 2dMVAR1 model the REML log-likelihood must be maximized subject to the three sets of constraints. The constraints can be imposed using Lagrange multipliers, as follows

$$C_r = [\mathsf{vechl}(\boldsymbol{\Psi}_r)]^T[\mathsf{vechl}(\boldsymbol{\Omega}_r\boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma}\boldsymbol{\Omega}_r^T)]$$

where $\boldsymbol{\Psi}_r$ is the $t \times t$ symmetric matrix of Lagrange multipliers Ψ_{rij} with $\Psi_{rii} = 0$ for the constraint involving $\boldsymbol{\Omega}_r\boldsymbol{\Sigma}$,

$$C_c = [\mathsf{vechl}(\boldsymbol{\Psi}_c)]^T[\mathsf{vechl}(\boldsymbol{\Omega}_c\boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma}\boldsymbol{\Omega}_c^T)]$$

where $\boldsymbol{\Psi}_c$ is the $t \times t$ symmetric matrix of Lagrange multipliers Ψ_{cij} with $\Psi_{cii} = 0$ for the constraint involving $\boldsymbol{\Omega}_c\boldsymbol{\Sigma}$, and

$$C_{rc} = [\mathsf{vechl}(\boldsymbol{\Psi}_{rc})]^T[\mathsf{vechl}(\boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma}\boldsymbol{\Omega}_{rc}^T)]$$

where $\boldsymbol{\Psi}_{rc}$ is the $t \times t$ symmetric matrix of Lagrange multipliers Ψ_{rcij} with $\Psi_{rcii} = 0$ for the constraint involving $\boldsymbol{\Omega}_{rc}\boldsymbol{\Sigma}$.

Therefore the constrained log-likelihood l_r^* needs to be maximized, where

$$l_r^* = l_r + C_r + C_c + C_{rc}$$

To maximize the log-likelihood the derivatives of l_r^* are taken with respect to the variance parameters and are equated to zero. Therefore for parameters in $\boldsymbol{\Sigma}$, $\boldsymbol{\Omega}_r$ and $\boldsymbol{\Omega}_c$ for example σ_{rs}, this will mean calculating

$$\frac{\partial l_r^*}{\partial \sigma_{rs}} = \boldsymbol{U}(\sigma_{rs}) + \frac{\partial C_r}{\partial \sigma_{rs}} + \frac{\partial C_c}{\partial \sigma_{rs}} + \frac{\partial C_{rc}}{\partial \sigma_{rs}}$$

This requires the derivatives of the full 2dMVAR1 covariance matrix $\boldsymbol{R}$, where

$$\boldsymbol{R}^{-1} = \boldsymbol{\Gamma}^{-1}(\boldsymbol{I} - \boldsymbol{\Omega})$$

The derivatives for the constraints C are also required. These derivatives are presented below.

The derivatives of $\boldsymbol{R}^{-1}$ with respect to the variance parameters $\{\sigma_{rs}\}$ and $\{\omega_{rij}\}$ $\{\omega_{cij}\}$ (combined in vector $\kappa_k, k = 1, \ldots, t(t+1)/2 + 2t^2$) are given by

$$\frac{\partial \boldsymbol{R}^{-1}}{\partial \kappa_k} = \frac{\partial \boldsymbol{R}_1}{\partial \kappa_k}\boldsymbol{R}_2 + \boldsymbol{R}_1\frac{\partial \boldsymbol{R}_2}{\partial \kappa_k}$$

where

$$\boldsymbol{R}_1 = \boldsymbol{\Gamma}^{-1}$$

and

$$\begin{aligned} \boldsymbol{R_2} &= (\boldsymbol{I} - \boldsymbol{\Omega}) \\ &= \boldsymbol{I}_{rct} - \boldsymbol{I}_c \otimes \boldsymbol{B}_1 - (\boldsymbol{F}_{1c} - \boldsymbol{D}_c) \otimes \boldsymbol{B}_3 - \boldsymbol{D}_c \otimes \boldsymbol{B}_2 \end{aligned}$$

For the parameters in κ_k, with $k = 1, \ldots, t(t+1)/2$ (parameters σ_{rs} from $\boldsymbol{\Sigma}$) the

derivative of $\boldsymbol{R}_1$ wrt σ_{rs} is given by

$$\frac{\partial \boldsymbol{R}_1}{\partial \sigma_{rs}} = [\boldsymbol{I}_c \otimes ([\boldsymbol{I}_r \otimes (\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_{rs}}(\boldsymbol{I} - \boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I} - \boldsymbol{\Omega}_c^2)^{-1})][\boldsymbol{I}_{rt} + \boldsymbol{E}_{1r} \otimes \boldsymbol{\Omega}_r^2])][\boldsymbol{I}_{rct} + \boldsymbol{E}_{1c} \otimes \boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c^2]$$

The derivative of $\boldsymbol{R}_2$ wrt σ_{rs} is given by

$$\frac{\partial \boldsymbol{R}_2}{\partial \sigma_{rs}} = 0 \times \boldsymbol{I}_{rct} = \boldsymbol{0}$$

For the parameters in $\boldsymbol{\Omega}_r$, that is in κ_k for $k = (1 + t(t+1)/2, \ldots, t(t+1)/2 + t^2)$

$$\begin{aligned}\frac{\partial \boldsymbol{R}_1}{\partial \omega_{rij}} &= [\boldsymbol{I}_c \otimes ([\boldsymbol{I}_r \otimes (\boldsymbol{\Sigma}^{-1}\frac{\partial (\boldsymbol{I}_t - \boldsymbol{\Omega}_r^2)^{-1}}{\partial \omega_{rij}}(\boldsymbol{I}_t - \boldsymbol{\Omega}_c^2)^{-1})][\boldsymbol{I}_{rt} + \boldsymbol{E}_{1r} \otimes \boldsymbol{\Omega}_r^2])][\boldsymbol{I}_{rct} + \boldsymbol{E}_{1c} \otimes \boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c^2] \\ &+ [\boldsymbol{I}_c \otimes ([\boldsymbol{I}_r \otimes (\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_t - \boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I}_t - \boldsymbol{\Omega}_c^2)^{-1})][\boldsymbol{I}_{rt} + \boldsymbol{E}_{1r} \otimes \frac{\partial \boldsymbol{\Omega}_r^2}{\partial \omega_{rij}}])][\boldsymbol{I}_{rct} + \boldsymbol{E}_{1c} \otimes \boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c^2]\end{aligned}$$

where

$$\frac{\partial (\boldsymbol{I}_t - \boldsymbol{\Omega}_r^2)^{-1}}{\partial \omega_{rij}} = (\boldsymbol{I}_t - \boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{\Omega}_r \frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}} + \frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}}\boldsymbol{\Omega}_r)(\boldsymbol{I}_t - \boldsymbol{\Omega}_r^2)^{-1}$$

and

$$\frac{\partial \boldsymbol{\Omega}_r^2}{\partial \omega_{rij}} = \boldsymbol{\Omega}_r \frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}} + \frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}}\boldsymbol{\Omega}_r$$

The derivative of $\boldsymbol{\Omega}_r$ wrt ω_{rij} is given by

$$\frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}} = \boldsymbol{z}_i \boldsymbol{z}_j^T$$

where $\boldsymbol{z}_i$ is a $t \times 1$ vector containing all zeros except for the i^{th} element (row) which equals 1.

The derivative of $\boldsymbol{R}_2$ wrt ω_{rij} is given by

$$\frac{\partial \boldsymbol{R}_2}{\partial \omega_{rij}} = -\boldsymbol{I}_c \otimes \frac{\partial \boldsymbol{B}_1}{\partial \omega_{rij}} - (\boldsymbol{F}_{1c} - \boldsymbol{D}_c) \otimes \frac{\partial \boldsymbol{B}_3}{\partial \omega_{rij}} - \boldsymbol{D}_c \otimes \frac{\partial \boldsymbol{B}_2}{\partial \omega_{rij}}$$

where

$$\frac{\partial \boldsymbol{B}_1}{\partial \omega_{rij}} = (\boldsymbol{F}_{1r} - \boldsymbol{D}_r) \otimes (\frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}}(\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-1} + \boldsymbol{\Omega}_r \frac{\partial (\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-1}}{\partial \omega_{rij}}) + \boldsymbol{D}_r \otimes \frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}}$$

$$\frac{\partial \boldsymbol{B}_2}{\partial \omega_{rij}} = -(\boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c)\frac{\partial \boldsymbol{B}_1}{\partial \omega_{rij}}$$

and

$$\frac{\partial \boldsymbol{C}_3}{\partial \omega_{rij}} = \frac{\partial \boldsymbol{C}_2}{\partial \omega_{rij}}(\boldsymbol{I}_c \otimes (\boldsymbol{I}_t + \boldsymbol{\Omega}_c^2)^{-1})$$

where

$$\frac{\partial (\boldsymbol{I} + \boldsymbol{\Omega}_r^2)^{-1}}{\partial \omega_{rij}} = -(\boldsymbol{I}_t + \boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{\Omega}\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{rij}} + \frac{\partial \boldsymbol{\Omega}}{\partial \omega_{rij}}\boldsymbol{\Omega})(\boldsymbol{I}_t + \boldsymbol{\Omega}_r^2)^{-1}$$

and $\frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}}$ is defined as above.

The derivatives of $\boldsymbol{R}_1$ wrt ω_{cij} are as follows

$$\begin{aligned}
\frac{\partial \boldsymbol{R}_1}{\partial \omega_{cij}} &= [\boldsymbol{I}_c \otimes ([\boldsymbol{I}_r \otimes (\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_t - \boldsymbol{\Omega}_r^2)^{-1})\frac{\partial (\boldsymbol{I}_t - \boldsymbol{\Omega}_c^2)^{-1}}{\partial \omega_{cij}}][\boldsymbol{I}_{rt} + \boldsymbol{E}_{1r} \otimes \boldsymbol{\Omega}_r^2])] \\
&\quad [\boldsymbol{I}_{rct} + \boldsymbol{E}_{1c} \otimes \boldsymbol{I}_r \otimes \boldsymbol{\Omega}_c^2] \\
&+ [\boldsymbol{I}_c \otimes ([\boldsymbol{I}_r \otimes (\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_t - \boldsymbol{\Omega}_r^2)^{-1}(\boldsymbol{I}_t - \boldsymbol{\Omega}_c^2)^{-1})][\boldsymbol{I}_{rt} + \boldsymbol{E}_{1r} \otimes \boldsymbol{\Omega}_r^2])][\boldsymbol{E}_{1c} \otimes \boldsymbol{I}_r \otimes \frac{\partial \boldsymbol{\Omega}_c^2}{\partial \omega_{cij}}]
\end{aligned}$$

where

$$\frac{\partial (\boldsymbol{I}_t - \boldsymbol{\Omega}_c^2)^{-1}}{\partial \omega_{cij}} = (\boldsymbol{I}_t - \boldsymbol{\Omega}_c^2)^{-1}(\boldsymbol{\Omega}_c\frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cij}} + \frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cij}}\boldsymbol{\Omega}_c)(\boldsymbol{I}_t - \boldsymbol{\Omega}_c^2)^{-1}$$

and

$$\frac{\partial \boldsymbol{\Omega}_c^2}{\partial \omega_{cij}} = \boldsymbol{\Omega}_c\frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cij}} + \frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cij}}\boldsymbol{\Omega}_c$$

The derivative of $\boldsymbol{\Omega}_c$ wrt ω_{cij} is given by

$$\frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cij}} = \boldsymbol{z}_i\boldsymbol{z}_j^T$$

where $\boldsymbol{z}_i$ is a $t \times 1$ vector containing all zeros except for the i^{th} element (row) which equals 1.

The derivative of $\boldsymbol{R}_2$ wrt ω_{cij} is given by

$$\frac{\partial \boldsymbol{R}_2}{\partial \omega_{cij}} = -(\boldsymbol{F}_{1c} - \boldsymbol{D}_c) \otimes \frac{\partial \boldsymbol{B}_3}{\partial \omega_{cij}} - \boldsymbol{D}_c \otimes \frac{\partial \boldsymbol{B}_2}{\partial \omega_{cij}}$$

where

$$\frac{\partial \boldsymbol{B}_2}{\partial \omega_{cij}} = (\boldsymbol{I}_r \otimes \frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cij}})(\boldsymbol{I}_{rt} - \boldsymbol{B}_1)$$

and

$$\frac{\partial \boldsymbol{B}_3}{\partial \omega_{cij}} = \frac{\partial \boldsymbol{B}_2}{\partial \omega_{cij}}(\boldsymbol{I}_r \otimes (\boldsymbol{I}_t + \boldsymbol{\Omega}_c^2)^{-1}) + \boldsymbol{B}_2(\boldsymbol{I}_r \otimes \frac{\partial (\boldsymbol{I}_t + \boldsymbol{\Omega}_c^2)^{-1}}{\partial \omega_{cij}})$$

and where

$$\frac{\partial(\boldsymbol{I}_t + \boldsymbol{\Omega}_c^2)^{-1}}{\partial \omega_{cij}} = -(\boldsymbol{I}_t + \boldsymbol{\Omega}_c^2)^{-1} \frac{\partial \boldsymbol{\Omega}_c^2}{\partial \omega_{cij}} (\boldsymbol{I}_t + \boldsymbol{\Omega}_c^2)^{-1}$$

and $\frac{\partial \boldsymbol{\Omega}_c^2}{\partial \omega_{cij}}$ is defined above.

Derivatives of constraints

The derivative of the constraints $C = C_r + C_c + C_{rc}$ with respect to σ_{rs} is given by

$$\frac{\partial C}{\partial \sigma_{rs}} = \frac{\partial C_r}{\partial \sigma_{rs}} + \frac{\partial C_c}{\partial \sigma_{rs}} + \frac{\partial C_{rc}}{\partial \sigma_{rs}}$$

where

$$\begin{aligned} \frac{\partial C_r}{\partial \sigma_{rs}} &= \frac{\partial}{\partial \sigma_{rs}} ([\mathsf{vechl}(\boldsymbol{\psi}_r)]^T [\mathsf{vechl}(\boldsymbol{\Omega_r \Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma \Omega_r^T})]) \\ &= [\mathsf{vechl}(\boldsymbol{\psi}_r)]^T [\mathsf{vechl}(\boldsymbol{\Omega}_r \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{rs}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{rs}} \boldsymbol{\Omega}_r^T)] \end{aligned}$$

where $\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{rs}}$ is defined as above.

Similarly

$$\frac{\partial C_c}{\partial \sigma_{rs}} = [\mathsf{vechl}(\boldsymbol{\psi}_c)]^T [\mathsf{vechl}(\boldsymbol{\Omega}_c \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{rs}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{rs}} \boldsymbol{\Omega}_c^T)]$$

and

$$\frac{\partial C_{rc}}{\partial \sigma_{rs}} = [\mathsf{vechl}(\boldsymbol{\psi}_{rc})]^T [\mathsf{vechl}(\boldsymbol{\Omega}_r \boldsymbol{\Omega}_c \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{rs}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{rs}} \boldsymbol{\Omega}_c^T \boldsymbol{\Omega}_r^T)]$$

The derivative of the constraints C with respect to ω_{rij} is given by

$$\frac{\partial C}{\partial \omega_{rij}} = \frac{\partial C_r}{\partial \omega_{rij}} + \frac{\partial C_{rc}}{\partial \omega_{rij}}$$

where

$$\frac{\partial C_r}{\partial \omega_{rij}} = [\mathsf{vechl}(\boldsymbol{\psi}_r)]^T [\mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}} \boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma} \frac{\partial \boldsymbol{\Omega}_r^T}{\partial \omega_{rij}})]$$

and

$$\frac{\partial C_{rc}}{\partial \omega_{rij}} = [\mathsf{vechl}(\boldsymbol{\psi}_{rc})]^T [\mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}} \boldsymbol{\Omega}_c \boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma} \boldsymbol{\Omega}_c^T \frac{\partial \boldsymbol{\Omega}_r^T}{\partial \omega_{rij}})]$$

where $\frac{\partial \boldsymbol{\Omega}_r^T}{\partial \omega_{rij}} = \left(\frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}}\right)^T$ and $\frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}}$ is defined as above.

The derivative of the constraints C with respect to ω_{cij} is given by

$$\frac{\partial C}{\partial \omega_{cij}} = \frac{\partial C_c}{\partial \omega_{cij}} + \frac{\partial C_{rc}}{\partial \omega_{cij}}$$

where

$$\frac{\partial C_c}{\partial \omega_{cij}} = [\mathsf{vech}(\boldsymbol{\psi}_c)]^T[\mathsf{vech}(\frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cij}}\boldsymbol{\Sigma}) - \mathsf{vech}(\boldsymbol{\Sigma}\frac{\partial \boldsymbol{\Omega}_c^T}{\partial \omega_{cij}})]$$

and

$$\frac{\partial C_{rc}}{\partial \omega_{cij}} = [\mathsf{vech}(\boldsymbol{\psi}_{rc})]^T[\mathsf{vech}(\boldsymbol{\Omega}_r\frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cij}}\boldsymbol{\Sigma}) - \mathsf{vech}(\boldsymbol{\Sigma}\frac{\partial \boldsymbol{\Omega}_c^T}{\partial \omega_{cij}}\boldsymbol{\Omega}_r^T)]$$

where $\frac{\partial \boldsymbol{\Omega}_c^T}{\partial \omega_{cij}} = \left(\frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cij}}\right)^T$ and $\frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cij}}$ is defined as above.

The derivative of the constraints C with respect to ψ_{rij} is given by

$$\begin{aligned}\frac{\partial C}{\partial \psi_{rij}} &= [\mathsf{vech}(\frac{\partial \boldsymbol{\psi}_r}{\partial \psi_{rij}})]^T[\mathsf{vech}(\boldsymbol{\Omega}_r\boldsymbol{\Sigma}) - \mathsf{vech}(\boldsymbol{\Sigma}\boldsymbol{\Omega}_r^T)] \\ &= [\mathsf{vech}(\boldsymbol{z}_i\boldsymbol{z}_j^T)]^T[\mathsf{vech}(\boldsymbol{\Omega}_r\boldsymbol{\Sigma}) - \mathsf{vech}(\boldsymbol{\Sigma}\boldsymbol{\Omega}_r^T)]\end{aligned}$$

The derivative of the constraints C with respect to ψ_{cij} is given by

$$\frac{\partial C}{\partial \psi_{cij}} = [\mathsf{vech}(\boldsymbol{z}_i\boldsymbol{z}_j^T)]^T[\mathsf{vech}(\boldsymbol{\Omega}_c\boldsymbol{\Sigma}) - \mathsf{vech}(\boldsymbol{\Sigma}\boldsymbol{\Omega}_c^T)]$$

The derivative of the constraints C with respect to ψ_{rcij} is given by

$$\frac{\partial C}{\partial \psi_{rcij}} = [\mathsf{vech}(\boldsymbol{z}_i\boldsymbol{z}_j^T)]^T[\mathsf{vech}(\boldsymbol{\Omega}_r\boldsymbol{\Omega}_c\boldsymbol{\Sigma}) - \mathsf{vech}(\boldsymbol{\Sigma}\boldsymbol{\Omega}_c^T\boldsymbol{\Omega}_r^T)]$$

The second differentials of the constraints are also required to update the Average Information matrix. A number of these are zero but the non zero differentials are as follows.

The second derivative of the constraints C with respect to ψ_{rij} and ω_{ruv} is given by

$$\frac{\partial C}{\partial \psi_{rij}\partial \omega_{ruv}} = [\mathsf{vech}(\boldsymbol{z}_i\boldsymbol{z}_j^T)]^T[\mathsf{vech}(\frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{ruv}}\boldsymbol{\Sigma}) - \mathsf{vech}(\boldsymbol{\Sigma}\frac{\partial \boldsymbol{\Omega}_r^T}{\partial \omega_{ruv}})]$$

The second derivative of the constraints C with respect to ψ_{cij} and ω_{cuv} is given by

$$\frac{\partial C}{\partial \psi_{cij}\partial \omega_{cuv}} = [\mathsf{vech}(\boldsymbol{z}_i\boldsymbol{z}_j^T)]^T[\mathsf{vech}(\frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cuv}}\boldsymbol{\Sigma}) - \mathsf{vech}(\boldsymbol{\Sigma}\frac{\partial \boldsymbol{\Omega}_c^T}{\partial \omega_{cuv}})]$$

The second derivative of the constraints C with respect to ψ_{rcij} and ω_{ruv} is given by

$$\frac{\partial C}{\partial \psi_{rcij}\partial \omega_{ruv}} = [\mathsf{vech}(\boldsymbol{z}_i\boldsymbol{z}_j^T)]^T[\mathsf{vech}(\frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{ruv}}\boldsymbol{\Omega}_c\boldsymbol{\Sigma}) - \mathsf{vech}(\boldsymbol{\Sigma}\boldsymbol{\Omega}_c^T\frac{\partial \boldsymbol{\Omega}_r^T}{\partial \omega_{ruv}})]$$

The second derivative of the constraints C with respect to ψ_{rcij} and ω_{cuv} is given by

$$\frac{\partial C}{\partial \psi_{rcij}\partial \omega_{cuv}} = [\mathsf{vech}(\boldsymbol{z}_i\boldsymbol{z}_j^T)]^T[\mathsf{vech}(\boldsymbol{\Omega}_r\frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cuv}}\boldsymbol{\Sigma}) - \mathsf{vech}(\boldsymbol{\Sigma}\frac{\partial \boldsymbol{\Omega}_c^T}{\partial \omega_{cuv}}\boldsymbol{\Omega}_r^T)]$$

The second derivative of the constraints C with respect to ψ_{rij} and σ_{uv} is given by

$$\frac{\partial C}{\partial \psi_{rij} \partial \sigma_{uv}} = [\mathsf{vechl}(\boldsymbol{z}_i \boldsymbol{z}_j^T)]^T [\mathsf{vechl}(\boldsymbol{\Omega_r} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}} \boldsymbol{\Omega}_r^T)]$$

The second derivative of the constraints C with respect to ψ_{cij} and σ_{uv} is given by

$$\frac{\partial C}{\partial \psi_{cij} \partial \sigma_{uv}} = [\mathsf{vechl}(\boldsymbol{z}_i \boldsymbol{z}_j^T)]^T [\mathsf{vechl}(\boldsymbol{\Omega_c} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}} \boldsymbol{\Omega}_c^T)]$$

The second derivative of the constraints C with respect to ψ_{rcij} and σ_{uv} is given by

$$\frac{\partial C}{\partial \psi_{rcij} \partial \sigma_{uv}} = [\mathsf{vechl}(\boldsymbol{z}_i \boldsymbol{z}_j^T)]^T [\mathsf{vechl}(\boldsymbol{\Omega}_r \boldsymbol{\Omega}_c \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}} \boldsymbol{\Omega}_c^T \boldsymbol{\Omega}_r^T)]$$

The second derivative of the constraints C with respect to ω_{rij} and σ_{uv} is given by

$$\begin{aligned} \frac{\partial C}{\partial \omega_{rij} \partial \sigma_{uv}} &= [\mathsf{vechl}(\boldsymbol{\psi}_r)]^T [\mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}} \frac{\partial \boldsymbol{\Omega}_r^T}{\partial \omega_{rij}})] \\ &+ [\mathsf{vechl}(\boldsymbol{\psi}_{rc})]^T [\mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}} \boldsymbol{\Omega}_c \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}} \boldsymbol{\Omega}_c^T \frac{\partial \boldsymbol{\Omega}_r^T}{\partial \omega_{rij}})] \end{aligned}$$

The second derivative of the constraints C with respect to ω_{cij} and σ_{uv} is given by

$$\begin{aligned} \frac{\partial C}{\partial \omega_{cij} \partial \sigma_{uv}} &= [\mathsf{vechl}(\boldsymbol{\psi}_c)]^T [\mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cij}} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}} \frac{\partial \boldsymbol{\Omega}_c^T}{\partial \omega_{cij}})] \\ &+ [\mathsf{vechl}(\boldsymbol{\psi}_{rc})]^T [\mathsf{vechl}(\boldsymbol{\Omega}_r \frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cij}} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}}) - \mathsf{vechl}(\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{uv}} \frac{\partial \boldsymbol{\Omega}_c^T}{\partial \omega_{cij}} \boldsymbol{\Omega}_r^T)] \end{aligned}$$

The second derivative of the constraints C with respect to ω_{rij} and ω_{cuv} is given by

$$\frac{\partial C}{\partial \omega_{rij} \partial \omega_{cuv}} = [\mathsf{vechl}(\boldsymbol{\psi}_{rc})]^T [\mathsf{vechl}(\frac{\partial \boldsymbol{\Omega}_r}{\partial \omega_{rij}} \frac{\partial \boldsymbol{\Omega}_c}{\partial \omega_{cuv}} \boldsymbol{\Sigma}) - \mathsf{vechl}(\boldsymbol{\Sigma} \frac{\partial \boldsymbol{\Omega}_c^T}{\partial \omega_{cuv}} \frac{\partial \boldsymbol{\Omega}_r^T}{\partial \omega_{rij}})]$$

9.5 Application to multivariate examples with spatial correlation in row and column directions

To illustrate the 2dMVAR1 model presented in this chapter and it's suitability for modelling the residual covariance structure for multivariate data collected on a 2 dimensional lattice, a number of bivariate analyses have been performed on lucerne yield and persistence data collected at a number of times and sites. In each bivariate data set the yield and persistence measures were made on the same day or within a couple of weeks of each other. In these analyses the 2dMVAR1 model has been used to model the residual spatial and temporal covariance structure. Code has been written in R (R Development Core Team, 2012) to implement these models. This code is presented in Appendix E

(and is referred to as MCARUSgen). In these models the matrix $\mathbf{\Sigma}$ is assumed to be an unstructured symmetric covariance matrix (us) and $\mathbf{\Omega}_r$ and $\mathbf{\Omega}_c$ are fully parameterized non symmetric matrices (gen$\mathbf{\Omega}$).

The results from fitting the 2dMVAR1 models are presented in Table 9.1. Comparisons have been made between the log-likelihood from the 2dMVAR1 model and the separable us(Trait).ar1(Column).ar1(Row) model using REMLRT.

The first data set analysed was the yield and persistence data from Euloma measured on 1/11/2003 and 14/10/2003 for yield and persistence respectively. The spatial correlation parameters estimated from individual analyses of each of these traits are presented in Table 9.1. The spatial correlation parameters in the column direction were reasonably similar between the two traits (0.210 and 0.140 for persistence and yield respectively), while the spatial correlation parameters in the row direction were quite different (0.640 and 0.180 for persistence and yield respectively). These estimates were used as starting values for the diagonal elements of the spatial dependency matrices $\mathbf{\Omega}_c$ and $\mathbf{\Omega}_r$ in the 2dMVAR1 model, while the starting values for the off diagonal elements were set at a small number (0.001).

The log-likelihood obtained from fitting the separable us(Trait).ar1(Column).ar1(Row) model to the Euloma data set in ASReml-R was 54.852 (Table 9.1). The spatial correlation parameter estimates (here common for both traits) from this separable model were 0.18 and 0.52 for the column and row directions respectively. The estimates for the us temporal covariance matrix from this separable model were used as starting values for $\mathbf{\Sigma}$ in the 2dMVAR1 model.

The 2dMVAR1 model converged with a log-likelihood of 66.512, which was a significant improvement in fit than the separable model (REMLRT = 23.32 on 6 df, $P < 0.001$). The parameters estimated from this non-separable model are as follows:

$$\hat{\mathbf{\Sigma}} = \begin{bmatrix} 0.134 & 0.132 \\ 0.132 & 0.840 \end{bmatrix} \hat{\mathbf{\Omega}}_r = \begin{bmatrix} 0.169 & 0.015 \\ -0.435 & 0.709 \end{bmatrix} \hat{\mathbf{\Omega}}_c = \begin{bmatrix} 0.158 & 0.001 \\ -0.039 & 0.207 \end{bmatrix}$$

The spatial dependency parameters for the two traits differ in the row direction with the spatial dependency parameter for persistence on neighbouring plots in the row direction estimated as 0.169 while the spatial dependency between neighbouring plots for yield was 0.709. The parameters in the column direction were more similar.

A similar model fitting process was followed for the remaining eight data sets (Table 9.1). Out of the nine data sets, six showed a significant improvement in model fit with the non-separable 2dMVAR1 residual model than the separable residual model. The parameter estimates from the 2dMVAR1 models are presented in Table 9.2.

In the three analyses that did not show a significant improvement with the 2dMVAR1 residual model, the spatial correlation parameters (as estimated from the individual analyses of each trait) were fairly similar between the two traits for both column and row directions. Where the spatial parameters (from the individual analyses) differed by approximately 0.4 or more in either direction, the non-separable 2dMVAR1 model was a significant improvement on the separable model (see Table 9.1).

Table 9.1: Summary of models fitted for bivariate analysis of yield and persistence at each site using separable and non-separable 2dMVAR1 residual models.

Site	Harvest	Yld Spatial par		Persistence Spatial par		Separable us(Tr):ar1:ar1	2dMVAR1 MCARUSgen	P
		Col	Row	Col	Row	LL	LL	
Euloma	11/03	0.210	0.640	0.140	0.180	54.852	66.512	<0.001
Euloma	02/04	0.315	0.676	0.210	0.165	93.997	112.564	<0.001
Euloma	10/04	0.251	0.607	0.196	0.232	100.566	109.834	0.005
Lead	11/04	0.056	0.709	0.030	0.318	102.595	109.927	0.023
Lead	11/05	0.248	0.520	-0.046	0.314	149.926	154.479	0.168 n.s
Sand	11/03	0.292	0.627	0.234	0.251	127.139	133.729	0.040
TCCI	11/03	0.262	0.174	0.024	0.601	-64.108	-49.252	<0.001
TCCI	12/04	0.359	0.370	0.245	0.460	141.324	142.737	0.830 n.s
Terry	11/03	-0.017	0.270	-0.068	0.207	163.758	165.143	0.837 n.s

9.6 Summary

In this chapter the MVAR1 model of the previous chapter has been extended to the 2 directional (row and column) situation by using the theory of MCAR models. While examples of MCAR models can be found in the literature to model spatio-temporal data in alternate areas such as finance, real-estate, disease epidemics, mostly on irregular lattices and fitted in a Bayesian framework (for example Carlin & Banerjee (2003), Gelfand & Vounatsou, 2003), the covariance structure imposed by these MCAR models is not desirable for our situation of residuals from plots in a field trial. The CAR model has been shown to have a centro-symmetric correlation structure (see Wall (2004) and Assuncao & Krainski, 2009) where plot variances and correlations between neighbouring plots in a regular lattice are not stationary but instead increase towards the centre of the lattice. This structure carries over to the multivariate MCAR case.

Instead the 2dMVAR1 model we have developed imposes the desired covariance structure we require for our situation. That is, the variance matrix is constant for all plots and the matrix of covariances between plots at the same displacement is constant. The model is also easily implemented into the linear mixed model framework with estimation using REML.

The resulting 2dMVAR1 model has been applied to a number of bivariate data sets, modelling the residual spatio-temporal correlation structure in row and column directions. In most cases the non-separable 2dMVAR1 model was a significant improvement in model fit than the comparable separable us(Trait).ar1(Column).ar1(Row) model.

The approach allows for separate spatial dependency parameters to be estimated for each time (or trait) thereby providing a much more flexible range of residual models for multi-harvest (or multi-trait) data. The models allow for further insight into the spatial structure of the data by allowing for different spatial parameters rather than assuming a common spatial correlation parameter across times or between traits, as is the case in

Table 9.2: REML estimates of parameters from 2dMVAR1 residual model fitted to bivariate lucerne yield and persistence data sets at a number of sites and times.

Site	Harvest	2dMVAR1 estimates		
		$\hat{\boldsymbol{\Sigma}}$	$\hat{\boldsymbol{\Omega}}_r$	$\hat{\boldsymbol{\Omega}}_c$
Euloma	11/03	$\hat{\sigma}_{11} = 0.134$	$\hat{\omega}_{r11} = 0.169$	$\hat{\omega}_{c11} = 0.158$
		$\hat{\sigma}_{12} = 0.132$	$\hat{\omega}_{r12} = 0.015$	$\hat{\omega}_{c12} = 0.001$
		$\hat{\sigma}_{21} = 0.132$	$\hat{\omega}_{r21} = -0.435$	$\hat{\omega}_{c21} = -0.039$
		$\hat{\sigma}_{22} = 0.840$	$\hat{\omega}_{r22} = 0.709$	$\hat{\omega}_{c22} = 0.207$
Euloma	02/04	$\hat{\sigma}_{11} = 0.079$	$\hat{\omega}_{r11} = 0.044$	$\hat{\omega}_{c11} = 0.114$
		$\hat{\sigma}_{12} = 0.131$	$\hat{\omega}_{r12} = 0.071$	$\hat{\omega}_{c12} = 0.023$
		$\hat{\sigma}_{21} = 0.131$	$\hat{\omega}_{r21} = -0.140$	$\hat{\omega}_{c21} = -0.045$
		$\hat{\sigma}_{22} = 1.013$	$\hat{\omega}_{r22} = 0.680$	$\hat{\omega}_{c22} = 0.319$
Euloma	10/04	$\hat{\sigma}_{11} = 0.102$	$\hat{\omega}_{r11} = 0.118$	$\hat{\omega}_{c11} = 0.164$
		$\hat{\sigma}_{12} = 0.091$	$\hat{\omega}_{r12} = 0.081$	$\hat{\omega}_{c12} = 0.018$
		$\hat{\sigma}_{21} = 0.091$	$\hat{\omega}_{r21} = 0.108$	$\hat{\omega}_{c21} = 0.024$
		$\hat{\sigma}_{22} = 0.660$	$\hat{\omega}_{r22} = 0.580$	$\hat{\omega}_{c22} = 0.265$
Lead	11/04	$\hat{\sigma}_{11} = 0.093$	$\hat{\omega}_{r11} = 0.304$	$\hat{\omega}_{c11} = 0.019$
		$\hat{\sigma}_{12} = 0.025$	$\hat{\omega}_{r12} = -0.003$	$\hat{\omega}_{c12} = 0.000$
		$\hat{\sigma}_{21} = 0.025$	$\hat{\omega}_{r21} = -0.131$	$\hat{\omega}_{c21} = 0.020$
		$\hat{\sigma}_{22} = 0.669$	$\hat{\omega}_{r22} = 0.700$	$\hat{\omega}_{c22} = 0.079$
Lead	11/05	$\hat{\sigma}_{11} = 0.078$	$\hat{\omega}_{r11} = 0.305$	$\hat{\omega}_{c11} = -0.063$
		$\hat{\sigma}_{12} = 0.023$	$\hat{\omega}_{r12} = -0.016$	$\hat{\omega}_{c12} = 0.024$
		$\hat{\sigma}_{21} = 0.023$	$\hat{\omega}_{r21} = -0.008$	$\hat{\omega}_{c21} = 0.012$
		$\hat{\sigma}_{22} = 0.342$	$\hat{\omega}_{r22} = 0.519$	$\hat{\omega}_{c22} = 0.249$
Sand	11/03	$\hat{\sigma}_{11} = 0.068$	$\hat{\omega}_{r11} = 0.208$	$\hat{\omega}_{c11} = -0.221$
		$\hat{\sigma}_{12} = 0.044$	$\hat{\omega}_{r12} = -0.044$	$\hat{\omega}_{c12} = 0.007$
		$\hat{\sigma}_{21} = 0.044$	$\hat{\omega}_{r21} = -0.062$	$\hat{\omega}_{c21} = 0.010$
		$\hat{\sigma}_{22} = 0.528$	$\hat{\omega}_{r22} = 0.643$	$\hat{\omega}_{c22} = 0.288$
TCCI	11/03	$\hat{\sigma}_{11} = 0.181$	$\hat{\omega}_{r11} = 0.605$	$\hat{\omega}_{c11} = 0.040$
		$\hat{\sigma}_{12} = 0.267$	$\hat{\omega}_{r12} = -0.003$	$\hat{\omega}_{c12} = 0.001$
		$\hat{\sigma}_{21} = 0.267$	$\hat{\omega}_{r21} = 0.794$	$\hat{\omega}_{c21} = -0.338$
		$\hat{\sigma}_{22} = 2.181$	$\hat{\omega}_{r22} = 0.046$	$\hat{\omega}_{c22} = 0.278$
TCCI	12/04	$\hat{\sigma}_{11} = 0.090$	$\hat{\omega}_{r11} = 0.386$	$\hat{\omega}_{c11} = 0.197$
		$\hat{\sigma}_{12} = 0.131$	$\hat{\omega}_{r12} = 0.000$	$\hat{\omega}_{c12} = 0.009$
		$\hat{\sigma}_{21} = 0.131$	$\hat{\omega}_{r21} = 0.011$	$\hat{\omega}_{c21} = -0.236$
		$\hat{\sigma}_{22} = 0.618$	$\hat{\omega}_{r22} = 0.376$	$\hat{\omega}_{c22} = 0.400$
Terry	11/03	$\hat{\sigma}_{11} = 0.094$	$\hat{\omega}_{r11} = 0.228$	$\hat{\omega}_{c11} = -0.019$
		$\hat{\sigma}_{12} = 0.041$	$\hat{\omega}_{r12} = 0.035$	$\hat{\omega}_{c12} = -0.051$
		$\hat{\sigma}_{21} = 0.041$	$\hat{\omega}_{r21} = 0.090$	$\hat{\omega}_{c21} = -0.130$
		$\hat{\sigma}_{22} = 0.236$	$\hat{\omega}_{r22} = 0.222$	$\hat{\omega}_{c22} = -0.011$

the separable models. This flexibility makes more sense biologically as it is likely that the spatial correlation will differ between harvest times and also between traits (as shown in Chapter 3).

The 2dMVAR1 models are not without their potential disadvantages. For large numbers of harvests the number of parameters that are required to be estimated for the spatial dependency matrix ($\mathbf{\Omega}$) becomes very large. Interpretation of the parameters in this spatial dependency matrix may also be difficult. In most cases observed in the examples here, the diagonal terms of $\mathbf{\Omega}$ (the spatial dependency parameters for each trait) are the larger components (in magnitude), with the off-diagonal terms (the spatial dependency parameters between yield on one plot and persistence on the neighbouring plot) being very small. However there are cases where the diagonal terms are quite large and more difficult to interpret. Rather than dissecting the individual spatial and temporal components it is really the full spatial and temporal interaction ($\mathbf{\Omega\Sigma}$) which gives the covariances between neighbouring plots that is of most importance.

The main interest in these analyses for perennial crops is at the genetic level, however providing a better model fit at the residual level with better modelling of the spatial and temporal correlation structures will provide better estimates of genetic effects. The results from fitting these 2dMVAR1 residual models are very promising with the model providing a significant improvement in modelling the residual covariance structure than previously implemented methods. These models will provide a significant improvement in the analysis of multi-harvest and multi-trait data in perennial crops. The models also have the potential to provide significant improvement to multi-trait analyses in annual crops.

Chapter 10

General Discussion

10.1 Overview of thesis

The motivation behind this thesis was to investigate and develop efficient methods for analysing data from variety selection trials in perennial crops. The analysis of such data aims to be as accurate and comprehensive as possible in order for the best variety selections to be made across a range of environments (times and locations).

The main issues involved with the analysis of data from perennial crops over and above that of annual crops, are how to best model the spatial correlation between plots in the field as well as modelling the temporal correlation between multiple harvests, and how to provide insight into variety by time by environment interactions. In this thesis, methods have been proposed that successfully model both the residual spatial and temporal covariance structure and also model variety performance over time. It has been demonstrated that these methods are a significant improvement on the historical approach of individually analysing each harvest time separately.

While it is relatively common practice to model the spatial correlation between plots in annual crop field trials using the spatial analysis approach of Gilmour et al. (1997), these methods have not been widely adopted in the analysis of perennial crop data. Exceptions include Smith et al. (2007) and Stringer & Cullis (2002) in sugarcane, Davik & Honne (2005) in strawberry, Resende et al. (2006) in tea, Murison et al. (2006) in perennial pasture trials, and Jones et al. (2009) in grapes, who all found significant spatial correlation in their trials. In Chapter 3 of this thesis, spatial analysis techniques were applied to a number of data sets from multiple trials and harvests in the perennial crops lucerne and chicory at individual harvest times. The analysis methods were based on a combination of a randomisation (or design) based approach (which includes terms to account for the randomisation used in the experimental design) combined with that of Gilmour et al. (1997) which includes terms for global and extraneous spatial trends and models the local spatial correlation between neighbouring plots using an autoregressive process of order 1 in the row and column directions. This hybrid approach is now recommended (see Smith et al., 2005, Smith et al., 2007 and Beeck et al., 2010) as is using the additional diagnostic plots of Stefanova et al. (2009) of the row and column faces of the semivariogram of the residuals, together with 95% coverage intervals.

The spatial analyses performed on the lucerne and chicory data sets identified a number of spatial trends and significant local spatial correlation between plots and the models were shown to provide a significantly better fit to the data than traditional designed based (e.g. RCB) models. Of the 64 spatial analyses performed, 57 of those exhibited significant spatial trend and/or location spatial correlation. The local spatial correlation parameters were all positive, reflecting that neighbouring plots are more likely to be similar than those further apart. There was no evidence of interplot competition (which would be indicated by negative spatial correlation parameters) as found by Stringer & Cullis (2002) in sugarcane trials. Many harvests showed global or extraneous spatial trends associated with the columns of the trials and at Leadville there was significant global trend in the row direction for all harvests besides the first. This may be due to the harvesting pattern at this trial as the plots were harvested across columns (row 1 first, then row 2 etc).

One of the main aims of the analyses in Chapter 3 was to investigate the levels of spatial correlation found at each harvest time within a trial and to see how the spatial correlation differed between harvests. For some trials the local spatial correlation parameters differed substantially between harvests while at other trials the spatial correlation parameters were more similar. For example, for yield at Euloma the row spatial correlation parameters were quite similar (ranging from $0.59 - 0.63$) while at Tamworth the column spatial correlation parameters were more variable ($0.18 - 0.59$). A REMLRT test between models (for yield) assuming common spatial parameters across harvests versus separate spatial parameters (with no modelling of temporal correlation) showed the separate spatial parameter model to be significantly better than the common spatial model for 3 sites (Tamworth, Leadville and Sandi) while the models were not significantly different for Euloma and Terry Hie Hie (in Chapter 6). This provides new insight as there are no other known published studies investigating the changes in spatial correlation parameters across harvests in perennial crops.

The benefits of fitting these spatial analyses were that spatial trends were identified and spatial correlation parameters estimated for each harvest, providing a better insight into the spatial variation impacting each trial. This insight will aid in future trial design and trial practices. These analyses were successful in identifying spatial terms that will be required in more complex analyses across harvests. The analyses also showed that the spatial correlation is not constant across harvests within some trials and this fact needs to be taken into account in analyses across harvests. The spatial correlation parameters from the individual harvests will provide useful starting values for such analyses. Plots of the residuals from the spatial analyses within a trial also showed that the residuals were correlated between harvest times (evidence of temporal correlation). It must be noted that the trials investigated here involved simple Randomised Complete Block (RCB) designs (which are still commonly found in perennial crops). The abundant global and extraneous spatial trends may be reduced with the implementation of better, more efficient spatial designs (for example Cullis et al., 2006), however it has been found that the autoregressive residual structure still best represents the spatial variation than design terms in more sophisticated designs (see Beeck et al., 2010, Smith et al., 2007, Dutkowski et al., 2002).

While there have been many studies showing that spatial models provide a better model fit to data from field trials it is not apparent that there have been studies performed to evaluate how well the spatial models predict the genetic effects in variety selection trials, especially in the case of perennial crops. A simulation study was performed in Chapter 4 which showed that the spatial models not only fitted the data better, but provided more accurate variety predictions (closer to the true genetic effects) than the RCB models. It is therefore recommended that these spatial analysis methods should be applied in variety selection trials in perennial crops in order to get the most accurate variety predictions.

The other issue investigated in the simulations in Chapter 4 was the merit of fitting a measurement error term in the spatial analyses. A number of authors have recommended the inclusion of a measurement error component including Stefanova et al. (2009), Gilmour et al. (1997), Dutkowski et al. (2002) and Wilkinson et al. (1983) who suggest there are strong statistical and biological reasons for including a measurement error term. However Zimmerman & Harville (1991) suggest that it may be unnecessary to include such a term, based on empirical results found by Besag & Kempton (1986). Cullis et al. (1998) demonstrated that including a measurement error term with an autoregressive separable residual model did not necessarily cause computation problems but there may be problems when the autoregressive correlation parameters are small. The simulation study conducted in Chapter 4 showed that while the addition of a measurement error term to the spatial model resulted in a better model fit in many cases, this did not result in an improvement in variety predictions. There were a large number of situations where the measurement error component was unable to be fitted, with models failing to converge. It was observed that the number of simulations that did not converge increased as the measurement error component increased. It was also observed that the number that did not converge was higher in the simulations with high spatial correlation, than the simulations with moderate spatial correlation, and also generally increased as the genetic variance increased. Due to the difficulty in fitting the measurement error term and given the variety predictions were so similar under models with or without a measurement error term, the recommendation from these analyses is to omit the measurement error term from spatial analyses of individual harvests. However, it is shown in later chapters that there still may be a need to account for measurement error in the spatio-temporal sense when data is analysed across harvests.

While the analyses in Chapters 3 and 4 showed clearly that the spatial analysis methods should be implemented in the analysis of perennial crop data, complications arise because the data is usually obtained from multiple harvest times, and hence there is also a need to simultaneously model the temporal covariance structure between measurements. In Chapter 5 models were presented to model the residual spatial and temporal correlation between multiple harvests within a single trial, using separable residual models, incorporating global and extraneous spatial trends identified using the spatial analysis techniques of Chapter 3. In Chapter 6 these models were applied to the analysis of data from a single site of the lucerne yield and persistence data.

Prior to this thesis, separable spatial by temporal residual models were investigated

in perennial crops for only a very small number of harvests, by Smith et al. (2007). Their models were not directly applicable to the data sets in this thesis due to the larger numbers of harvests. In this thesis the separable residual models have been extended to cater for greater numbers of measurements, often typical in perennial variety selection trials, using an approach based on the model of Diggle (1988) for longitudinal analysis. Diggle's model (with terms for plot (or unit) effects, measurement error and serial temporal correlation) has been extended to the spatial context for repeated measurements on a row column lattice to create a new approach to modelling the residual covariance structure in multi-harvest data. These extended separable models provide a novel set of flexible residual models that accommodate both spatial and temporal correlation that has been shown to be evident in variety selection trials in perennial crops.

The concept of using separable spatial by temporal models for the residual correlation structure is appealing as it provides easy interpretation and clear insight into both the spatial component and the temporal component contributing to the overall residual covariance structure. The other advantage lies in the numerous computational benefits (discussed in Galecki, 1994 and Naik & Rao, 2001) of separable models. The limitations in the separable residual models are that they may not model the interaction between spatial and temporal components adequately. The added benefits of the new proposed separable residual models based on the model of Diggle (1988) are that they include terms for all the major identified components of spatial and temporal correlation and they can be fitted to data sets with large numbers of harvests.

The analyses in Chapter 6 showed that temporal correlation was clearly evident and important to model. The three components of temporal correlation discussed in Diggle (1988) were found across the two examples, namely plot effects, measurement error and serial correlation. The spatial extensions to the plot term and serial correlation component were significant however the spatial extension to the measurement error was not. The temporal serial component was best modelled in the yield analysis using an antedependence model which is similar to Núñez-Antón & Zimmerman (2000) where antedependence models were found to best model the serial correlation in their examples of longitudinal data. The autoregressive ar1 model was best for modelling the serial correlation in persistence, which is similar to the models of Patterson & Lowe (1970) and Bjornsson (1978) where they used models with autoregressive errors to model perennial crop data.

At the genetic (or variety) level, models were presented for modelling the genetic effects over time. In this thesis different models have been proposed for different situations and for different types of traits observed in perennial crops. One such model is the Factor Analytic (fa) model (which is routinely used in the analysis of MET data in annual crops), presented here to model variety effects at the different harvest times. Interpretation from this model and investigation into variety by harvest interactions may be enhanced by implementing a cluster analysis approach to the between harvest genetic correlation structure estimated from the factor analytic (fa) model. While this approach has been implemented for MET data in annual crops (Cullis et al., 2010) and forestry (Hardner et al., 2010), it is a novel approach to multi-harvest data in perennial crops. The fa model has advantages in it's

parsimony, however it may have limitations in this application in that it does not take into account the time ordering of the harvests. The groups of harvests formed from the cluster analysis on the results from the fa model were slightly difficult to interpret.

Another method presented for modelling the genetic effects over time was the random regression model. In this thesis the approach was to model the underlying trend over time using a cubic smoothing spline and then the variety deviations from this trend were modelled using random regressions. In the example data set of lucerne persistence, analysed in Chapter 6, linear random regressions were found sufficient to model the deviations from the overall underlying trend but in other situations cubic smoothing spline random regressions may be required and these models were also presented. The models were successful in modelling the persistence response for each variety over time and predictions made for the time taken to reach a certain level of persistence. These models provide an extension to the models for genetic effects in perennial crops presented by Smith et al. (2007) in that traits are able to be modelled using smooth functions over time. This allows further insight into the variety responses over time and allows for predictions to be made at any time in the trial, not just at the harvest times. Not all traits will be suitable for this type of modelling but growth traits such as trunk circumference, tree height, canopy diameter are just some of the other potential traits in perennial crops that could benefit from this type of model.

As variety selection in perennial crops is usually based on a number of trials grown in different environments (METs), approaches for analysis need to be applicable to data across sites. The methods introduced for analysis of multi-harvest data at a single site were extended to the multi-environment situation in Chapter 7. These approaches provided insight into variety by environment (time and location) interaction. These models extended the ideas of Smith et al. (2007) to include the new residual models developed in Chapter 5, enabling examples with large numbers of harvests to be analysed and genetic effects to be modelled over time. At the genetic level the factor analytic models and random regression models were applied to the variety by harvest by trial effects. One possible limitation with the factor analytic genetic models applied to lucerne yield is that there is no structure of harvests within trials being taken into account. This is mainly to do with the structure of the example data set with unequal harvest numbers and times being conducted at each trial. More structure may be included with alternative data sets. A result of this is that the groupings from the cluster analysis were slightly difficult to interpret. The random regression genetic models allowed insight into variety by environment interaction in establishment and persistence with some varieties shown to perform well at some sites and not so well at others.

While the separable residual models presented in Chapters 5, 6 and 7 provided an approach to modelling the residual spatial and temporal covariance structure in perennial crops that was significantly better than assuming independence between harvests, the separable models are quite restrictive. The models assume that the spatial correlation parameters are constant across all harvests, which may be a condition that may not hold in practice. In the spatial analyses of individual harvests in Chapter 3 the spatial correlation

parameters were found to vary substantially across the harvests at a number of sites. This is to be expected given the different environmental conditions and the different stage of growth of the plants at the different harvests. Hence, it was identified that a model that allowed for different spatial correlation parameters across the harvests, such as a non-separable spatio-temporal model that could be implemented into the linear mixed model may be more desirable.

Initially in Chapter 8, the multivariate autoregressive (MVAR1) model (with the h harvest measurements forming the multi-variate vector on each plot) was investigated as a possible suitable non-separable model for modelling the residual spatial and temporal covariance structure in perennial crop data. The model was investigated for data from plots spatially correlated in a single direction, for example plots in a long row or plots in a lattice where spatial correlation was evident in only one direction (e.g. row but not column direction).

The multivariate autoregressive model can be found in the literature, most commonly to model time series applications, where a number of traits have been measured on units at repeated times, for example Harison et al. (2003) modelling brain activity over time and Hytti et al. (2006) modelling cardiovascular measurements over time. The temporal component is therefore assumed to follow an autoregressive of order 1 (ar1) process. These published applications differ from our approach in that we require a model that models the spatial correlation between plots using the ar1 process while the multi-variate vector of measurements on each plot consists of the multi-harvest (or multi-trait) measurements. This application raises an additional issue than the standard multivariate autoregressive model over time, as time has a distinct order (first to last) whereas we require our model to be the same whether we move in either direction across the field. We also require the model to be fitted as the residual component in a linear mixed model using REML. An extensive literature search failed to find any such applications or theory developed for multivariate autoregressive models suitable for spatially correlated plots in a field, with estimation using REML.

The MVAR1 model formulation developed in this thesis allows the model to be invariant to direction. This is facilitated by a symmetry constraint that is applied to the model. This constraint makes the model suitable for data measured on spatially correlated plots in the field, differing from the usual application of multivariate autoregressive models in time series applications.

Code was written in R (R Development Core Team, 2012) to implement these MVAR1 models using REML for estimation and they were applied to a selection of multi-harvest and multi-trait examples using the lucerne yield and persistence data. These models performed significantly better than their comparable separable model forms implemented in Chapters 5 and 6. Prior to this thesis MVAR1 models had not been used to model the spatial and temporal covariance structure in multi-harvest data, or the spatial and between trait covariance structure in multi-trait data, in field trials.

The MVAR1 residual model has many benefits over the separable residual models in Chapters 5 and 6 including the ability to model the interaction between spatial and tem-

poral processes rather than assuming them to be separable. The model has been shown to provide a better fit than the separable model in a variety of examples in the thesis and so is likely to provide more accurate estimates of genetic effects. The downside to this model is that it is only applicable to plots in one direction (for example in the row direction) and cannot directly be extended to the two directional lattice case. Another issue is that the parameters are less interpretable than in the separable models. The model is defined in terms of spatial dependency parameters and conditional correlations rather than spatial correlations as in the separable model. Another potential limitation is the large number of parameters that may need to be estimated as the number of harvests becomes large. For example in the 10 harvest example of yield at Terry Hie Hie presented in the thesis, there were 100 parameters that needed to be estimated for the spatial dependency matrix. This was made possible by using good starting values from the initial individual harvest analyses in Chapter 3. By taking the individual harvest spatial parameters as starting values for the diagonal elements of the spatial dependency matrix and allowing the off-diagonals to be very small and constant for a reduced number of harvests, the model was slowly built up, one harvest at a time for the symmetrical spatial dependency model. Once the full symmetric spatial dependency matrix was estimated these parameter values were used as starting values to fit the full general spatial dependency structure. Using this process the model was able to be successfully fitted to the 10 harvest case, however the model may be difficult to fit for examples with many more harvests. It is of particular interest that for this data set at Terry Hie Hie, the non-separable model provided a significantly better fit than the separable form, even though the spatial correlation parameters (from individual harvest analyses) did not differ greatly between harvests. In Chapter 6 it was shown that for this data set, allowing for separate spatial parameters for each harvest rather than common spatial parameters (prior to modelling the temporal correlation) was not a significant improvement. There is clearly benefit in the non-separable models even for trials where spatial parameters may not differ greatly, and it is the improved modelling of the interaction between spatial and temporal processes that is important.

To be entirely suitable for modelling the covariance structure in perennial crop field trials planted in a row by column lattice in the field, the non-separable MVAR1 residual models needed to be extended to 2 spatial dimensions. To extend the model to accommodate spatial correlation in both the row and column directions was not straightforward. In the univariate case the covariance structure on a row column lattice can simply be modelled by taking the kronecker product of two autoregressive ar1 processes (ar1(Column)$\otimes$ar1(Row)) but this is not possible using the MVAR1 model in the multivariate case, as the MVAR1$\otimes$MVAR1 model is of the wrong dimension. An alternative approach was investigated based on the theory of Multivariate Conditional Autoregressive Models (MCAR) which are defined in terms of neighbouring plots in two directions on a two dimensional lattice. This allowed the MVAR1 model to be extended to a 2dMVAR1 model which was suitable for modelling the covariance structure for data from spatially correlated plots in both row and column directions, as well as being correlated across times (or traits). Code for fitting the resulting 2dMVAR1 model in a linear mixed model using

REML was written in R (R Development Core Team, 2012) and applied to a number of bivariate lucerne yield and persistence examples. The model was a significant improvement in fit for most applications than the separable models of the previous chapters.

MCAR models can be found in the literature to model multivariate data on a lattice (mostly irregular lattices) over space and time, for example Billheimer et al. (1997) for modelling biological ecosystems, Jin et al. (2005) for disease mapping, and Carlin & Banerjee (2003) for modelling cancer survival data, all using Bayesian methods for estimation. One of the most important issues in MCAR models is how the spatial dependency structure is defined. The 2dMVAR1 model presents a different formulation of the MCAR model than other published cases in that it provides a MVAR1 structure to hold in the row and column directions, resulting in a multivariate analogue of the univariate ar1(Column)$\otimes$ar1(Row) spatial process. It also is fitted in a linear mixed model with estimation of parameters by REML, unlike most other applications of MCAR models which are fitted using Bayesian methods.

A common formulation of the spatial dependency structure in MCAR models found in the literature is to define the spatial dependency matrices for each plot to be a diagonal matrix weighted for the number of neighbours of each plot (for example Billheimer et al., 1997, Gelfand & Vounatsou, 2003 and Carlin & Banerjee, 2003). This model gives a very restrictive form for the spatial dependency, which in some ways is similar to the special case in Section 8.3.2 of a separable model, however it does not follow the ar1 process that we have assumed. A more flexible and sophisticated model was presented by Sain & Cressie (2007) which allows for a non-symmetric spatial dependency matrix. This results in non-symmetric cross correlations (which implies the correlation between trait 1 on plot i and trait 2 on neighbouring plot $i+1$ does not have to equal the correlation between trait 2 on plot i and trait 1 on plot $i+1$). This model assumes a common general form of the spatial dependency for all plots (with small departures allowed due to differing degrees of variability at each plot), which results in an undesirable covariance structure for the application of modelling multi-harvest data in field trials. This is similar to the covariance issues discussed by Wall (2004) and Assuncao & Krainski (2009) implied in univariate CAR models. In contrast the 2dMVAR1 model presented in this thesis assumes a different form for spatial dependency in the row and column directions and for corner, edge and internal plots. This results in the covariance structure that is required to model the residuals from multi-harvest data in field trials. The 2dMVAR1 model also allows for non-symmetric spatial dependency matrices, but unlike Sain & Cressie (2007) the model requires the cross correlations to be equal, as the results are required to be the same in either direction across the field. The symmetry conditions introduced in Section 9.4 ensure this condition.

The flexibility of the 2dMVAR1 model comes with some disadvantages, similar to those identified for the MVAR1 model. The main issue is the large number of parameters that need to be estimated. For example just in the spatial dependency matrices for a 10 harvest example there would be 200 spatial dependency parameters to be estimated in the general case. The examples presented in this thesis were based on multi-trait bivariate data sets

which did not suffer from this limitation. The models fitted quickly and without any convergence issues. The speed and success of model fitting will be aided by good initial starting values which can be obtained from the initial analyses of individual harvest (or trait) data. The other potential limitation is the difficulty in interpretation of spatial and temporal parameters (as compared to a separable residual model) as the model is modelling the spatial by temporal interaction rather than just the separate individual components.

Despite the potential limitations, the 2dMVAR1 models have great potential. The models developed in this thesis provide a novel, flexible approach for modelling the residual spatial and temporal correlation from multi-harvest and multi-trait data from perennial crop variety selection trials, that are a significant improvement on previously implemented models. The models allow modelling of the interaction between spatial and temporal processes and allow for differing spatial correlations for different harvests (or traits) which makes more sense biologically than assuming common spatial parameters. More efficient modelling of the residual covariance structure will result in more accurate predictions of the genetic effects. Hence, together with the proposed genetic models presented in this thesis, analyses implementing these models have the potential to have a significant impact on improving the accuracy of variety selections made in perennial crops.

The application of these models is not only restricted to perennial crops, as these models also have the potential to be very useful for analyses of multi-trait data in annual crops.

10.1.1 Future research

Multi-harvest / multi-trait

While the MVAR1 model in this thesis was presented for modelling the *residual* covariance structure in multi-harvest data in perennial crops, it is also clear how this model could be useful at the *genetic* level for multi-trait, multi-harvest data. In this situation the multi-trait measurements made at each time would form the multivariate vector and the harvest times the "units" on which the multivariate measurements are made. There is a possible issue with the symmetry constraints discussed in Chapter 7 as while these constraints may not be strictly needed (as time is an ordered variable, in contrast to plots in a row) it would seem necessary to impose the constraints to be able to write the model in a simple form. Code would need to be written for this application but the theory has been developed in this thesis. Suitable non-separable models for multi-harvest, multi-trait data have been proposed for the residual level in Chapter 8.

METs

The MVAR1 model may also prove a suitable non-separable model for modelling MET multi-harvest data at the genetic level. This would be suitable only for balanced data sets where each site is measured at the same harvest times. In many situations the harvest times are not conducted at the same times across sites, for example in the lucerne

data set, and hence this approach may not be applicable for MET analyses of such data. Separable harvest by site models are also not practical in this situation. More research into modelling the genetic effects over time for such unbalanced MET data sets would be desirable.

More parsimonious forms for Ω in MVAR1 and 2dMVAR1 models

The non-separable MVAR1 and 2dMVAR1 models developed in this thesis require suitable forms for $\mathbf{\Omega}$ (the dependency matrix linking observations across times or traits). For multi-trait data the suitable form for $\mathbf{\Omega}$ is clearly a general form with full parameterisation, but for the multi-harvest case there may be more parsimonious forms of $\mathbf{\Omega}$ that may be more suitable. In this thesis models were successfully fitted using a symmetric form for $\mathbf{\Omega}$ but further research may reveal more suitable forms that require less parameters. More parsimonious forms will aid in the speed of model fitting.

Simulations for non-separable residual models

It has been shown that the non-separable 2dMVAR1 residual models outperformed their comparable separable models in most of the analyses performed here. The examples used in this thesis were mostly with data sets showing quite different spatial correlation parameters for the different harvests or traits. It would be of interest to do a detailed simulation study for a range of spatial correlations and temporal (or between trait) correlations to see how these models perform and also together with more complex genetic models to see the impact of these models on the prediction of variety effects.

Conclusion

In conclusion, the models presented in this thesis provide new approaches for the analysis of variety selection trials that will be of significant practical benefit for both multi-harvest and multi-trait data in perennial crops. Further research into the topics raised in this discussion may add further insight into these approaches and may allow them to be introduced to a wider range of applications.

Appendix

A Matrix results

Result A.1

If $\boldsymbol{A}$ is a $m \times m$ matrix, $\boldsymbol{B}$ is a $m \times n$ matrix, $\boldsymbol{C}$ is a $n \times m$ matrix and $\boldsymbol{D}$ is a $m \times m$ matrix

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}$$

Result A.2

$$(\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1} = \boldsymbol{G}\boldsymbol{Z}^T\boldsymbol{H}^{-1}$$

Proof
Using Result A.1

$$\begin{aligned}
\boldsymbol{Z}^T\boldsymbol{H}^{-1} &= \boldsymbol{Z}^T(\boldsymbol{R} + \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T)^{-1} \\
&= \boldsymbol{Z}^T\boldsymbol{R}^{-1} - \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1} \\
&= ((\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1}) - \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z})(\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1} \\
&= \boldsymbol{G}^{-1}(\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1}
\end{aligned}$$

Therefore

$$\boldsymbol{G}\boldsymbol{Z}^T\boldsymbol{H}^{-1} = (\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1}$$

Result A.3

If $\boldsymbol{A}$ is a $n \times n$ matrix, $\boldsymbol{a}$ is a $n \times 1$ vector and $\boldsymbol{B}$ is a $n \times n$ matrix whose elements are functions of η_i and is non-singular then the following hold:

- $\partial \boldsymbol{A}^T = (\partial \boldsymbol{A})^T$
- $\frac{\partial \boldsymbol{a}^T\boldsymbol{A}}{\partial \boldsymbol{a}} = \boldsymbol{A}$
- $\frac{\partial \boldsymbol{a}^T\boldsymbol{A}\boldsymbol{a}}{\partial \boldsymbol{a}} = 2\boldsymbol{A}\boldsymbol{a}$
- $\frac{\partial \boldsymbol{B}^{-1}}{\partial \eta_i} = -\boldsymbol{B}^{-1}\frac{\partial \boldsymbol{B}}{\partial \eta_i}\boldsymbol{B}^{-1}$
- $\frac{\partial log|\boldsymbol{B}|}{\partial \eta_i} = tr(\boldsymbol{B}^{-1}\frac{\partial \boldsymbol{B}}{\partial \eta_i})$

also

- $\frac{\partial \boldsymbol{P}}{\partial \eta_i} = -\boldsymbol{P}\frac{\partial \boldsymbol{H}}{\partial \eta_i}\boldsymbol{P}$ where $\boldsymbol{P} = \boldsymbol{H}^{-1} - \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}$

Result A.4

Conditional mean and variance.
If

$$\begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$$

then

$$\boldsymbol{y_1}|\boldsymbol{y_2} \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{y_2} - \boldsymbol{\mu_2}), \Sigma_{11} - \Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{21})$$

Result A.5

If the matrices $\boldsymbol{X}$ and $\boldsymbol{L}_2$ satisfy $\boldsymbol{L}_2^T\boldsymbol{X} = 0$ and $\boldsymbol{H}$ is positive definite then

$$\boldsymbol{H} - \boldsymbol{H}\boldsymbol{L}_2(\boldsymbol{L}_2^T\boldsymbol{H}\boldsymbol{L}_2)^{-1}\boldsymbol{L}_2^T\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T$$

Proof See Verbyla (1990)

Result A.6

The Kronecker product of two matrices $\boldsymbol{A}$ $(p \times q)$ and $\boldsymbol{B}$ $(n \times m)$ is defined as the $pn \times qm$ matrix

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \dots & a_{1q}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \dots & a_{2q}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1}\boldsymbol{B} & a_{p2}\boldsymbol{B} & \dots & a_{pq}\boldsymbol{B} \end{bmatrix}$$

B Variance Models for R and G

The Residual and Genetic variance matrices $\boldsymbol{R}$ and $\boldsymbol{G}$ as specified in the mixed model 2.2.1, can have many possible forms. The following section outlines some of the possibilities.

Autoregressive (AR) models

A zero mean, p^{th} order autoregressive process (AR(p)) may be written as

$$y_t = \sum_{s=1}^{p} \phi_s y_{t-s} + \epsilon_t \tag{B.1}$$

$t = 1, 2, \dots T$, where the ϕ's are the autoregressive parameters and the errors ϵ_t are assumed to be normally distributed (and defined below).

We consider the cases where $p = 1$, $p = 2$, $p = 3$ and general p, and present the covariance matrix, inverse covariance matrix and matrix of derivatives with respect to ϕ_s. In all cases the matrices are symmetric and hence the lower half (given) is sufficient to determine the complete matrix.

To form these matrices we need to use the following results for the autocorrelation function (see above) between observations at lag k apart. Box & Jenkins (1970) show that by multiplying both sides of Equation(B.1) by y_{t-k} and taking expected values gives

$$\begin{aligned} E[(y_{t-k}y_t)] &= E[\phi_1 y_{t-k}y_{t-1} + \phi_2 y_{t-k}y_{t-2} + \dots + \phi_p y_{t-k}y_{t-p} + y_{t-k}\epsilon_t] \\ &= \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \dots \phi_p\gamma_{k-p} = \gamma_k \end{aligned} \tag{B.2}$$

where γ_k is the autocovariance between observations k apart. Note $E[y_{t-k}\epsilon_t] = 0$ as ϵ_t is uncorrelated with previous values of the process. If we now divide through by γ_0 then we

obtain an equation for the autocorrelation function at lag k.

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p} \tag{B.3}$$

where $\rho_0 = 1$. If we take $k = 0$ in (B.2) we note that $E[y_{t-k}\epsilon_t] = E[\epsilon_t^2] = \sigma^2$. Hence

$$\gamma_0 = \phi_1\gamma_{-1} + \phi_2\gamma_{-2} + \ldots + \phi_p\gamma{-p} + \sigma^2$$

If we divide through by $\gamma_0 = \sigma_y^2$ we obtain the following expression for the variance of y_t.

$$\sigma_y^2 = \frac{\sigma^2}{1 - \phi_1\rho_1 - \phi_2\rho_2 - \ldots - \phi_p\rho_p} \tag{B.4}$$

AR(1)

A first order autoregressive process $\{y_t\}$ with mean zero is defined by

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \epsilon_t \\ y_1 &= \epsilon_1 \end{aligned}$$

where

$$\epsilon_t \sim N(0, \sigma^2)$$

for $t > 1$. The condition $|\phi_1| < 1$ is assumed for the process to be stationary.
Using (B.3) and (B.4) we have $\rho_1 = \phi_1$ and the variance of y_t is given by

$$var(y_t) = \frac{\sigma^2}{1 - \rho_1\phi_1} = \frac{\sigma^2}{1 - \phi_1^2}.$$

It follows that

$$\epsilon_1 \sim N(0, \frac{\sigma^2}{1 - \phi_1^2}).$$

The process has covariance matrix given by $\sigma^2\mathbf{\Sigma}$ where the $(i,j)^{th}$ element of $\mathbf{\Sigma}$ is given by

$$\Sigma_{ij} = \frac{1}{1 - \phi_1^2} \begin{cases} 1 & \text{for } i = j \\ \phi_1 & |i - j| = 1 \\ \phi_1\Sigma_{i-1,j} & i > j + 1 \end{cases}$$

Therefore the matrix $\mathbf{\Sigma}$ is given by

$$\mathbf{\Sigma} = \frac{1}{1 - \phi_1^2} \begin{bmatrix} 1 & \phi_1 & \phi_1^2 & \ldots & \phi_1^{T-1} \\ \phi_1 & 1 & \phi_1 & \ldots & \phi_1^{T-2} \\ \phi_1^2 & \phi_1 & 1 & \ldots & \phi_1^{T-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \phi_1^{T-1} & \ldots & \phi_1^2 & \phi_1 & 1 \end{bmatrix}$$

Using the results of Verbyla (1985) we can show that the inverse covariance matrix is

given by $\frac{1}{\sigma^2}\boldsymbol{\Sigma}^{-1}$, where

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 1 & -\phi_1 & 0 & \dots & 0 \\ -\phi_1 & 1+\phi_1^2 & -\phi_1 & \dots & 0 \\ 0 & -\phi_1 & 1+\phi_1^2 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -\phi_1 & 1 \end{bmatrix} \tag{B.5}$$

We can write the inverse using the following closed form expression

$$\frac{1}{\sigma^2}\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2}(\boldsymbol{I}_T + \phi_1^2\boldsymbol{E}_1 - \phi_1\boldsymbol{F}_1)$$

where $\boldsymbol{I}_T$ is the $T \times T$ identity matrix, $\boldsymbol{E}_1$ is like the identity matrix but with the first and last one set to zero, and $\boldsymbol{F}_1$ has ones along the upper and lower 1^{st} minor diagonals and zeroes elsewhere.
We can also write the inverse covariance matrix in the form of

$$\frac{1}{\sigma^2}\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2}(\boldsymbol{M}^{-1} - \boldsymbol{W})$$

where $\boldsymbol{M}$ is a diagonal matrix and $\boldsymbol{W}$ is a neighbour matrix with zeros on the main diagonal and non zero elements on the diagonals one element removed from the main diagonal.
The matrices are given by

$$\boldsymbol{M}^{-1} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1+\phi_1^2 & 0 & \dots & 0 \\ 0 & 0 & 1+\phi_1^2 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{W} = \begin{bmatrix} 0 & \phi_1 & 0 & \dots & 0 \\ \phi_1 & 0 & \phi_1 & \dots & 0 \\ 0 & \phi_1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \phi_1 \\ 0 & \dots & 0 & \phi_1 & 0 \end{bmatrix} = \phi_1 \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix}$$

So that

$$\boldsymbol{M}^{-1} = \boldsymbol{I}_T + \phi_1^2\boldsymbol{E}_1 \tag{B.6}$$
$$\boldsymbol{W} = \phi_1\boldsymbol{F}_1 \tag{B.7}$$

If the matrix of derivatives of $\boldsymbol{\Sigma}$ is required we can use the result $\frac{\partial\boldsymbol{\Sigma}}{\partial\phi} = -\boldsymbol{\Sigma}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\phi}\boldsymbol{\Sigma}$. It is

easy to calculate the derivatives of $\mathbf{\Sigma}^{-1}$ with respect to ϕ_1, using (B.6) and (B.7), as

$$\begin{aligned}\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \phi_1} &= \frac{\partial \boldsymbol{M}^{-1}}{\partial \phi_1} - \frac{\partial \boldsymbol{W}}{\partial \phi_1} \\ &= 2\phi_1 \boldsymbol{E}_1 - \boldsymbol{F}_1\end{aligned}$$

The logarithm of the determinant (log det) of $\mathbf{\Sigma}$ is given by

$$\log|\mathbf{\Sigma}| = \log(\frac{1}{1-\phi_1^2}) = -\log(1-\phi_1^2)$$

.

AR(2)

The second order autoregressive process is given by

$$\begin{aligned}y_{t+2} &= \phi_1 y_{t+1} + \phi_2 y_t + \epsilon_{t+2} \\ y_1 &= \epsilon_1 \\ y_2 &= \phi_1 y_1 + \epsilon_2\end{aligned}$$

where $|\phi_1 \pm \phi_2| < 1, |\phi_1| < 1$ and $|\phi_2| < 1$ are assumed to ensure the process is stationary. Using (B.3) the autocorrelation functions of order 1 and 2 (ρ_1 and ρ_2) are given by

$$\rho_1 = \phi_1 + \phi_2 \rho_1$$

hence

$$\rho_1 = \frac{\phi_1}{1-\phi_2}$$

and

$$\rho_2 = \phi_1 \rho_1 + \phi_2$$

hence

$$\rho_2 = \frac{\phi_1^2}{1-\phi_2} + \phi_2$$

.

The ϵ's are defined as follows

$$\epsilon_t \sim N(0, \sigma^2)$$

for $t > 2$

From B.4 the variance of y_t is given by

$$Var(y_t) = \frac{\sigma^2}{(1-\phi_1\rho_1 - \phi_2\rho_2)}$$

It follows that

$$\begin{aligned}\epsilon_1 &\sim N(0, \frac{\sigma^2}{(1-\phi_1\rho_1 - \phi_2\rho_2)} \\ \epsilon_2 &\sim N(0, \frac{\sigma^2}{(1-\phi_1\rho_1 - \phi_2\rho_2)}(1-\phi_1^2))\end{aligned}$$

The process has covariance matrix given by $\sigma^2\mathbf{\Sigma}$ where the $(i,j)^{th}$ element of $\mathbf{\Sigma}$ is

given by

$$\Sigma_{ij} = \frac{1}{(1-\phi_1\rho_1-\phi_2\rho_2)} \begin{cases} 1 & i=j \\ \frac{\phi_1}{1-\phi_2} & |i-j|=1 \\ \phi_1\Sigma_{i-1,j}+\phi_2\Sigma_{i-2,j} & i>j+1 \end{cases}$$

The inverse covariance matrix is given by $\frac{1}{\sigma^2}\Sigma^{-1}$ where

$$\Sigma^{-1} = \boldsymbol{I}_T + \sum_{j=1}^{2}\phi_j^2\boldsymbol{E}_j - \sum_{j=1}^{2}\phi_j\boldsymbol{F}_j + \phi_1\phi_2\boldsymbol{G}_{1,1}$$

where $\boldsymbol{I}_T$ is the $T \times T$ identity matrix, $\boldsymbol{E}_j$ is like the identity matrix but with the first and last j ones set to zero, $\boldsymbol{F}_j$ has ones along the upper and lower j^{th} minor diagonals and zeroes elsewhere and $\boldsymbol{G}_{1,1} = \boldsymbol{E}_1\boldsymbol{F}_1\boldsymbol{E}_1$ which equals $\boldsymbol{F}_1$ except the top and bottom 1 ones along the 1^{st} minor diagonals are replaced by zeroes.

Explicitly, the inverse covariance matrix is given by

$$\Sigma^{-1} = \begin{bmatrix} 1 & & & & & \\ -\phi_1 & 1+\phi_1^2 & & & & \\ -\phi_2 & -\phi_1+\phi_1\phi_2 & 1+\phi_1^2+\phi_2^2 & & & \\ 0 & -\phi_2 & -\phi_1+\phi_1\phi_2 & 1+\phi_1^2+\phi_2^2 & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \\ & & -\phi_2 & -\phi_1+\phi_1\phi_2 & 1+\phi_1^2 & \\ 0 & \dots & 0 & -\phi_2 & -\phi_1 & 1 \end{bmatrix}$$

We can also write the inverse covariance matrix in the form of

$$\frac{1}{\sigma^2}\Sigma^{-1} = \frac{1}{\sigma^2}(\boldsymbol{M}^{-1} - \boldsymbol{W})$$

where

$$\boldsymbol{M}^{-1} = \boldsymbol{I}_T + \sum_{j=1}^{2}\phi_j^2\boldsymbol{E}_j \tag{B.8}$$

$$\boldsymbol{W} = \sum_{j=1}^{2}\phi_j\boldsymbol{F}_j - \phi_1\phi_2\boldsymbol{G}_{1,1} \tag{B.9}$$

If the matrix of derivatives of $\boldsymbol{\Sigma}$ is required we can use the result $\frac{\partial\boldsymbol{\Sigma}}{\partial\phi} = -\boldsymbol{\Sigma}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\phi}\boldsymbol{\Sigma}$. It is easy to calculate the derivatives of $\boldsymbol{\Sigma}^{-1}$ with respect to ϕ_1 and ϕ_2, using (B.8) and (B.9), as

$$\begin{aligned} \frac{\partial\Sigma^{-1}}{\partial\phi_1} &= \frac{\partial M^{-1}}{\partial\phi_1} - \frac{\partial W}{\partial\phi_1} \\ &= 2\phi_1\boldsymbol{E}_1 - \boldsymbol{F}_1 + \phi_2\boldsymbol{G}_{1,1} \end{aligned}$$

$$\begin{aligned} \frac{\partial\Sigma^{-1}}{\partial\phi_2} &= \frac{\partial M^{-1}}{\partial\phi_2} - \frac{\partial W}{\partial\phi_2} \\ &= 2\phi_2\boldsymbol{E}_2 - \boldsymbol{F}_2 + \phi_1\boldsymbol{G}_{1,1} \end{aligned}$$

The logarithm of the determinant (log det) of Σ is given by Siddiqui (1958)

$$log|\Sigma| = log(\frac{1}{(1+\phi_2)^2[(1-\phi_2)^2 - \phi_1^2]})$$

which can also be obtained using Haddad (1998) from the roots (r_1 and r_2) of the characteristic equation

$$z^2 - \phi_1 z - \phi_2 = 0$$

that is

$$r_1 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

$$r_2 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

and using the following expression for the determinant of Σ

$$det(\Sigma) = \prod_{i,j=1}^{2} (1 - r_i r_j)^{-1}$$

AR(p)

The p^{th} order autoregressive process (AR(p)) may be written as

$$y_t = \sum_{s=1}^{p} \phi_s y_{t-s} + \epsilon_t$$

$t = 1, 2, \ldots, T$, where

$$\epsilon_t \sim N(0, \sigma^2)$$

and

$$Var(y_t) = \frac{\sigma^2}{1 - \sum_{i=1}^{p} \phi_i \rho_i}$$

where ρ_i is the autocorrelation between observations i apart, defined previously.
We can write the general form for the inverse covariance matrix of the AR(p) model as

$$\Sigma^{-1} = \frac{1}{\sigma^2} \left(I_T + \sum_{j=1}^{p} \phi_j^2 E_j - \sum_{j=1}^{p} \phi_j F_j + \sum_{j=1}^{p-1} \sum_{i=1}^{p-j} \phi_i \phi_{i+j} G_{i,j} \right)$$

where $\boldsymbol{E}_j$ is like the identity matrix but with the first and last j ones set to zero, $\boldsymbol{F}_j$ has ones along the upper and lower j^{th} minor diagonals and zeroes elsewhere, and $\boldsymbol{G}_{i,j} = \boldsymbol{E}_i \boldsymbol{F}_j \boldsymbol{E}_i$ which equals $\boldsymbol{F}_j$ except the top and bottom i ones along the j^{th} minor diagonals are replaced by zeroes. We can also write this as

$$\frac{1}{\sigma^2}\Sigma^{-1} = \frac{1}{\sigma^2}(M^{-1} - W)$$

where

$$M^{-1} = I_T + \sum_{j=1}^{p} \phi_j^2 E_j$$

and

$$W = \sum_{j=1}^{p} \phi_j F_j - \sum_{j=1}^{p-1} \sum_{i=1}^{p-j} \phi_i \phi_{i+j} G_{i,j}$$

To calculate the matrix of derivatives of Σ w.r.t $\phi_1 \dots \phi_p$ it is easier to calculate the derivatives of Σ^{-1} and use the result $\frac{\partial \Sigma}{\partial \phi} = -\Sigma \frac{\partial \Sigma^{-1}}{\partial \phi} \Sigma$.

$$\begin{aligned} \frac{\partial \Sigma^{-1}}{\partial \phi_s} &= \frac{\partial M^{-1}}{\partial \phi_s} - \frac{\partial W}{\partial \phi_s} \\ &= 2\phi_s \boldsymbol{E}_s - \boldsymbol{F}_s + \sum_{j=1}^{p-s} \phi_{s+j} \boldsymbol{G}_{s,j} + \sum_{j=1}^{s-1} \phi_{s-j} \boldsymbol{G}_{s-j,j} \end{aligned}$$

The logarithm of the determinant (log det) of Σ can be obtained (using Haddad (1998)) from the roots $(r_1, r_2, \dots r_p)$ of the characteristic equation

$$z^p - \phi_1 z^{p-1} - \dots - \phi_{p-1} z - \phi_p = 0$$

and using

$$det(\Sigma) = \prod_{i,j=1}^{p} (1 - r_i r_j)^{-1}$$

SAR symmetric simultaneous autoregressive model

A stationary symmetric simultaneous autoregressive SAR(p) model can be written as

$$e_i = \sum_{j=1}^{p} \lambda_j (e_{i-j} + e_{i+j}) + \xi_i$$

Hence a SAR(1) model is given by

$$e_i = \lambda (e_{i-1} + e_{i+1}) + \xi_i$$

where $\xi_i \sim N(0, \sigma^2)$ and $|\lambda| < 0.5$.

This model has the same covariance structure as an AR(2) model where

$$e_i = \phi_1 e_{i-1} + \phi_2 e_{i-2} + \xi_i^*$$

with $\phi_1 = 2\phi$, $\phi_2 = -\phi^2 = -\frac{\phi_1^2}{4}$, $\lambda = \frac{\phi}{1+\phi^2}$ and $\xi_i^* \sim N(0, \nu^2)$.
The variance covariance matrix is given by $\nu^2 \Sigma = \nu^2 \{\Sigma_{ij}\}$ where

$$\Sigma_{ij} = \begin{cases} 1 & i = j \\ \frac{\phi_1}{1+\frac{\phi_1^2}{4}} & |i-j| = 1 \\ \phi_1 \Sigma_{i-1,j} + \frac{\phi_1^2}{4} \Sigma_{i-2,j} & i > j+1 \end{cases} \tag{B.10}$$

The inverse covariance matrix is given by

$$\frac{1}{\nu^2}\Sigma^{-1} = \frac{1}{\nu^2}\begin{bmatrix} 1 & & & & & \\ -\phi_1 & 1+\phi_1^2 & & & & \\ \frac{\phi_1^2}{4} & -\phi_1-\frac{\phi_1^3}{4} & 1+\phi_1^2+\frac{\phi_1^4}{16} & & & \\ 0 & \frac{\phi_1^2}{4} & -\phi_1-\frac{\phi_1^3}{4} & 1+\phi_1^2+\frac{\phi_1^4}{16} & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \\ & & \frac{\phi_1^2}{4} & -\phi_1-\frac{\phi_1^3}{4} & 1+\phi_1^2 & \\ 0 & \dots & 0 & \frac{\phi_1^2}{4} & -\phi_1 & 1 \end{bmatrix} \tag{B.11}$$

and hence

$$\Sigma^{-1} = \boldsymbol{I}_T + \phi_1^2\boldsymbol{E}_1 + \frac{\phi_1^4}{16}\boldsymbol{E_2} - \phi_1\boldsymbol{F}_1 - \frac{\phi_1^3}{4}\boldsymbol{G}_{1,1} + \frac{\phi_1^2}{4}\boldsymbol{F}_2$$

Again we can write

$$\Sigma^{-1} = (M^{-1} - W)$$

where

$$M^{-1} = \boldsymbol{I}_T + \phi_1^2\boldsymbol{E}_1 + \frac{\phi_1^4}{16}\boldsymbol{E}_2$$

$$W = \phi_1\boldsymbol{F}_1 + \frac{\phi_1^3}{4}\boldsymbol{G}_{1,1} - \frac{\phi_1^2}{4}\boldsymbol{F}_2$$

To calculate the matrix of derivatives of Σ we use the standard result $\frac{\partial\boldsymbol{\Sigma}}{\partial\phi} = -\boldsymbol{\Sigma}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\phi}\boldsymbol{\Sigma}$. and calculate the derivatives of the inverse covariance matrix Σ^{-1}. That is

$$\frac{\partial\Sigma^{-1}}{\partial\phi_1} = \frac{\partial M^{-1}}{\partial\phi_1} - \frac{\partial W}{\partial\phi_1}$$

where

$$\frac{\partial M^{-1}}{\partial\phi_1} = 2\phi_1\boldsymbol{E}_1 + \frac{4\phi_1^3}{16}\boldsymbol{E}_2$$

$$\frac{\partial W}{\partial\phi_1} = \boldsymbol{F}_1 + \frac{3\phi_1^2}{4}\boldsymbol{G}_{1,1} - \frac{\phi_1}{2}\boldsymbol{F}_2$$

The logarithm of the determinant (log det) of Σ is given by

$$log|\Sigma| = log[\frac{1}{(1-\frac{\phi_1^2}{4})^2[(1+\frac{\phi_1^2}{4})^2-\phi_1^2]}]$$

CAR Conditional Autoregressive Models

CAR models are defined by conditional distributions as given by

$$\begin{aligned} E[y_i|y_{-i}] &= \mu_i + \sum_{j=1}^{n} c_{ij}(y_j - \mu_j) \\ Var[y_i|y_{-i}] &= \sigma_i^2 \end{aligned}$$

for $i, j = 1 \dots n$ where y_{-i} indicates all y_j such that $j \neq i$.

It can be shown in the Gaussian case

$$\boldsymbol{y} \sim N(\boldsymbol{\mu}, (\boldsymbol{I}-\boldsymbol{C})^{-1}\boldsymbol{M}) \tag{B.12}$$

provided $(\boldsymbol{I}-\boldsymbol{C})$ is invertible and $(\boldsymbol{I}-\boldsymbol{C})^{-1}\boldsymbol{M}$ is symmetric and positive definite, where $\boldsymbol{M} = diag(\sigma_1^2, \dots, \sigma_n^2)$ and $\boldsymbol{C}$ is a $n \times n$ matrix specifying spatial dependencies between

locations with $c_{ii} = 0$ and $c_{ij} \neq 0$ if j a neighbour of i, $j \in N_i$. The requirement that $(\boldsymbol{I} - \boldsymbol{C})^{-1}\boldsymbol{M}$ is symmetric gives the condition $c_{ij}\sigma_j^2 = c_{ji}\sigma_i^2$.

Therefore the inverse covariance matrix is given by

$$\Sigma^{-1} = M^{-1}(I - C) = M^{-1} - M^{-1}C$$

A CAR model of order 1 has C given by

$$\boldsymbol{C_{ij}} = \begin{cases} 0 & for\ i = j \\ c_{ij} \quad where\ c_{ij}\sigma_j^2 = c_{ji}\sigma_i^2 & |i-j| = 1 \\ 0 & |i-j| > 1 \end{cases} \tag{B.13}$$

The logarithm of the determinant of Σ^{-1} is given by

$$log|\Sigma^{-1}| = log(\prod_{i=1}^{n} \frac{1}{\sigma_i^2}) + log|I - C)|$$

Exponential model

An exponential covariance model (based on distance between observations ($d_{ij} = |x_i - x_j|$) is given by $\sigma^2\Sigma = \sigma^2\{\Sigma_{ij}\}$ where

$$\Sigma_{ij} = \begin{cases} 1 & for\ i = j \\ b_1 exp(-\frac{|x_i - x_j|}{b_2}) \quad = \phi^{|x_i - x_j|} & i \neq j \end{cases} \tag{B.14}$$

where $|\phi| < 1$ and x_i are co-ordinates.
The matrix of derivatives is given by

$$\frac{\partial\Sigma}{\partial\phi} = \begin{cases} 0 & i = j \\ |x_i - x_j|\phi^{|x_i - x_j|-1} & i \neq j \end{cases}$$

AR1 $\times$ AR1

The separable AR(1) by AR(1) model in a grid of c columns by r rows has covariance matrix given by

$$\Sigma = \sigma^2(\Sigma_c \otimes \Sigma_r)$$

where Σ_r is the covariance matrix in the row dimension and Σ_c is the covariance matrix in the column direction.
Explicitly

$$\Sigma_r = \begin{bmatrix} 1 & \phi_r & \phi_r^2 & \dots & \phi_r^{r-1} \\ \phi_r & 1 & \phi_r & \dots & \phi_r^{r-2} \\ \phi_r^2 & \phi_r & 1 & \dots & \phi_r^{r-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \phi_r^{r-1} & \dots & \phi_r^2 & \phi_r & 1 \end{bmatrix}$$

and

$$\Sigma_c = \begin{bmatrix} 1 & \phi_c & \phi_c^2 & \dots & \phi_c^{c-1} \\ \phi_c & 1 & \phi_c & \dots & \phi_c^{c-2} \\ \phi_c^2 & \phi_c & 1 & \dots & \phi_c^{c-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \phi_c^{c-1} & \dots & \phi_c^2 & \phi_c & 1 \end{bmatrix}$$

and hence for the case of 9 plots in 3 rows and 3 columns

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \phi_r & \phi_r^2 & \phi_c & \phi_c\phi_r & \phi_c\phi_r^2 & \phi_c^2 & \phi_c^2\phi_r & \phi_c^2\phi_r^2 \\ \phi_r & 1 & \phi_r & \phi_c\phi_r & \phi_c & \phi_c\phi_r & \phi_c^2\phi_r & \phi_c^2 & \phi_c^2\phi_r \\ \phi_r^2 & \phi_r & 1 & \phi_c\phi_r^2 & \phi_c\phi_r & \phi_c & \phi_c^2\phi_r^2 & \phi_c^2\phi_r & \phi_c^2 \\ \phi_c & \phi_c\phi_r & \phi_c\phi_r^2 & 1 & \phi_r & \phi_r^2 & \phi_c & \phi_c\phi_r & \phi_c\phi_r^2 \\ \phi_c\phi_r & \phi_c & \phi_c\phi_r & \phi_r & 1 & \phi_r & \phi_c\phi_r & \phi_c & \phi_c\phi_r \\ \phi_c\phi_r^2 & \phi_c\phi_r & \phi_c & \phi_r^2 & \phi_r & 1 & \phi_c\phi_r^2 & \phi_c\phi_r & \phi_c \\ \phi_c^2 & \phi_c^2\phi_r & \phi_c^2\phi_r^2 & \phi_c & \phi_c\phi_r & \phi_c\phi_r^2 & 1 & \phi_r & \phi_r^2 \\ \phi_c^2\phi_r & \phi_c^2 & \phi_c^2\phi_r & \phi_c\phi_r & \phi_c & \phi_c\phi_r & \phi_r & 1 & \phi_r \\ \phi_c^2\phi_r^2 & \phi_c^2\phi_r & \phi_c^2 & \phi_c\phi_r^2 & \phi_c\phi_r & \phi_c & \phi_r^2 & \phi_r & 1 \end{bmatrix}$$

The inverse covariance matrix is given by

$$\Sigma^{-1} = \frac{1}{\sigma^2}(\Sigma_c^{-1} \otimes \Sigma_r^{-1})$$

where

$$\Sigma_c^{-1} = \begin{bmatrix} 1 & -\phi_c & 0 & \dots & 0 \\ -\phi_c & 1+\phi_c^2 & -\phi_c & \dots & 0 \\ 0 & -\phi_c & 1+\phi_c^2 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -\phi_c & 1 \end{bmatrix}$$

Similarly for Σ_r^{-1}.

Hence

$$\begin{aligned} \Sigma^{-1} &= \frac{1}{\sigma^2}[(I_{Tc} + \phi_c^2 E_1 - \phi_c F_1) \otimes (I_{Tr} + \phi_r^2 E_1 - \phi_r F_1)] \\ &= I_{Tc} \otimes I_{Tr} + \phi_r^2 I_{Tc} \otimes E_{r1} - \phi_r I_{Tc} \otimes F_{r1} + \phi_c^2 E_{c1} \otimes I_{Tr} + \phi_r^2\phi_c^2 E_{c1} \otimes E_{r1} - \phi_c^2\phi_r E_{c1} \otimes F_{r1} \\ &\quad -\phi_c F_{c1} \otimes I_{Tr} - \phi_c\phi_r^2 F_{c1} \otimes E_{r1} + \phi_r\phi_c F_{c1} F_{r1} \end{aligned}$$

That is

$$\Sigma^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} \boldsymbol{\Sigma}_r^{-1} & \boldsymbol{A}_2 & 0 & \dots & 0 \\ \boldsymbol{A}_2 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \dots & 0 \\ 0 & \boldsymbol{A}_2 & \boldsymbol{A}_1 & & 0 \\ \vdots & \ddots & \ddots & \ddots & \boldsymbol{A}_2 \\ 0 & \dots & 0 & \boldsymbol{A}_2 & \boldsymbol{\Sigma}_r^{-1} \end{bmatrix}$$

where

$$\boldsymbol{A}_1 = (1+\phi_c^2)\Sigma_r^{-1}$$
$$\boldsymbol{A}_2 = -\phi_c\Sigma_r^{-1}$$

To calculate the matrix of derivatives of Σ w.r.t ϕ_r and ϕ_c, we use the result $\frac{\partial\Sigma}{\partial\phi} = -\Sigma\frac{\partial\Sigma^{-1}}{\partial\phi}\Sigma$ and calculate the derivatives of Σ^{-1}.

$$\begin{aligned}\frac{\partial\Sigma^{-1}}{\partial\phi_r} &= 2\phi_r I_{Tc}\otimes E_{r1} - I_{Tc}\otimes F_{r1} + 2\phi_r\phi_c^2 E_{c1}\otimes E_{r1} && \text{(B.15)}\\ &\quad -\phi_c^2 E_{c1}\otimes F_{r1} - 2\phi_c\phi_r F_{c1}\otimes E_{r1} + \phi_c F_{c1}\otimes F_{r1} && \text{(B.16)}\end{aligned}$$

$$\begin{aligned}\frac{\partial\Sigma^{-1}}{\partial\phi_c} &= 2\phi_c E_{c1}\otimes I_{Tr} + 2\phi_r^2\phi_c E_{c1}\otimes E_{r1} && \text{(B.17)}\\ &\quad -2\phi_c\phi_r E_{c1}\otimes F_{r1} - F_{c1}\otimes I_{Tr} - \phi_r^2 F_{c1}\otimes E_{r1} + \phi_r F_{c1}\otimes F_{r1} && \text{(B.18)}\end{aligned}$$

The determinant of Σ is given by

$$\begin{aligned}det(\Sigma) &= det(\sigma^2(\Sigma_c\otimes\Sigma_r))\\ &= (\sigma^2)^{rc}(det(\Sigma_c^{-1})^r(det(\Sigma_r^{-1}))^c\\ &= (\sigma^2)^{rc}(\frac{1}{1-\phi_c^2})^r(\frac{1}{1-\phi_r^2})^c\end{aligned}$$

CAR1 $\otimes$ CAR1

We now consider a separable $CAR1\otimes CAR1$ model on a 2 dimensional array of r rows and c columns. If $\Sigma_r^{-1} = M_r^{-1}(I_r - C_r)$ gives the inverse covariance matrix of a CAR model (of order1) in the row direction and $\Sigma_c^{-1} = M_c^{-1}(I_c - C_c)$ is the inverse covariance matrix in the column direction, then the separable $CAR1\otimes CAR1$ model has inverse covariance matrix given by

$$\begin{aligned}\Sigma^{-1} &= \Sigma_r^{-1}\otimes\Sigma_c^{-1}\\ &= M_r^{-1}(I_r - C_r)\otimes M_c^{-1}(I_c - C_c)\\ &= (M_r^{-1}\otimes M_c^{-1})((I_r - C_r)\otimes(I_c - C_c))\\ &= (M_r^{-1}\otimes M_c^{-1})(I_r\otimes(I_c - C_c) - C_r\otimes(I_c - C_c))\\ &= (M_r^{-1}\otimes M_c^{-1})(I_r\otimes I_c - I_r\otimes C_c - C_r\otimes I_c + C_r\otimes C_c)\end{aligned}$$

This can be written as

$$\Sigma^{-1} = M^{-1}(I_{rc} - C)$$

with $M^{-1} = M_r^{-1}\otimes M_c^{-1}$ and $C = I_r\otimes C_c + C_r\otimes I_c - C_r\otimes C_c$.
The spatial dependency matrix $\boldsymbol{C}_c$ has $(i,j)^{th}$ element given by

$$\boldsymbol{C}_c = \begin{cases} 0 & for\ i = j\\ c_{ij}^c \quad where\ c_{ij}^c\sigma_j^2 = c_{ji}^c\sigma_i^2 & |i-j| = 1\\ 0 & |i-j| > 1\end{cases}$$

for $i,j = 1\ldots c$. Similarly for $\boldsymbol{C}_r$.

We can write the model in terms of conditional distributions as

$$\begin{aligned}E[y_{ij}|y_{-ij}] = \mu_{ij} + c^r_{i-1,j}(y_{i-1,j} - \mu_{i-1,j}) + \\ c^r_{i+1,j}(y_{i+1,j} - \mu_{i+1,j}) + \\ c^c_{i-1,j}(y_{i,j-1} - \mu_{i,j-1}) + \\ c^c_{i+1,j}(y_{i,j+1} - \mu_{i,j+1}) + \\ -c^r_{i-1,j}c^c_{i-1,j}(y_{i-1,j-1} - \mu_{i-1,j-1}) \\ -c^r_{i-1,j}c^c_{i+1,j}(y_{i-1,j+1} - \mu_{i-1,j+1}) \\ -c^r_{i+1,j}c^c_{i-1,j}(y_{i+1,j-1} - \mu_{i+1,j-1}) \\ -c^r_{i+1,j}c^c_{i+1,j}(y_{i+1,j+1} - \mu_{i+1,j+1})\end{aligned}$$

The $rc \times rc$ matrix $\boldsymbol{C}$ can be written as

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_c & \boldsymbol{C}_{12} & 0 & \dots & 0 \\ \boldsymbol{C}_{21} & \boldsymbol{C}_c & \boldsymbol{C}_{23} & \dots & 0 \\ 0 & \boldsymbol{C}_{32} & \boldsymbol{C}_c & & 0 \\ \vdots & \ddots & \ddots & \ddots & \boldsymbol{C}_{r-1,r} \\ 0 & \dots & 0 & \boldsymbol{C}_{r,r-1} & \boldsymbol{C_c} \end{bmatrix}$$

where $\boldsymbol{C}_c$ is the $c \times c$ matrix defined above and

$$\boldsymbol{C}_{ij} = c^r_{ij}(\boldsymbol{I} + \boldsymbol{C}_c)$$

where $|i - j| = 1$.

proper 2d CAR

If the data Y are observed on a regular 2 dimensional lattice of n plots in r rows and c columns ($n = r \times c$) then we can define a 2d first order CAR model similar to that of Cressie (1993), using just three spatial dependency parameters (in C), with two (γ_r and γ_c) for adjacent plots in the row and column direction respectively and γ_{rc} for neighbouring diagonal plots. Therefore if we have plots

$$\begin{matrix} y_{i-1,j-1} & y_{i-1,j} & y_{i-1,j+1} \\ y_{i,j-1} & y_{i,j} & y_{i,j+1} \\ y_{i+1,j-1} & y_{i+1,j} & y_{i+1,j+1} \end{matrix}$$

Then we define the 2d CAR model of order 1 as

$$\begin{aligned}E[y_{ij}|y_{-ij}] = \mu_{ij} + \gamma_r(y_{i-1,j} - \mu_{i-1,j}) + \\ \gamma_r(y_{i+1,j} - \mu_{i+1,j}) + \\ \gamma_c(y_{i,j-1} - \mu_{i,j-1}) + \\ \gamma_c(y_{i,j+1} - \mu_{i,j+1}) + \\ \gamma_{rc}[(y_{i-1,j-1} - \mu_{i-1,j-1}) \\ +(y_{i-1,j+1} - \mu_{i-1,j+1}) \\ +(y_{i+1,j-1} - \mu_{i+1,j-1}) \\ +(y_{i+1,j+1} - \mu_{i+1,j+1})]\end{aligned} \tag{B.19}$$

We assume a common conditional variance for all plots. That is

$$Var(y_{ij}|y_{-ij}) = \sigma^2$$

Now the variance matrix of y is given by

$$Var(y) = (I - C)^{-1}M = \Sigma$$

and hence

$$\Sigma^{-1} = M^{-1}(I - C)$$

where M is a diagonal matrix containing the conditional variances $M = diag(\sigma^2)$ and C is the spatial dependency matrix which can be written as

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 & 0 & \dots & 0 \\ \boldsymbol{C}_2 & \boldsymbol{C}_1 & \boldsymbol{C}_2 & \dots & 0 \\ 0 & \boldsymbol{C}_2 & \boldsymbol{C}_1 & & 0 \\ \vdots & \ddots & \ddots & \ddots & \boldsymbol{C}_2 \\ 0 & \dots & 0 & \boldsymbol{C}_2 & \boldsymbol{C}_1 \end{bmatrix}$$

where

$$\boldsymbol{C}_1 = \gamma_c \boldsymbol{F}_{1c}$$

and

$$\boldsymbol{C}_2 = \gamma_{rc} \boldsymbol{F}_{1r} + \gamma_r \boldsymbol{I}_r.$$

Therefore we can write

$$\boldsymbol{I} - \boldsymbol{C} = (\boldsymbol{I}_c \otimes (\boldsymbol{I}_r - \gamma_r \boldsymbol{F}_{1r}) + \boldsymbol{F}_{1c} \otimes (-\gamma_c \boldsymbol{I}_r - \gamma_{rc} \boldsymbol{F}_{1r}))$$

and hence

$$\begin{aligned} \Sigma^{-1} &= \frac{1}{\sigma^2}[I_c \otimes (I_r - \gamma_r F_{1r}) + F_{1c} \otimes (-\gamma_c I_r - \gamma_{rc} F_{1r})] \\ &= \frac{1}{\sigma^2}[I_c \otimes I_r - \gamma_r (I_c \otimes F_{1r}) - \gamma_c (F_{1c} \otimes I_r) - \gamma_{rc} (F_{1c} \otimes F_{1r})] \end{aligned} \tag{B.20}$$

If we take $\gamma_{rc} = -\gamma_c \gamma_r$ we can get a separable form for Σ^{-1} as follows

$$\begin{aligned} \Sigma^{-1} &= \frac{1}{\sigma^2}[I_c \otimes I_r - \gamma_r (I_c \otimes F_{1r}) - \gamma_c (F_{1c} \otimes I_r) + \gamma_c \gamma_r (F_{1c} \otimes F_{1r})] \\ &= \frac{1}{\sigma^2}[I_c \otimes (I_r - \gamma_r F_{1r}) - \gamma_c (F_{1c} \otimes (I_r - \gamma_r F_{1r}))] \\ &= \frac{1}{\sigma^2}[(I_c - \gamma_c F_{1c}) \otimes (I_r - \gamma_r F_{1r})] \end{aligned} \tag{B.21}$$

If the matrix of derivatives of Σ w.r.t γ_r, γ_c and γ_{rc} is required then we use the standard result $\frac{\partial \Sigma}{\partial \phi} = -\Sigma \frac{\partial \Sigma^{-1}}{\partial \phi} \Sigma$ and calculate the derivatives of Σ^{-1}. Hence, using B.20,

$$\frac{\partial \Sigma^{-1}}{\partial \gamma_r} = -\frac{1}{\sigma^2}(I_c \otimes F_{1r})$$

$$\frac{\partial \Sigma^{-1}}{\partial \gamma_c} = -\frac{1}{\sigma^2}(F_{1c} \otimes I_c)$$

$$\frac{\partial \Sigma^{-1}}{\partial \gamma_{rc}} = -\frac{1}{\sigma^2}(F_{1c} \otimes F_{1r})$$

US Unstructured general covariance

An unstructured general covariance matrix is given by $\sigma^2\Sigma = \sigma^2\{\Sigma_{ij}\}$ where

$$\Sigma_{ij} = \phi_{ij}$$

where $|\phi| < 1$.

COR Uniform correlation

A uniform correlation model is given by $\sigma^2\Sigma = \sigma^2\{\Sigma_{ij}\}$ where

$$\Sigma_{ij} = \begin{cases} 1 & for\ i = j \\ \theta & i \neq j \end{cases} \tag{B.22}$$

where $\theta < 1$.
The matrix of derivatives is given by

$$\frac{\partial\Sigma}{\partial\theta} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$$

ANTE Antedependence order k

An antedependence model of order s is defined by the fact that the j^{th} observation ($j > s$) given the s proceeding observations is independent of all other proceeding observations.

An antedependence variance model is similar to an autoregressive model but it does not restrict the variances to be constant and it allows correlations between measurements at same lags to vary. We can write a s^{th} order antedependence model for $Y = (y_1, \ldots y_T)$ where $Y \sim N(\mu, \Sigma)$ as

$$Y_j = \mu_j + \sum_{k=1}^{s_j} \gamma_{kj}(Y_{j-k} - \mu_{j-k}) + \epsilon_j \tag{B.23}$$

where $\epsilon_j \sim N(0, d_j)$ and $s_j = min(j-1, s)$.
The inverse covariance matrix of Y can be written as

$$\Sigma^{-1} = U^T D U$$

where D is defined as a $T \times T$ diagonal matrix with d_j (conditional variances) as diagonal elements and U is a $t \times t$ lower triangular matrix with elements

$$u_{ij} = \begin{cases} 1 & i = j \\ -\gamma_{i,j} & 1 \leq i - j < s \\ 0 & elsewhere \end{cases}$$

The determinant of Σ^{-1} is then equal to the product of the diagonal elements of D.

That is

$$det(\Sigma^{-1}) = det(U^T DU) = \prod_{i=1}^{T} d_j$$

.

FA Factor analytic

A factor analytic model (Smith et al. (2001) is a multiplicative model (based on the multivariate technique of factor analysis) that gives a good approximation to the fully unstructured model but with many less parameters. The covariance model is given by

$$\Sigma = \Gamma\Gamma^T + \Psi \tag{B.24}$$

where Γ is a matrix of environmental loadings and Ψ is a diagonal matrix containing the specific variances, one for each trial.

The inverse covariance model is given by

$$\Sigma^{-1} = \Psi^{-1} - \Psi^{-1}\Gamma(\Gamma^T\Psi^{-1}\Gamma + I)^{-1}\Gamma^T\Psi^{-1}$$

DIAG diagonal

A diagonal covariance matrix is given by

$$\Sigma_{ij} = \begin{cases} \phi_i & for\ i = j \\ 0 & i \neq j \end{cases} \tag{B.25}$$

C R code for fitting random regression model in Chapter 6

```
# Code for fitting overall cubic smoothing spline plus linear random regression model
# to the lucerne persistence data and subsequently estimating the time for each
# variety to decline to 30% persistence
#
# fit model with overall spline to mean response and linear random regressions for
# variety deviations from this overall trend
# dev(years) - fits years as a factor which is given by ExptTime
# random intercepts and slopes need to be correlated
 terrycnew.m51.sv <- asreml(logisP~1+ years ,
                            random=~spl(years)+ExptTime+ str(~ID+ years:ID,
                            ~corh(2):id(60)) +
                            ExptTime:ID+ at(ExptTime):Rep + diag(ExptTime):Range
                            +Range:ar1v(Row) + ExptTime:Range:Row,
                            rcov=~ ar1h(ExptTime):Range:ar1(Row), start.values=T,
                            G.param=terrycnew.mod5o.asr$G.param,
                            R.param=terrycnew.mod5o.asr$R.param,
                            data=terryc,maxit=30)
gammasm51.tab <- terrycnew.m51.sv$gammas.table
gammasm51.tab[c(19,20,21),'Value'] <- c(-0.2,0.03,0.01)
terrycnew.m51.asr <- asreml(logisP~1+ years ,
                            random=~spl(years)+ExptTime+ str(~ID+ years:ID,
```

```
                            ~corh(2):id(60)) +
                            ExptTime:ID+ at(ExptTime):Rep + diag(ExptTime):Range
                            +Range:ar1v(Row) + ExptTime:Range:Row,
                            rcov=~ ar1h(ExptTime):Range:ar1(Row),
                            G.param=gammasm51.tab, R.param=gammasm51.tab,
                            data=terryc,maxit=30)
#
# now predict time to 30% based on terrycnew.m51.asr
########################################
# try and predict time to 30% - based on mod51 (with correlation between
# intercept & slope terms)
# calculate time to p% persistence
# calc gamma = R^{-1}Q^Tg
# calc h, Q, G_s all as above
# calc overall gamma = R^{-1}Q^Tg   - a p-2 x 1 matrix
# # overall spline g
terrycnew.m51.pred<- predict(terrycnew.m51.asr, classify="years",
levels=list(years=unique(terryc$years)),
         only=c("years","spl(years)","(Intercept)"))
t5.pred<-terrycnew.m51.pred$pred$pvals
g5<-t5.pred$predicted.value
# calculate time to p% persistence
# calc gamma = R^{-1}Q^Tg
hdat.df<-data.frame(yr=unique(terryc$years))
# calc h
h<-c()
ty<-length(unique(terryc$years))
for (i in 1:(ty-1))
{h <- c(h, hdat.df$yr[i+1] - hdat.df$yr[i])}
# calc Q matrix - from Green Silverman
q<-c(h[1]^{-1}, 0,0,0,-(h[1]^{-1}+h[2]^{-1}), h[2]^{-1}, 0,0, h[2]^{-1},
-(h[2]^{-1}+h[3]^{-1}),
 h[3]^{-1},0,0,h[3]^{-1},-(h[3]^{-1}+h[4]^{-1}),h[4]^{-1},
0,0,h[4]^{-1}, -(h[4]^{-1}+h[5]^{-1}), 0,0,0,h[5]^{-1})
qmat<-matrix(nrow=6,ncol=4,byrow=TRUE,data=q)
# calc G_s matrix (also called R in book)
gs<-c((h[1]+h[2])/3, h[2]/6, 0,0,h[2]/6, (h[2]+h[3])/3, h[3]/6,0,0,h[3]/6,
(h[3]+h[4])/3,
h[4]/6,0,0,h[4]/6,(h[4]+h[5])/3)
gsmat<-matrix (nrow = 4, ncol = 4 , byrow = TRUE , data = gs )
# calc gamma = R^{-1}Q^Tg   - a p-2 x 1 matrix
library(MASS)
gamm5<-ginv(gsmat)%*%t(qmat)%*%g
# set first and last elements to zero
gamm5<- rbind(0,gamm5,0)

# get intercepts & slopes from linear random regression
# and add line values to spline
t51.sum<- summary(terrycnew.m51.asr,all=T)
t5rr<-t51.sum$coef.random[grep("ID", dimnames(t51.sum$coef.random)[[1]]),]
t5rrint<-t5rr[1:60,1]
t5rrsl<-t5rr[61:120,1]
yval<-unique(terryc$years)
line.df<-data.frame(yr=rep(unique(terryc$years),each=60),
ID=rep(levels(terryc$ID),6),
```

```
int=rep(t5rrint,6), sl=rep(t5rrsl,6))
line.df$lin<-line.df$int+line.df$sl*line.df$yr
line.df$g5<-rep(g5,each=60)
line.df$val<-line.df$lin+line.df$g5

# get values for overall spline and individual RR together
# gi in order years then ID so need to change everything else
line.df$yro<-rep(unique(terryc$yearso),each=60)
xyplot(val~yro|ID,type="b", xlab='Years from sowing',
ylab='Predicted persistence (logit) from final model',
par.strip.text=list(cex=0.6),as.table=T,data=line.df)

# calc target value on transformed scale - 30%?
p<-30
target<-log((p+0.5)/(100-p+0.5))
# sort line.df so on ID then yr
line.dfo<-line.df[order(line.df$ID,line.df$yr),]
# find interval where each variety crosses target
#(decreases to target - val_l > target and val_r < target
s<-0
time<-list()
for (i in 1:60)
{ for (h in 1:6)
  {if ((line.dfo$val[h+(i-1)*6] > target) &  (line.dfo$val[h+1+(i-1)*6] < target))
  {time[[i]]<-h}
}
 }
 t5_l<-t5_r<-t_m<-g5_l<-g5_r<-gam5_l<-gam5_r<-result5<-yrres5<-list()

 for (i in 1:60)
 {t5_l[[i]] <- yval[time[[i]]]
  t5_r[[i]] <- yval[time[[i]]+1]
  # get g(t) and gam at t_l and t_r
  g5_l[[i]] <- g5[time[[i]]]
  g5_r[[i]] <- g5[time[[i]]+1]
  gam5_l[[i]] <- gamm5[time[[i]]]
  gam5_r[[i]] <- gamm5[time[[i]]+1]
#z(t)=g(t)+h(t)
# g(t) use eqn 2.19 in Green Silverman
# h(t) = int + slope*t for each i
#h_t[[i]] <- t5rrint[i] + t5rrsl[i]*t
result5[[i]] <- uniroot(function(t) (((t-t5_l[[i]])*g5_r[[i]]) +
((t5_r[[i]]-t)*g5_l[[i]]))/(t5_r[[i]]-t5_l[[i]])
-(1/6)*(t-t5_l[[i]])*(t5_r[[i]]-t)*((1+(t-t5_l[[i]])/(t5_r[[i]]-t5_l[[i]]))*gam5_r[[i]]
 + (1+(t5_r[[i]]-t)/(t5_r[[i]]-t5_l[[i]]))*gam5_l[[i]]) +t5rrint[i] + t5rrsl[i]*t
 - target,
lower=t5_l[[i]], upper=t5_r[[i]] )
# get year value for each variety (add on mean year)
yrres5[[i]]<- mean(terryc$yearso)+result5[[i]]$root
  }
results5.df<-data.frame(ID=levels(terryc$ID), yrres5=unlist(yrres5))
resord5.df<-results5.df[order(results5.df$yrres,decreasing=F),]
library(xtable)
xtable(resord5.df)
resord5.df$count<-c(1:60)
```

```
#plot results
stripplot(resord5.df$count ~ resord5.df$yrres5, data=resord5.df,
scales=list(y=list(labels=resord5.df$ID, cex=0.6)), ylab="Variety",
xlab="Time (years) till 30% persistence")


##############################################################################################
```

D R Code for implementing MVAR models in Chapter 8

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
##############################################################################################
# getsigimodmv
# for multivariate data in Row dir (MCAR x ID)
# modified by JDF Mar2010
##############################################################################################
getsigimodmv <- function(rows,times, rowmod, dirname = "Row")
{
    partype <- ifelse(dirname == "Row", 1, 2)
    psi<-rowmod$psi

    if(rowmod$name == 'USI') {
        asig <- USI(rows, times,phi = rowmod$phi)
        sig <- asig$sig
        sigi <- asig$sigi
        dsig <- asig$dsig
        phi <- rowmod$phi
        partype <- rep(partype, length(phi))
        }
         else if(rowmod$name == 'USgen') {
        asig <- USgen(rows, times,phi = rowmod$phi, psi=rowmod$psi)
        sig <- asig$sig
        sigi <- asig$sigi
        dsig <- asig$dsig
        phi <- rowmod$phi
        psi <- rowmod$psi
        partype <- rep(partype, length(phi))
        }
         else if(rowmod$name == 'USsym') {
        asig <- USsym(rows, times,phi = rowmod$phi, psi=rowmod$psi)
        sig <- asig$sig
        sigi <- asig$sigi
        dsig <- asig$dsig
        phi <- rowmod$phi
        psi <- rowmod$psi
        partype <- rep(partype, length(phi))
        }
          else if(rowmod$name == 'ar1hgen') {
        asig <- ar1hgen(rows, times,phi = rowmod$phi, psi=rowmod$psi)
        sig <- asig$sig
        sigi <- asig$sigi
        dsig <- asig$dsig
```

```
        phi <- rowmod$phi
        psi <- rowmod$psi
        partype <- rep(partype, length(phi))
        }
        else if(rowmod$name == 'ar1hsym') {
        asig <- ar1hsym(rows, times,phi = rowmod$phi, psi=rowmod$psi)
        sig <- asig$sig
        sigi <- asig$sigi
        dsig <- asig$dsig
        phi <- rowmod$phi
        psi <- rowmod$psi
        partype <- rep(partype, length(phi))
           }
        else if(rowmod$name == 'anteband') {
        asig <- anteband(rows, times,phi = rowmod$phi, psi=rowmod$psi)
        sig <- asig$sig
        sigi <- asig$sigi
        dsig <- asig$dsig
        phi <- rowmod$phi
        psi <- rowmod$psi
        partype <- rep(partype, length(phi))
       }
    list(sig=sig,sigi = sigi, dsig = dsig, phi = phi, psi=psi, partype = partype)
}
# end getsigimodmv
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

##############################################################################
#
# USgen
##############################################################################
USgen<-function(r,t,phi=c(rep(0.1,(t*(t+1)/2+t^2))), psi=NULL)
{
# forms the inverse sig and derivatives of sigma for the MVAR1 model
# with general Lambda (t times)and Sigma = US (symmetric) with constraints)
# r=rows t=times phi = sigma_11 sigma_12, sigma_13, ...sigma_1t, sigma_22,
#sigma_23, ... lambda_11, lambda_12,...lambda_1t,lambda_21, lambda_22,... lambda_2t
# Mar2010
#JDF
##############################################################################
F1<-mydiag(1,1,r-1,r-1)+mydiag(1,-1,r-1,r-1)
E1<- diag(c(0,rep(1,r-2),0))
Ir<-diag(r)
It<-diag(t)
Irt<-diag(r*t)
Jt<-matrix(1,nrow=t,ncol=t)
#sigt<-matrix(ncol=2,nrow=2,data=c(phi[1],phi[2],phi[2],phi[3]))
#lamt<-matrix(ncol=2,nrow=2,data=c(phi[4], phi[5],phi[6],phi[7]))
# sigt = Sigma for times  dS - derivatives of Sigma wrt phi[k]
#phi<-c(1:(t*(t+1)/2+t^2))
sigphi<-t*(t+1)/2
totphi<-t*(t+1)/2 + t^2
lamphi<- totphi-sigphi
Sigma<-matrix(0,t,t)
ei<-ej<-matrix(0,t,1)
```

```
dS<-dSo<-list()
k<-0
for (i in 1:t)
{for (j in i:t)
{k<-k+1
Sigma[i,j]<- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dSo[[k]]<-ei%*%t(ej)
dS[[k]]<- dSo[[k]] + t(dSo[[k]]) - diag(diag(dSo[[k]]))
}
}
sigt<-Sigma+t(Sigma)-diag(diag(Sigma))
# Lambda
lamt<-matrix(0,t,t)
dL<-list()
k<-(t*(t+1)/2)
for (i in 1:t)
{ for (j in 1:t)
{k <- k+1
lamt[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dL[[k]]<-ei%*%t(ej)
}
}


sti<-ginv.new(sigt)$invx
s0<-sti%*%ginv.new(It-(lamt%*%lamt))$invx
s1<-kronecker(Ir,s0)
s2<-kronecker(F1,lamt)
s3<-kronecker(E1,(lamt%*%lamt))
sigi<-s1%*%(Irt-s2+s3)
sig<-ginv.new(sigi)$invx
P1<-s1
P2<-Irt-s2+s3
# diff wrt phi[k]
dP1<-list()
dP2<-list()
for (k in 1:sigphi)
{dP1[[k]]<- kronecker(Ir, -sti%*%(dS[[k]])%*%s0)
dP2[[k]]<-0*Irt
}
for (k in (1+sigphi):totphi)
{dP1[[k]]<- kronecker(Ir, s0%*%(lamt%*%dL[[k]]+
 dL[[k]]%*%lamt)%*%ginv.new(It-(lamt%*%lamt))$invx )
dP2[[k]]<-kronecker(-F1,dL[[k]]) + kronecker(E1,(lamt%*%dL[[k]]+dL[[k]]%*%lamt))
}
dsigi<-list()
dsig<-list()
for (k in 1:totphi)
{dsigi[[k]]<-dP1[[k]]%*%P2 + P1%*%dP2[[k]]
```

```
dsig[[k]]<- -sig%*%dsigi[[k]]%*%sig
}
list(sig=sig,sigi=sigi,dsig=dsig,phi=phi,partype=rep(1,length(phi)))
}
#
#end USgen
################################################################
#
# derivatives of constraints using vech  - tdconsgen
################################################
dconsgen<-function(t,tp,qran,rowmod)
{
phi<-rowmod$phi
psi<-rowmod$psi
sigphi<-t*(t+1)/2
totphi<-t*(t+1)/2 + t^2
lamphi<- totphi-sigphi
# function to get constraints and differentials of constraints
Sigma<-matrix(0,t,t)
dS<-dSo<-list()
k<-0
for (i in 1:t)
{for (j in i:t)
  {k<-k+1
  Sigma[i,j]<- phi[k]
  ei<-ej<-matrix(0,t,1)
  ei[i,1]<-1
  ej[j,1]<-1
  dSo[[k]]<-ei%*%t(ej)
  dS[[k]]<- dSo[[k]] + t(dSo[[k]]) - diag(diag(dSo[[k]]))
   }
}
sigt<-Sigma+t(Sigma)-diag(diag(Sigma))

lamt<-matrix(0,t,t)
dL<-list()
k<-(t*(t+1)/2)
for (i in 1:t)
{ for (j in 1:t)
{k <- k+1
lamt[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
  ei[i,1]<-1
  ej[j,1]<-1
  dL[[k]]<-ei%*%t(ej)
}
}

# now for constraints
dcons<-matrix(0,t,t)
psimat<-matrix(0,t,t)
cons<-matrix(0,t,t)
dcon<-list()
dconpsi<-list()
c<-list()
```

```
l<-0
# allocate psi[k] across top half of psimat (then make as a symmetric matrix)
for (i in 1:(t-1))
{ for (j in (i+1):t)
  { l<-l+1
    psimat[i,j]<-psi[l]
#    cons[i,j]<-sum(lamt[i,1:t]*sigt[1:t,j]) - sum(sigt[i,1:t]*lamt[j,1:t])
#    c[[l]]<- cons[i,j]
  }
}
# now create full matrices
psimatf<-psimat+t(psimat)-diag(diag(psimat))
psimat<-psimatf
#cons<-consf
# now differentiate the constraints wrt phi[k] - from Sigma  - sigma_rs
pvec<-mvec<-list()
M<-list()
 for (k in 1:sigphi)
 {
   pvec[[k]]<-matrix(0,t*(t+1)/2,1)
   mvec[[k]]<-matrix(0,t*(t+1)/2,1)
   M[[k]] <- lamt%*%dS[[k]] - dS[[k]]%*%t(lamt)
   pvec[[k]]<-as.vector(psimat[lower.tri(psimat)])
  mvec[[k]]<-as.vector(M[[k]][lower.tri(M[[k]])])
  dcon[[k]] <-(t(pvec[[k]])%*%mvec[[k]])
  }

# now differentiate C wrt phi[k] - from Lambda - lambda_ij
for (k in (sigphi+1):totphi)
{
  pvec[[k]]<-matrix(0,t*(t+1)/2,1)
  mvec[[k]]<-matrix(0,t*(t+1)/2,1)
   M[[k]] <- dL[[k]]%*%sigt - sigt%*%t(dL[[k]])
  pvec[[k]]<-as.vector(psimat[lower.tri(psimat)])
  mvec[[k]]<-as.vector(M[[k]][lower.tri(M[[k]])])
  dcon[[k]] <-(t(pvec[[k]])%*%mvec[[k]])
  }

# do using vech
P<-list()
l<-0
for (s in 1:(t-1))
 {for (r in (s+1):t)
  {l<-l+1
  pvec[[l]]<-mvec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  P[[l]]<-er%*%t(es)
  M[[l]]<-lamt%*%sigt - sigt%*%t(lamt)
  pvec[[l]]<-as.vector(P[[l]][lower.tri(P[[l]])])
  mvec[[l]]<-as.vector(M[[l]][lower.tri(M[[l]])])
  dconpsi[[l]]<-t(pvec[[l]])%*%mvec[[l]]
  # match up constraints with psis in correct order
```

```
  c[[l]]<-dconpsi[[l]]
  }
}



# now need to adjust AI matrix with second derivatives where relevant
# use matrix O to add extra bits to AI matrix (AI in order sigma2, gammas,
# phi[k] , phi_col, psi[l])
# tp= total number of parameters - O is tpxtp

# second differentials wrt psi_rs and sigma_uv
O<-matrix(0,tp,tp)
evec<-mvec<-list()
E<-M<-list()
l<-(tp-length(psi))
for (s in 1:(t-1))
{ for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for ( k in (1+qran+1):(1+qran+sigphi))
  {q<- k-(1+qran)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]]<-lamt%*%dS[[q]] - dS[[q]]%*%t(lamt)
   evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
   mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
   }
}
}
#stop (print (mvec))
#

# second differentials wrt psi_rs and lambda_uv
l<-(tp-length(psi))
for (s in 1:(t-1))
{ for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for (k in (1+qran+sigphi+1):(1+qran+totphi))
   {q<- k-(1+qran)
      mvec[[k]]<-matrix(0,t*(t-1)/2,1)
      M[[k]]<-dL[[q]]%*%sigt - sigt%*%t(dL[[q]])

     evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
```

```
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
    }
    }
 }




# second differential wrt sigma_rs and lambda_uv
for (l in (1+qran+1):(1+qran+sigphi))
  { p<-l-(1+qran)
  pvec[[l]]<-matrix(0,t*(t-1)/2,1)
   for (k in (1+qran+sigphi+1):(1+qran+totphi))
    {q<-k - (1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-dL[[q]]%*%dS[[p]] - dS[[p]]%*%t(dL[[q]])
     pvec[[l]] <-as.vector(psimat[lower.tri(psimat)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(pvec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
     }
    }



list(phi=phi,constraint=c,dcons=dcon,dconspsi=dconpsi,psi=psi,Omat=O)
}

#
# end dconsgen
#############################################################################################
#
# USsym
#############################################################################################
USsym<-function(r,t,phi=NULL, psi=NULL)
{
# forms the inverse sig and derivatives of sigma for the MVAR1 model
# with symmetric Lambda (t times)and Sigma = US (symmetric) with constraints)
# r=rows t=times phi = sigma_11 sigma_12, sigma_13, ...sigma_1t, sigma_22,
#sigma_23, ... lambda_11, lambda_12,...lambda_1t,lambda_22, lambda_23,...
# Mar2010
#JDF
#############################################################################################
F1<-mydiag(1,1,r-1,r-1)+mydiag(1,-1,r-1,r-1)
E1<- diag(c(0,rep(1,r-2),0))
Ir<-diag(r)
It<-diag(t)
Irt<-diag(r*t)
Jt<-matrix(1,nrow=t,ncol=t)

sigphi<-t*(t+1)/2
```

```
totphi<-t*(t+1)
lamphi<- totphi-sigphi
Sigma<-matrix(0,t,t)
ei<-ej<-matrix(0,t,1)
dS<-dSo<-list()
k<-0
for (i in 1:t)
{for (j in i:t)
{k<-k+1
Sigma[i,j]<- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dSo[[k]]<-ei%*%t(ej)
dS[[k]]<- dSo[[k]] + t(dSo[[k]]) - diag(diag(dSo[[k]]))
}
}
sigt<-Sigma+t(Sigma)-diag(diag(Sigma))
# Lambda
lam<-lamt<-matrix(0,t,t)
dL<-dLo<-list()
k<-(sigphi)
for (i in 1:t)
{ for (j in i:t)
{k <- k+1
lam[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dLo[[k]]<-ei%*%t(ej)
dL[[k]]<-dLo[[k]] + t(dLo[[k]]) - diag(diag(dLo[[k]]))
}
}
lamt<-lam+t(lam)-diag(diag(lam))
sti<-ginv.new(sigt)$invx
s0<-sti%*%ginv.new(It-(lamt%*%lamt))$invx
s1<-kronecker(Ir,s0)
s2<-kronecker(F1,lamt)
s3<-kronecker(E1,(lamt%*%lamt))
sigi<-s1%*%(Irt-s2+s3)
sig<-ginv.new(sigi)$invx
P1<-s1
P2<-Irt-s2+s3
# diff wrt phi[k]
dP1<-list()
dP2<-list()
for (k in 1:sigphi)
{dP1[[k]]<- kronecker(Ir, -sti%*%(dS[[k]])%*%s0)
dP2[[k]]<-0*Irt
}
for (k in (1+sigphi):totphi)
{dP1[[k]]<- kronecker(Ir, s0%*%(lamt%*%dL[[k]]+
 dL[[k]]%*%lamt)%*%ginv.new(It-(lamt%*%lamt))$invx )
dP2[[k]]<-kronecker(-F1,dL[[k]]) + kronecker(E1,(lamt%*%dL[[k]]+dL[[k]]%*%lamt))
}
```

```
dsigi<-list()
dsig<-list()
for (k in 1:totphi)
{dsigi[[k]]<-dP1[[k]]%*%P2 + P1%*%dP2[[k]]
dsig[[k]]<- -sig%*%dsigi[[k]]%*%sig
}
list(sig=sig,sigi=sigi,dsig=dsig,phi=phi,partype=rep(1,length(phi)))
}
#
# end USsym
#############################################################
#
# USsym constraints
################################################
USsymcons<-function(r,t,tp,qran,rowmod)
{
phi<-rowmod$phi
psi<-rowmod$psi
sigphi<-t*(t+1)/2
totphi<-t*(t+1)
lamphi<- totphi-sigphi
# function to get constraints and differentials of constraints

# allocate sigmas across top half of matrix (then make into a symmetric matrix)
Sigma<-matrix(0,t,t)
dS<-dSo<-list()
k<-0
for (i in 1:t)
{for (j in i:t)
  {k<-k+1
  Sigma[i,j]<- phi[k]
  ei<-ej<-matrix(0,t,1)
  ei[i,1]<-1
  ej[j,1]<-1
  dSo[[k]]<-ei%*%t(ej)
  dS[[k]]<- dSo[[k]] + t(dSo[[k]]) - diag(diag(dSo[[k]]))
   }
}
sigt<-Sigma+t(Sigma)-diag(diag(Sigma))

# allocate lambdas across matrix (symmetric)
# Lambda
lam<-lamt<-matrix(0,t,t)
dL<-dLo<-list()
k<-(sigphi)
for (i in 1:t)
{ for (j in i:t)
{k <- k+1
lam[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dLo[[k]]<-ei%*%t(ej)
dL[[k]]<-dLo[[k]] + t(dLo[[k]]) - diag(diag(dLo[[k]]))
}
```

```
}
lamt<-lam+t(lam)-diag(diag(lam))

# now for constraints
dcons<-matrix(0,t,t)
psimat<-matrix(0,t,t)
cons<-matrix(0,t,t)
dcon<-list()
dconpsi<-list()
c<-list()

l<-0
# allocate psi[k] across top half of psimat (then make as a symmetric matrix)
for (i in 1:(t-1))
{ for (j in (i+1):t)
  { l<-l+1
    psimat[i,j]<-psi[l]
#    cons[i,j]<-sum(lamt[i,1:t]*sigt[1:t,j]) - sum(sigt[i,1:t]*lamt[j,1:t])
#    c[[l]]<- cons[i,j]
  }
}
# now create full matrices
psimatf<-psimat+t(psimat)-diag(diag(psimat))
psimat<-psimatf
#cons<-consf
# now differentiate the constraints wrt phi[k] - from Sigma  - sigma_rs
pvec<-mvec<-list()
M<-list()
 for (k in 1:sigphi)
 {
   pvec[[k]]<-matrix(0,t*(t+1)/2,1)
   mvec[[k]]<-matrix(0,t*(t+1)/2,1)
   M[[k]] <- lamt%*%dS[[k]] - dS[[k]]%*%t(lamt)
   pvec[[k]]<-as.vector(psimat[lower.tri(psimat)])
  mvec[[k]]<-as.vector(M[[k]][lower.tri(M[[k]])])
  dcon[[k]] <-(t(pvec[[k]])%*%mvec[[k]])
  }

# now differentiate C wrt phi[k] - from Lambda - lambda_ij
for (k in (sigphi+1):totphi)
{
  pvec[[k]]<-matrix(0,t*(t+1)/2,1)
  mvec[[k]]<-matrix(0,t*(t+1)/2,1)
   M[[k]] <- dL[[k]]%*%sigt - sigt%*%t(dL[[k]])
  pvec[[k]]<-as.vector(psimat[lower.tri(psimat)])
  mvec[[k]]<-as.vector(M[[k]][lower.tri(M[[k]])])
  dcon[[k]] <-(t(pvec[[k]])%*%mvec[[k]])
  }

# do using vech
P<-list()
l<-0
for (s in 1:(t-1))
 {for (r in (s+1):t)
  {l<-l+1
```

```
  pvec[[l]]<-mvec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  P[[l]]<-er%*%t(es)
  M[[l]]<-lamt%*%sigt - sigt%*%t(lamt)
  pvec[[l]]<-as.vector(P[[l]][lower.tri(P[[l]])])
  mvec[[l]]<-as.vector(M[[l]][lower.tri(M[[l]])])
  dconpsi[[l]]<-t(pvec[[l]])%*%mvec[[l]]
  # match up constraints with psis in correct order
  c[[l]]<-dconpsi[[l]]
  }
}



# now need to adjust AI matrix with second derivatives where relevant
# use matrix O to add extra bits to AI matrix (AI in order sigma2, gammas,
# phi[k] , phi_col, psi[l])
# tp= total number of parameters - O is tpxtp

# second differentials wrt psi_rs and sigma_uv
O<-matrix(0,tp,tp)
evec<-mvec<-list()
E<-M<-list()
l<-(tp-length(psi))
for (s in 1:(t-1))
{ for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for ( k in (1+qran+1):(1+qran+sigphi))
  {q<- k-(1+qran)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]]<-lamt%*%dS[[q]] - dS[[q]]%*%t(lamt)
   evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
   mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
   }
}
}
#stop (print (mvec))
#

# second differentials wrt psi_rs and lambda_uv
l<-(tp-length(psi))
for (s in 1:(t-1))
{ for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
```

```
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for (k in (1+qran+sigphi+1):(1+qran+totphi))
   {q<- k-(1+qran)
      mvec[[k]]<-matrix(0,t*(t-1)/2,1)
      M[[k]]<-dL[[q]]%*%sigt - sigt%*%t(dL[[q]])

     evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
    }
    }
 }


# second differential wrt sigma_rs and lambda_uv
for (l in (1+qran+1):(1+qran+sigphi))
  { p<-l-(1+qran)
  pvec[[l]]<-matrix(0,t*(t-1)/2,1)
   for (k in (1+qran+sigphi+1):(1+qran+totphi))
    {q<-k - (1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-dL[[q]]%*%dS[[p]] - dS[[p]]%*%t(dL[[q]])
     pvec[[l]] <-as.vector(psimat[lower.tri(psimat)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(pvec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
     }
    }



list(phi=phi,constraint=c,dcons=dcon,dconspsi=dconpsi,psi=psi,Omat=O)
}
#
#end USsymcons
#
##############################################################################
##############################################################################
#
# ar1hgen
##############################################################################
ar1hgen<-function(r,t,phi=c(rep(0.1,(t+1+(t*(t+1)/2)))), psi=NULL)
{
# forms the inverse sigma and derivatives of sigma for
# the MVAR1 model with
# general nonsymmetric Lambda= (t times)and
# Sigma = ar1h heterogeneous ar1 process with constraints)
# r=rows t=times phi = sigma_11 sigma_22, sigma_33, ...sigma_tt,
# phi_ar1, ... lambda_11, lambda_12,..lambda_1t,lambda_21,....
# July2010
#JDF
##############################################################################
```

```
F1<-mydiag(1,1,r-1,r-1)+mydiag(1,-1,r-1,r-1)
E1<- diag(c(0,rep(1,r-2),0))
F1t<-mydiag(1,1,t-1,t-1)+mydiag(1,-1,t-1,t-1)
E1t<- diag(c(0,rep(1,t-2),0))
Ir<-diag(r)
It<-diag(t)
Irt<-diag(r*t)
Jt<-matrix(1,nrow=t,ncol=t)
sigphi<-(t+1)
totphi<-(t+1)+ (t*t)
lamphi<- totphi-sigphi
#Sigma<-matrix(0,t,t)
D12<-matrix(0,t,t)
SigAR1in<-matrix(0,t,t)
ei<-ej<-matrix(0,t,1)
dSin<-dSo<-dD12<-list()
k<-0
for (i in 1:t)
{k<-k+1
ei<-matrix(0,t,1)
ei[i,1]<-1
D12[i,i]<-(phi[k])^(-0.5)
dD12[[k]]<-(ei%*%t(ei))*(-0.5)*(phi[k])^(-1.5)
}
k<-t+1
SigAR1in<- (1-phi[k]^2)^(-1)*(It + phi[k]^2*E1t - phi[k]*F1t)
sti<-D12%*%SigAR1in%*%D12
sigt<-ginv.new(sti)$invx
# Lambda
lamt<-matrix(0,t,t)
dL<-list()
k<-(t+1)
for (i in 1:t)
{ for (j in 1:t)
{k <- k+1
lamt[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dL[[k]]<-ei%*%t(ej)
}
}
s0<-sti%*%ginv.new(It-(lamt%*%lamt))$invx
s1<-kronecker(Ir,s0)
s2<-kronecker(F1,lamt)
s3<-kronecker(E1,(lamt%*%lamt))
sigi<-s1%*%(Irt-s2+s3)
sig<-ginv.new(sigi)$invx
P1<-s1
P2<-Irt-s2+s3
# diff wrt phi[k]
dP1<-list()
dP2<-list()
for (k in 1:t)
{dSin[[k]]<-dD12[[k]]%*%(SigAR1in%*%D12) + D12%*%SigAR1in%*%dD12[[k]]
```

```
dP1[[k]]<- kronecker(Ir, dSin[[k]]%*%ginv.new(It-(lamt%*%lamt))$invx)
dP2[[k]]<-0*Irt
}

for (k in (t+1):(t+1))
{dSin[[k]]<-D12%*%((1-phi[k]^2)^(-1)*(2*phi[k]*E1t-F1t)+
 (2*phi[k]*(1-phi[k]^2)^(-2))*(It + phi[k]^2*E1t-phi[k]*F1t))%*%D12
dP1[[k]]<- kronecker(Ir, dSin[[k]]%*%ginv.new(It-(lamt%*%lamt))$invx)
dP2[[k]]<-0*Irt}
for (k in (1+sigphi):totphi)
{dP1[[k]]<- kronecker(Ir, s0%*%(lamt%*%dL[[k]]+
 dL[[k]]%*%lamt)%*%ginv.new(It-(lamt%*%lamt))$invx )
dP2[[k]]<-kronecker(-F1,dL[[k]]) + kronecker(E1,(lamt%*%dL[[k]]+dL[[k]]%*%lamt))
}
dsigi<-list()
dsig<-list()
for (k in 1:totphi)
{dsigi[[k]]<-dP1[[k]]%*%P2 + P1%*%dP2[[k]]
dsig[[k]]<- -sig%*%dsigi[[k]]%*%sig
}
list(sig=sig,sigi=sigi,dsig=dsig,phi=phi,partype=rep(1,length(phi)))
}
#
#end ar1hgen
################################################################
#
# ar1hgen constraints
################################################
ar1hgencons<-function(r,t,tp,qran,rowmod)

{
phi<-rowmod$phi
psi<-rowmod$psi

# function to get constraints and differentials of constraints

# Sigma & SigmaIn
F1<-mydiag(1,1,r-1,r-1)+mydiag(1,-1,r-1,r-1)
E1<- diag(c(0,rep(1,r-2),0))
F1t<-mydiag(1,1,t-1,t-1)+mydiag(1,-1,t-1,t-1)
E1t<- diag(c(0,rep(1,t-2),0))
Ir<-diag(r)
It<-diag(t)
Irt<-diag(r*t)
Jt<-matrix(1,nrow=t,ncol=t)
sigphi<-(t+1)
totphi<-(t+1)+ t*t
lamphi<- totphi-sigphi
#Sigma<-matrix(0,t,t)
D12<-matrix(0,t,t)
SigAR1in<-matrix(0,t,t)
ei<-ej<-matrix(0,t,1)
dSin<-dSo<-dD12<-dD122<-dSinps<-d2Sin<-list()
k<-0
for (i in 1:t)
```

```
{k<-k+1
ei<-matrix(0,t,1)
ei[i,1]<-1
D12[i,i]<-(phi[k])^(-0.5)
dD12[[k]]<-(ei%*%t(ei))*-0.5*(phi[k])^(-1.5)
dD122[[k]]<- 0.75*(phi[k]^(-2.5))*(ei%*%t(ei))
}
k<-t+1
SigAR1in<- (1-phi[k]^2)^(-1)*(It + phi[k]^2*E1t - phi[k]*F1t)

sti<-D12%*%SigAR1in%*%D12
sigt<-ginv.new(sti)$invx
# Lambda
lamt<-matrix(0,t,t)
dL<-list()
k<-(t+1)
for (i in 1:t)
{ for (j in 1:t)
{k <- k+1
lamt[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dL[[k]]<-ei%*%t(ej)
}
}

# dSin - wrt sigma_ii
for (k in 1:t)
{dSin[[k]]<-dD12[[k]]%*%(SigAR1in%*%D12) + D12%*%SigAR1in%*%dD12[[k]]
dSAR1v<- (1-phi[t+1]^2)^(-1)*(2*phi[t+1]*E1t-F1t)+
  (2*phi[t+1]/(1-phi[t+1]^2)^2)*(It + phi[t+1]^2*E1t-phi[t+1]*F1t)
dSinps[[k]]<-dD12[[k]]%*%dSAR1v%*%D12 + D12%*%dSAR1v%*%dD12[[k]]
d2Sin[[k]]<- dD122[[k]]%*%SigAR1in%*%D12 + D12%*%SigAR1in%*%dD122[[k]]+
  2*dD12[[k]]%*%SigAR1in%*%dD12[[k]]
}
# diff Sigma inverse wrt phi
for (k in (t+1):(t+1))
{dSin[[k]]<-D12%*%dSAR1v%*%D12
 d2Sin[[k]]<-D12%*%(2*(1-phi[k]^2)^(-1)*E1t +
 4*phi[k]*(1-phi[k]^2)^(-2)*(2*phi[k]*E1t-F1t) +
 (2*(1-phi[k]^2)^(-2) + 8*phi[k]^2*(1-phi[k]^2)^(-3))*(It+phi[k]^2*E1t-phi[k]*F1t) )%*%D
}


# psimat

# now for constraints
dcons<-matrix(0,t,t)
psimat<-matrix(0,t,t)
cons<-matrix(0,t,t)
dcon<-list()
dconpsi<-list()
c<-list()
l<-0
```

```
# allocate psi[k] across top half of psimat not diagonals then make symmetric
for (i in 1:(t-1))
{ for (j in (i+1):t)
  { l<-l+1
    psimat[i,j]<-psi[l]
  }
}

# now create full matrices
psimatf<-psimat+t(psimat)-diag(diag(psimat))
psimat<-psimatf
# diff C wrt phi[[k]]
# sigma_rs
pvec<-mvec<-list()
M<-list()
 for (k in 1:sigphi)
{  pvec[[k]]<-matrix(0,t*(t-1)/2,1)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]] <- dSin[[k]]%*%lamt - t(lamt)%*%dSin[[k]]
   pvec[[k]]<-as.vector(psimat[lower.tri(psimat)])
   mvec[[k]]<-as.vector(M[[k]][lower.tri(M[[k]])])
   dcon[[k]] <-(t(pvec[[k]])%*%mvec[[k]])
}


# phi
  k<-t+1
  pvec[[k]]<-matrix(0,t*(t-1)/2,1)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]] <- dSin[[k]]%*%lamt - t(lamt)%*%dSin[[k]]
   pvec[[k]]<-as.vector(psimat[lower.tri(psimat)])
   mvec[[k]]<-as.vector(M[[k]][lower.tri(M[[k]])])
   dcon[[k]] <-(t(pvec[[k]])%*%mvec[[k]])

# lambda_rs
for (k in (sigphi+1):totphi)
{  pvec[[k]]<-matrix(0,t*(t-1)/2,1)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]] <- sti%*%dL[[k]] - t(dL[[k]])%*%sti
   pvec[[k]]<-as.vector(psimat[lower.tri(psimat)])
   mvec[[k]]<-as.vector(M[[k]][lower.tri(M[[k]])])
   dcon[[k]] <-(t(pvec[[k]])%*%mvec[[k]])
 }
#diff C wrt psi[[l]]  - gives the constraints c
P<-list()
l<-0
for (s in 1:(t-1))
 {for (r in (s+1):t)
  {l<-l+1
  pvec[[l]]<-mvec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  P[[l]]<-er%*%t(es)
  M[[l]]<-sti%*%lamt - t(lamt)%*%sti
```

```
  pvec[[l]]<-as.vector(P[[l]][lower.tri(P[[l]])])
  mvec[[l]]<-as.vector(M[[l]][lower.tri(M[[l]])])
  dconpsi[[l]]<-t(pvec[[l]])%*%mvec[[l]]
 # match up constarints with psis in correct order
  c[[l]]<-dconpsi[[l]]
  }
}

# second derivatives of C
# now need to adjust AI matrix with second derivatives where relevant
# use matrix O to add extra bits to AI matrix (AI in order sigma2, gammas,
# phi[k] , phi_col, psi[l])
# tp= total number of parameters - O is tpxtp
O<-matrix(0,tp,tp)
evec<-mvec<-list()
E<-M<-list()

# second differential wrt sigma_rr^2
   for (k in (1+qran+1):(1+qran+t))
    {q<-k-(1+qran)
     pvec[[k]]<-mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-d2Sin[[q]]%*%lamt - t(lamt)%*%d2Sin[[q]]
     pvec[[k]] <-as.vector(psimat[lower.tri(psimat)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
     O[k,k]<- -(t(pvec[[k]])%*%mvec[[k]])
    }

# second differential wrt phi^2
     k <-1+qran+t+1
     q<-k - (1+qran)
     pvec[[k]]<-mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-d2Sin[[q]]%*%lamt - t(lamt)%*%d2Sin[[q]]
     pvec[[k]] <-as.vector(psimat[lower.tri(psimat)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
     O[k,k]<- -(t(pvec[[k]])%*%mvec[[k]])

# second differential wrt sigma_rr and lambda_uv
for (l in (1+qran+1):(1+qran+t))
{p<-l-(1+qran)
  pvec[[l]]<-matrix(0,t*(t-1)/2,1)
  for (k in (1+qran+sigphi+1):(1+qran+totphi))
 {   q<-k -(1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-dSin[[p]]%*%dL[[q]] - t(dL[[q]])%*%dSin[[p]]
     pvec[[l]] <-as.vector(psimat[lower.tri(psimat)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
     O[l,k]<- -(t(pvec[[l]])%*%mvec[[k]])
     O[k,l]<-O[l,k]
 }
}

# second differential wrt phi and lambda_uv
l<-1+ qran+t+1
p<-l-(1+qran)
  pvec[[l]]<-matrix(0,t*(t-1)/2,1)
```

```
  for (k in (1+qran+sigphi+1):(1+qran+totphi))
   {q<-k-(1+qran)
    mvec[[k]]<-matrix(0,t*(t-1)/2,1)
    M[[k]]<-dSin[[p]]%*%dL[[q]] - t(dL[[q]])%*%dSin[[p]]
    pvec[[l]] <-as.vector(psimat[lower.tri(psimat)])
    mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
    O[l,k]<- -(t(pvec[[l]])%*%mvec[[k]])
    O[k,l]<-O[l,k]
   }


   # second differential wrt phi and sigma_rr
  l<-1+ qran+t+1
  pvec[[l]]<-matrix(0,t*(t-1)/2,1)
  for (k in (1+qran+1):(1+qran+t))
   {q<-k-(1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-dSinps[[q]]%*%lamt - t(lamt)%*%dSinps[[q]]
     pvec[[l]] <-as.vector(psimat[lower.tri(psimat)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
     O[l,k]<- -(t(pvec[[l]])%*%mvec[[k]])
     O[k,l]<-O[l,k]
   }

# second differentials wrt psi_rs and sigma_uu
# phi[[l]] and phi[[k]]
evec<-mvec<-list()
E<-M<-list()
l<-(tp-length(psi))
for (s in 1:(t-1))
 { for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
 for (k in (1+qran+1):(1+qran+t))
   {q<-k-(1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-dSin[[q]]%*%lamt - t(lamt)%*%dSin[[q]]
     evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
     O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
     O[k,l]<-O[l,k]
   }
 }
}



# second differentials wrt psi_rs and lambda_uv
l<-(tp-length(psi))
for (s in 1:(t-1))
{ for (r in (s+1):t)
```

```
 {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
  for (k in (1+qran+sigphi+1):(1+qran+totphi))
  { q<-k - (1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-sti%*%dL[[q]] - t(dL[[q]])%*%sti
     evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
     O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
     O[k,l]<-O[l,k]
  }
 }
}
# second differentials wrt psi_rs and phi
# phi[[l]] and phi[[t+1]]
evec<-mvec<-list()
E<-M<-list()
l<-(tp-length(psi))
for (s in 1:(t-1))
 { for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
    k<-1+qran+t+1
    q<-k - (1+qran)
    mvec[[k]]<-matrix(0,t*(t-1)/2,1)
    M[[k]]<-dSin[[q]]%*%lamt - t(lamt)%*%dSin[[q]]
    evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
    mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
    O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
    O[k,l]<-O[l,k]
   }
}
#print (O)
list(phi=phi,constraint=c,dcons=dcon,dconspsi=dconpsi,psi=psi,Omat=O)
}

#
end ar1hgencons
#############################################################################
#
# ar1hsym
#############################################################################
ar1hsym<-function(r,t,phi=c(rep(0.1,(t+1+(t*(t+1)/2)))), psi=NULL)
{
# forms the inverse sig and derivatives of sigma for the MVAR1 model with
#symmetric Lambda= (t times)and
# Sigma = ar1h heterogeneous ar1 process with constraints)
```

```
# r=rows t=times phi = sigma_11 sigma_22, sigma_33, ...sigma_tt,
# phi_ar1, ... lambda_11, lambda_12,..lambda_1t,lambda_22,....
# July2010
#JDF
#############################################################################
F1<-mydiag(1,1,r-1,r-1)+mydiag(1,-1,r-1,r-1)
E1<- diag(c(0,rep(1,r-2),0))
F1t<-mydiag(1,1,t-1,t-1)+mydiag(1,-1,t-1,t-1)
E1t<- diag(c(0,rep(1,t-2),0))
Ir<-diag(r)
It<-diag(t)
Irt<-diag(r*t)
Jt<-matrix(1,nrow=t,ncol=t)
sigphi<-(t+1)
totphi<-(t+1)+ t*(t+1)/2
lamphi<- totphi-sigphi
#Sigma<-matrix(0,t,t)
D12<-matrix(0,t,t)
SigAR1in<-matrix(0,t,t)
ei<-ej<-matrix(0,t,1)
dSin<-dSo<-dD12<-list()
k<-0
for (i in 1:t)
{k<-k+1
ei<-matrix(0,t,1)
ei[i,1]<-1
D12[i,i]<-(phi[k])^(-0.5)
dD12[[k]]<-(ei%*%t(ei))*(-0.5)*(phi[k])^(-1.5)
}
k<-t+1
SigAR1in<- (1-phi[k]^2)^(-1)*(It + phi[k]^2*E1t - phi[k]*F1t)
sti<-D12%*%SigAR1in%*%D12
sigt<-ginv.new(sti)$invx
# Lambda
lam<-lamt<-matrix(0,t,t)
dL<-dLo<-list()
k<-(sigphi)
for (i in 1:t)
{ for (j in i:t)
{k <- k+1
lam[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dLo[[k]]<-ei%*%t(ej)
dL[[k]]<-dLo[[k]] + t(dLo[[k]]) - diag(diag(dLo[[k]]))
}
}
lamt<-lam+t(lam)-diag(diag(lam))

s0<-sti%*%ginv.new(It-(lamt%*%lamt))$invx
s1<-kronecker(Ir,s0)
s2<-kronecker(F1,lamt)
s3<-kronecker(E1,(lamt%*%lamt))
sigi<-s1%*%(Irt-s2+s3)
```

```
sig<-ginv.new(sigi)$invx
P1<-s1
P2<-Irt-s2+s3
# diff wrt phi[k]
dP1<-list()
dP2<-list()
for (k in 1:t)
{dSin[[k]]<-dD12[[k]]%*%(SigAR1in%*%D12) + D12%*%SigAR1in%*%dD12[[k]]
dP1[[k]]<- kronecker(Ir, dSin[[k]]%*%ginv.new(It-(lamt%*%lamt))$invx)
dP2[[k]]<-0*Irt
}

for (k in (t+1):(t+1))
{dSin[[k]]<-D12%*%((1-phi[k]^2)^(-1)*(2*phi[k]*E1t-F1t)+
 (2*phi[k]*(1-phi[k]^2)^(-2))*(It + phi[k]^2*E1t-phi[k]*F1t))%*%D12
dP1[[k]]<- kronecker(Ir, dSin[[k]]%*%ginv.new(It-(lamt%*%lamt))$invx)
dP2[[k]]<-0*Irt}
for (k in (1+sigphi):totphi)
{dP1[[k]]<- kronecker(Ir, s0%*%(lamt%*%dL[[k]]+
 dL[[k]]%*%lamt)%*%ginv.new(It-(lamt%*%lamt))$invx )
dP2[[k]]<-kronecker(-F1,dL[[k]]) + kronecker(E1,(lamt%*%dL[[k]]+dL[[k]]%*%lamt))
}
dsigi<-list()
dsig<-list()
for (k in 1:totphi)
{dsigi[[k]]<-dP1[[k]]%*%P2 + P1%*%dP2[[k]]
dsig[[k]]<- -sig%*%dsigi[[k]]%*%sig
}
list(sig=sig,sigi=sigi,dsig=dsig,phi=phi,partype=rep(1,length(phi)))
}
#
# end ar1hsym
#
#############################################################
#
# ar1hsym constraints
###############################################
ar1hsymcons<-function(r,t,tp,qran,rowmod)

{
phi<-rowmod$phi
psi<-rowmod$psi

# function to get constraints and differentials of constraints

# Sigma & SigmaIn
F1<-mydiag(1,1,r-1,r-1)+mydiag(1,-1,r-1,r-1)
E1<- diag(c(0,rep(1,r-2),0))
F1t<-mydiag(1,1,t-1,t-1)+mydiag(1,-1,t-1,t-1)
E1t<- diag(c(0,rep(1,t-2),0))
Ir<-diag(r)
It<-diag(t)
Irt<-diag(r*t)
Jt<-matrix(1,nrow=t,ncol=t)
sigphi<-(t+1)
```

```
totphi<-(t+1)+ t*(t+1)/2
lamphi<- totphi-sigphi
#Sigma<-matrix(0,t,t)
D12<-matrix(0,t,t)
SigAR1in<-matrix(0,t,t)
ei<-ej<-matrix(0,t,1)
dSin<-dSo<-dD12<-dD122<-dSinps<-d2Sin<-list()
k<-0
for (i in 1:t)
{k<-k+1
ei<-matrix(0,t,1)
ei[i,1]<-1
D12[i,i]<-(phi[k])^(-0.5)
dD12[[k]]<-(ei%*%t(ei))*-0.5*(phi[k])^(-1.5)
dD122[[k]]<- 0.75*(phi[k]^(-2.5))*(ei%*%t(ei))
}
k<-t+1
SigAR1in<- (1-phi[k]^2)^(-1)*(It + phi[k]^2*E1t - phi[k]*F1t)

sti<-D12%*%SigAR1in%*%D12
sigt<-ginv.new(sti)$invx
# Lambda
# Lambda
lam<-lamt<-matrix(0,t,t)
dL<-dLo<-list()
k<-(sigphi)
for (i in 1:t)
{ for (j in i:t)
{k <- k+1
lam[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dLo[[k]]<-ei%*%t(ej)
dL[[k]]<-dLo[[k]] + t(dLo[[k]]) - diag(diag(dLo[[k]]))
}
}
lamt<-lam + t(lam)-diag(diag(lam))
# dSin - wrt sigma_ii
for (k in 1:t)
{dSin[[k]]<-dD12[[k]]%*%(SigAR1in%*%D12) + D12%*%SigAR1in%*%dD12[[k]]
dSAR1v<- (1-phi[t+1]^2)^(-1)*(2*phi[t+1]*E1t-F1t)+
 (2*phi[t+1]/(1-phi[t+1]^2)^2)*(It + phi[t+1]^2*E1t-phi[t+1]*F1t)
dSinps[[k]]<-dD12[[k]]%*%dSAR1v%*%D12 + D12%*%dSAR1v%*%dD12[[k]]
d2Sin[[k]]<- dD122[[k]]%*%SigAR1in%*%D12 + D12%*%SigAR1in%*%dD122[[k]]+
  2*dD12[[k]]%*%SigAR1in%*%dD12[[k]]
}
# diff Sigma inverse wrt phi
for (k in (t+1):(t+1))
{dSin[[k]]<-D12%*%dSAR1v%*%D12
 d2Sin[[k]]<-D12%*%(2*(1-phi[k]^2)^(-1)*E1t +
  4*phi[k]*(1-phi[k]^2)^(-2)*(2*phi[k]*E1t-F1t) +
  (2*(1-phi[k]^2)^(-2) + 8*phi[k]^2*(1-phi[k]^2)^(-3))*
  (It+phi[k]^2*E1t-phi[k]*F1t) )%*%D12
}
```

```
# psimat

# now for constraints
dcons<-matrix(0,t,t)
psimat<-matrix(0,t,t)
cons<-matrix(0,t,t)
dcon<-list()
dconpsi<-list()
c<-list()
l<-0
# allocate psi[k] across top half of psimat not diagonals then make symmetric
for (i in 1:(t-1))
{ for (j in (i+1):t)
  { l<-l+1
    psimat[i,j]<-psi[l]
  }
}

# now create full matrices
psimatf<-psimat+t(psimat)-diag(diag(psimat))
psimat<-psimatf
# diff C wrt phi[[k]]
# sigma_rs
pvec<-mvec<-list()
M<-list()
 for (k in 1:sigphi)
{  pvec[[k]]<-matrix(0,t*(t-1)/2,1)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]] <- dSin[[k]]%*%lamt - t(lamt)%*%dSin[[k]]
   pvec[[k]]<-as.vector(psimat[lower.tri(psimat)])
   mvec[[k]]<-as.vector(M[[k]][lower.tri(M[[k]])])
   dcon[[k]] <-(t(pvec[[k]])%*%mvec[[k]])
}


# phi
  k<-t+1
  pvec[[k]]<-matrix(0,t*(t-1)/2,1)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]] <- dSin[[k]]%*%lamt - t(lamt)%*%dSin[[k]]
   pvec[[k]]<-as.vector(psimat[lower.tri(psimat)])
   mvec[[k]]<-as.vector(M[[k]][lower.tri(M[[k]])])
   dcon[[k]] <-(t(pvec[[k]])%*%mvec[[k]])

# lambda_rs
for (k in (sigphi+1):totphi)
{  pvec[[k]]<-matrix(0,t*(t-1)/2,1)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]] <- sti%*%dL[[k]] - t(dL[[k]])%*%sti
   pvec[[k]]<-as.vector(psimat[lower.tri(psimat)])
   mvec[[k]]<-as.vector(M[[k]][lower.tri(M[[k]])])
   dcon[[k]] <-(t(pvec[[k]])%*%mvec[[k]])
 }
```

```
#diff C wrt psi[[l]]  - gives the constraints c
P<-list()
l<-0
for (s in 1:(t-1))
 {for (r in (s+1):t)
  {l<-l+1
  pvec[[l]]<-mvec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  P[[l]]<-er%*%t(es)
  M[[l]]<-sti%*%lamt - t(lamt)%*%sti
  pvec[[l]]<-as.vector(P[[l]][lower.tri(P[[l]])])
  mvec[[l]]<-as.vector(M[[l]][lower.tri(M[[l]])])
  dconpsi[[l]]<-t(pvec[[l]])%*%mvec[[l]]
 # match up constarints with psis in correct order
  c[[l]]<-dconpsi[[l]]
  }
}

# second derivatives of C
# now need to adjust AI matrix with second derivatives where relevant
# use matrix O to add extra bits to AI matrix (AI in order sigma2, gammas,
# phi[k] , phi_col, psi[l])
# tp= total number of parameters - O is tpxtp
O<-matrix(0,tp,tp)
evec<-mvec<-list()
E<-M<-list()

# second differential wrt sigma_rr^2
   for (k in (1+qran+1):(1+qran+t))
    {q<-k-(1+qran)
     pvec[[k]]<-mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-d2Sin[[q]]%*%lamt - t(lamt)%*%d2Sin[[q]]
     pvec[[k]] <-as.vector(psimat[lower.tri(psimat)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
     O[k,k]<- -(t(pvec[[k]])%*%mvec[[k]])
    }

# second differential wrt phi^2
     k <-1+qran+t+1
     q<-k - (1+qran)
     pvec[[k]]<-mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-d2Sin[[q]]%*%lamt - t(lamt)%*%d2Sin[[q]]
     pvec[[k]] <-as.vector(psimat[lower.tri(psimat)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
     O[k,k]<- -(t(pvec[[k]])%*%mvec[[k]])

# second differential wrt sigma_rr and lambda_uv
for (l in (1+qran+1):(1+qran+t))
{p<-l-(1+qran)
  pvec[[l]]<-matrix(0,t*(t-1)/2,1)
  for (k in (1+qran+sigphi+1):(1+qran+totphi))
 {   q<-k -(1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
```

```
     M[[k]]<-dSin[[p]]%*%dL[[q]] - t(dL[[q]])%*%dSin[[p]]
     pvec[[l]] <-as.vector(psimat[lower.tri(psimat)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
     O[l,k]<- -(t(pvec[[l]])%*%mvec[[k]])
     O[k,l]<-O[l,k]
 }
}

# second differential wrt phi and lambda_uv
l<-1+ qran+t+1
p<-l-(1+qran)
  pvec[[l]]<-matrix(0,t*(t-1)/2,1)
  for (k in (1+qran+sigphi+1):(1+qran+totphi))
   {q<-k-(1+qran)
    mvec[[k]]<-matrix(0,t*(t-1)/2,1)
    M[[k]]<-dSin[[p]]%*%dL[[q]] - t(dL[[q]])%*%dSin[[p]]
    pvec[[l]] <-as.vector(psimat[lower.tri(psimat)])
    mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
    O[l,k]<- -(t(pvec[[l]])%*%mvec[[k]])
    O[k,l]<-O[l,k]
   }


   # second differential wrt phi and sigma_rr
  l<-1+ qran+t+1
  pvec[[l]]<-matrix(0,t*(t-1)/2,1)
  for (k in (1+qran+1):(1+qran+t))
   {q<-k-(1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-dSinps[[q]]%*%lamt - t(lamt)%*%dSinps[[q]]
     pvec[[l]] <-as.vector(psimat[lower.tri(psimat)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
     O[l,k]<- -(t(pvec[[l]])%*%mvec[[k]])
     O[k,l]<-O[l,k]
   }

# second differentials wrt psi_rs and sigma_uu
# phi[[l]] and phi[[k]]
evec<-mvec<-list()
E<-M<-list()
l<-(tp-length(psi))
for (s in 1:(t-1))
 { for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
 for (k in (1+qran+1):(1+qran+t))
   {q<-k-(1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-dSin[[q]]%*%lamt - t(lamt)%*%dSin[[q]]
     evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
```

```
     O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
     O[k,l]<-O[l,k]
   }
 }
}



# second differentials wrt psi_rs and lambda_uv
l<-(tp-length(psi))
for (s in 1:(t-1))
{ for (r in (s+1):t)
 {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
  for (k in (1+qran+sigphi+1):(1+qran+totphi))
  { q<-k - (1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-sti%*%dL[[q]] - t(dL[[q]])%*%sti
     evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
     O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
     O[k,l]<-O[l,k]
  }
 }
}
# second differentials wrt psi_rs and phi
# phi[[l]] and phi[[t+1]]
evec<-mvec<-list()
E<-M<-list()
l<-(tp-length(psi))
for (s in 1:(t-1))
 { for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
    k<-1+qran+t+1
    q<-k - (1+qran)
    mvec[[k]]<-matrix(0,t*(t-1)/2,1)
    M[[k]]<-dSin[[q]]%*%lamt - t(lamt)%*%dSin[[q]]
    evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
    mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
    O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
    O[k,l]<-O[l,k]
   }
}
#print (O)
list(phi=phi,constraint=c,dcons=dcon,dconspsi=dconpsi,psi=psi,Omat=O)
}
```

```
#
#end ar1hsymcons
#

 ##############################################################################
#
# anteband
##############################################################################
anteband<-function(r,t,phi=NULL, psi=NULL)
{
# forms the inverse sig and derivatives of sigma for the MVAR1 model
# with banded tri diagonal NONSYMMETRIC Lambda (t times)and
# Sigma = ante1 with constraints)
# r=rows t=times
# phi = ds = 1/sigma_11 1/sigma_22, 1/sigma_33, ...1/sigma_tt,
#       cs = c_1,c_2,...c_t-1 (antedependence parameters)
#       lambdas = lambda_11 lambda_12 lambda_21 lambda_22 lambda_23
# lambda_32 lambda_33 .... lambda_rr
# Apr2011
#JDF
##############################################################################
F1<-mydiag(1,1,r-1,r-1)+mydiag(1,-1,r-1,r-1)
E1<- diag(c(0,rep(1,r-2),0))
Ir<-diag(r)
It<-diag(t)
Irt<-diag(r*t)
Jt<-matrix(1,nrow=t,ncol=t)

sigphi<-2*t-1
lamphi<-3*t-2
totphi<- lamphi+sigphi
Sigma<-matrix(0,t,t)
D<-matrix(0,t,t)
ei<-ej<-matrix(0,t,1)
dS<-dSo<-dD<-list()
k<-0
# di's in Sigma
for (i in 1:t)
{k<-k+1
ei<-matrix(0,t,1)
ei[i,1]<-1
D[i,i]<-phi[k]
dD[[k]]<-ei%*%t(ei)
}
k<-t
# ci's in Sigma
C<-matrix(0,t,t)
dC<-list()
for (i in 1:(t-1))
{k<-k+1
ei<-ei1<-matrix(0,t,1)
ei[i,1]<-1
ei1[i+1,1]<- 1
C[i+1,i]<- -phi[k]
C[i,i]<-1
```

```
C[t,t]<-1
dC[[k]]<- -ei1%*%t(ei)
}
sti<-matrix(0,t,t)
sti<-t(C)%*%D%*%C
#sigt<-ginv.new(sti)$invx

# Lambda
lam<-lamt<-matrix(0,t,t)
dL<-dLo<-list()
lam[1,1]<-phi[sigphi+1]
lam[1,2]<-phi[sigphi+2]
e1<-e2<-matrix(0,t,1)
e1[1,1]<-1
e2[2,1]<-1
dL[[sigphi+1]]<-e1%*%t(e1)
dL[[sigphi+2]]<-e1%*%t(e2)
k<-(sigphi+2)
for (i in 2:(t-1))
{ for (j in (i-1):(i+1))
{k <- k+1
lam[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1

dL[[k]]<-ei%*%t(ej)
}
}
et<-et1<-matrix(0,t,1)
et[t,1]<-1
et1[t-1,1]<-1
lam[t,t-1]<-phi[totphi-1]
lam[t,t]<-phi[totphi]
dL[[totphi-1]]<-et%*%t(et1)
dL[[totphi]]<-et%*%t(et)



s0<-sti%*%ginv.new(It-(lamt%*%lamt))$invx
s1<-kronecker(Ir,s0)
s2<-kronecker(F1,lamt)
s3<-kronecker(E1,(lamt%*%lamt))
sigi<-s1%*%(Irt-s2+s3)
sig<-ginv.new(sigi)$invx
P1<-s1
P2<-Irt-s2+s3
# diff wrt phi[k]
dP1<-list()
dP2<-list()
dSin<-list()
for (k in 1:t)
{dSin[[k]]<-t(C)%*%dD[[k]]%*%C
dP1[[k]]<- kronecker(Ir, dSin[[k]]%*%ginv.new(It-(lamt%*%lamt))$invx)
dP2[[k]]<-0*Irt
```

```
}
for (k in (t+1):(2*t-1))
{dSin[[k]]<-t(dC[[k]])%*%D%*%C + t(C)%*%D%*%dC[[k]]
dP1[[k]]<- kronecker(Ir, dSin[[k]]%*%ginv.new(It-(lamt%*%lamt))$invx)
dP2[[k]]<-0*Irt
}

for (k in (1+sigphi):totphi)
{dP1[[k]]<- kronecker(Ir, s0%*%(lamt%*%dL[[k]]+
 dL[[k]]%*%lamt)%*%ginv.new(It-(lamt%*%lamt))$invx )
dP2[[k]]<-kronecker(-F1,dL[[k]]) + kronecker(E1,(lamt%*%dL[[k]]+dL[[k]]%*%lamt))
}
dsigi<-list()
dsig<-list()
for (k in 1:totphi)
{dsigi[[k]]<-dP1[[k]]%*%P2 + P1%*%dP2[[k]]
dsig[[k]]<- -sig%*%dsigi[[k]]%*%sig
}
list(sig=sig,sigi=sigi,dsig=dsig,phi=phi,partype=rep(1,length(phi)))
}
#
# end anteband
############################################################
#
# anteband constraints
###############################################
antebandcons<-function(r,t,tp,qran,rowmod)
{
phi<-rowmod$phi
psi<-rowmod$psi
sigphi<-2*t-1
lamphi<-3*t-2
totphi<- lamphi+sigphi
# function to get constraints and differentials of constraints

# Sigma
Sigma<-matrix(0,t,t)
D<-matrix(0,t,t)
ei<-ej<-matrix(0,t,1)
dS<-dSo<-dD<-dC<-list()
k<-0
# di's in Sigma
for (i in 1:t)
{k<-k+1
ei<-matrix(0,t,1)
ei[i,1]<-1
D[i,i]<-phi[k]
dD[[k]]<-ei%*%t(ei)
}
k<-t
# ci's in Sigma
C<-matrix(0,t,t)
dC<-list()
for (i in 1:(t-1))
{k<-k+1
```

```
ei<-ei1<-matrix(0,t,1)
ei[i,1]<-1
ei1[i+1,1]<- 1
C[i+1,i]<- -phi[k]
C[i,i]<-1
C[t,t]<-1
dC[[k]]<- -ei1%*%t(ei)
}
sti<-matrix(0,t,t)
sti<-t(C)%*%D%*%C
sigt<-ginv.new(sti)$invx


# allocate lambdas across matrix (symmetric) tridiagonal
# Lambda
# Lambda
lam<-lamt<-matrix(0,t,t)
dL<-dLo<-list()
lam[1,1]<-phi[sigphi+1]
lam[1,2]<-phi[sigphi+2]
e1<-e2<-matrix(0,t,1)
e1[1,1]<-1
e2[2,1]<-1
dL[[sigphi+1]]<-e1%*%t(e1)
dL[[sigphi+2]]<-e1%*%t(e2)
k<-(sigphi+2)
for (i in 2:(t-1))
{ for (j in (i-1):(i+1))
{k <- k+1
lam[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1

dL[[k]]<-ei%*%t(ej)
}
}
et<-et1<-matrix(0,t,1)
et[t,1]<-1
et1[t-1,1]<-1
lam[t,t-1]<-phi[totphi-1]
lam[t,t]<-phi[totphi]
dL[[totphi-1]]<-et%*%t(et1)
dL[[totphi]]<-et%*%t(et)




# now for constraints
dcons<-matrix(0,t,t)
psimat<-matrix(0,t,t)
cons<-matrix(0,t,t)
dcon<-list()
```

```
dconpsi<-list()
c<-list()

l<-0
# allocate psi[k] across top half of psimat (then make as a symmetric matrix)
for (i in 1:(t-1))
{ for (j in (i+1):t)
  { l<-l+1
    psimat[i,j]<-psi[l]
#    cons[i,j]<-sum(lamt[i,1:t]*sigt[1:t,j]) - sum(sigt[i,1:t]*lamt[j,1:t])
#    c[[l]]<- cons[i,j]
  }
}
# now create full matrices
psimatf<-psimat+t(psimat)-diag(diag(psimat))
psimat<-psimatf
#cons<-consf
# now differentiate the constraints wrt phi[k] - from Sigma
# di's
pvec<-mvec<-dSin<-list()
M<-list()
 for (k in 1:t)
 { dSin[[k]]<-t(C)%*%dD[[k]]%*%C
   pvec[[k]]<-matrix(0,t*(t-1)/2,1)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]] <- dSin[[k]]%*%lamt - t(lamt)%*%dSin[[k]]
   pvec[[k]]<-as.vector(psimat[lower.tri(psimat)])
  mvec[[k]]<-as.vector(M[[k]][lower.tri(M[[k]])])
  dcon[[k]] <-(t(pvec[[k]])%*%mvec[[k]])
  }
# ci's
for (k in (t+1):(2*t-1))
{ dSin[[k]]<-t(dC[[k]])%*%D%*%C + t(C)%*%D%*%dC[[k]]
 pvec[[k]]<-matrix(0,t*(t-1)/2,1)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]] <- dSin[[k]]%*%lamt - t(lamt)%*%dSin[[k]]
   pvec[[k]]<-as.vector(psimat[lower.tri(psimat)])
  mvec[[k]]<-as.vector(M[[k]][lower.tri(M[[k]])])
  dcon[[k]] <-(t(pvec[[k]])%*%mvec[[k]])
  }

# now differentiate C wrt phi[k] - from Lambda - lambda_ij
for (k in (sigphi+1):totphi)
{
  pvec[[k]]<-matrix(0,t*(t-1)/2,1)
  mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]] <- sti%*%dL[[k]] - t(dL[[k]])%*%sti
  pvec[[k]]<-as.vector(psimat[lower.tri(psimat)])
  mvec[[k]]<-as.vector(M[[k]][lower.tri(M[[k]])])
  dcon[[k]] <-(t(pvec[[k]])%*%mvec[[k]])
  }

# diff C wrt psi[[l]]
P<-list()
l<-0
```

```
for (s in 1:(t-1))
 {for (r in (s+1):t)
  {l<-l+1
  pvec[[l]]<-mvec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  P[[l]]<-er%*%t(es)
  M[[l]]<-sti%*%lamt - t(lamt)%*%sti
  pvec[[l]]<-as.vector(P[[l]][lower.tri(P[[l]])])
  mvec[[l]]<-as.vector(M[[l]][lower.tri(M[[l]])])
  dconpsi[[l]]<-t(pvec[[l]])%*%mvec[[l]]
  # match up constraints with psis in correct order
  c[[l]]<-dconpsi[[l]]
  }
}



 # now need to adjust AI matrix with second derivatives where relevant
# use matrix O to add extra bits to AI matrix (AI in order sigma2,
# gammas, phi[k] , phi_col, psi[l])
# tp= total number of parameters - O is tpxtp

# second differentials wrt psi_rs and di
O<-matrix(0,tp,tp)
evec<-mvec<-list()
E<-M<-list()
l<-(tp-length(psi))
for (s in 1:(t-1))
{ for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for ( k in (1+qran+1):(1+qran+t))
  {q<- k-(1+qran)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]]<-dSin[[q]]%*%lamt - t(lamt)%*%dSin[[q]]
   evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
   mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
   }
}
}
# second differentials wrt psi_rs and ci
O<-matrix(0,tp,tp)
evec<-mvec<-list()
E<-M<-list()
l<-(tp-length(psi))
for (s in 1:(t-1))
{ for (r in (s+1):t)
```

```
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for ( k in (1+qran+t+1):(1+qran+2*t-1))
  {q<- k-(1+qran)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]]<-dSin[[q]]%*%lamt - t(lamt)%*%dSin[[q]]
   evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
   mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
   }
}
}


#stop (print (mvec))
#

# second differentials wrt psi_rs and lambda_uv
l<-(tp-length(psi))
for (s in 1:(t-1))
{ for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for (k in (1+qran+sigphi+1):(1+qran+totphi))
   {q<- k-(1+qran)
      mvec[[k]]<-matrix(0,t*(t-1)/2,1)
      M[[k]]<-sti%*%dL[[q]] - t(dL[[q]])%*%sti

     evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
    }
    }
 }

# second differential wrt di and lambda_uv
for (l in (1+qran+1):(1+qran+t))
  { p<-l-(1+qran)
  pvec[[l]]<-matrix(0,t*(t-1)/2,1)
   for (k in (1+qran+sigphi+1):(1+qran+totphi))
    {q<-k - (1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-dSin[[p]]%*%dL[[q]] - t(dL[[q]])%*%dSin[[p]]
     pvec[[l]] <-as.vector(psimat[lower.tri(psimat)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
```

```
      O[l,k]<- -(t(pvec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
     }
    }
# second differential wrt ci and lambda_uv
for (l in (1+qran+t+1):(1+qran+2*t-1))
  { p<-l-(1+qran)
  pvec[[l]]<-matrix(0,t*(t-1)/2,1)
   for (k in (1+qran+sigphi+1):(1+qran+totphi))
    {q<-k - (1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-dSin[[p]]%*%dL[[q]] - t(dL[[q]])%*%dSin[[p]]
     pvec[[l]] <-as.vector(psimat[lower.tri(psimat)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(pvec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
     }
    }

# second differential wrt ci and di
for (l in (1+qran+1):(1+qran+t))
  { p<-l-(1+qran)
  pvec[[l]]<-matrix(0,t*(t-1)/2,1)
   for (k in (1+qran+t+1):(1+qran+2*t-1))
    {q<-k - (1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-(t(dC[[q]])%*%dD[[p]]%*%C +
      t(C)%*%dD[[p]]%*%dC[[q]])%*%lamt-t(lamt)%*%(t(dC[[q]])%*%dD[[p]]%*%C +
      t(C)%*%dD[[p]]%*%dC[[q]])
     pvec[[l]] <-as.vector(psimat[lower.tri(psimat)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(pvec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
     }
    }
list(phi=phi,constraint=c,dcons=dcon,dconspsi=dconpsi,psi=psi,Omat=O)
}
#
# end antebandcons
#
#############################################################################
#
# USI
#############################################################################
USI<-function(r,t,phi=c(rep(0.1,(t*(t+1)/2+t^2))))
{
# forms the inverse sig and derivatives of sigma for the MVAR1 model with
# general Sigma = US (symmetric) with Lambda = diag(lambda)
# r=rows t=times phi = sigma_11 sigma_12, sigma_13, ...sigma_1t, sigma_22, sigma_23,
... lambda
# same as separable case USxar1
# Mar2010
#JDF
#############################################################################
F1<-mydiag(1,1,r-1,r-1)+mydiag(1,-1,r-1,r-1)
```

```
E1<- diag(c(0,rep(1,r-2),0))
Ir<-diag(r)
It<-diag(t)
Irt<-diag(r*t)
Jt<-matrix(1,nrow=t,ncol=t)
sigphi<-t*(t+1)/2
Sigma<-matrix(0,t,t)
ei<-ej<-matrix(0,t,1)
dS<-dSo<-list()
k<-0
for (i in 1:t)
{for (j in i:t)
{k<-k+1
Sigma[i,j]<- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dSo[[k]]<-ei%*%t(ej)
dS[[k]]<- dSo[[k]] + t(dSo[[k]]) - diag(diag(dSo[[k]]))
}
}
sigt<-Sigma+t(Sigma)-diag(diag(Sigma))
lamt<-phi[sigphi+1]*It

sti<-ginv.new(sigt)$invx
s0<-sti%*%ginv.new(It-(lamt%*%lamt))$invx
s1<-kronecker(Ir,s0)
s2<-kronecker(F1,lamt)
s3<-kronecker(E1,(lamt%*%lamt))
sigi<-s1%*%(Irt-s2+s3)
sig<-ginv.new(sigi)$invx
P1<-s1
P2<-Irt-s2+s3
# diff wrt phi[k]
dP1<-list()
dP2<-list()
for (k in 1:sigphi)
{dP1[[k]]<- kronecker(Ir, -sti%*%(dS[[k]])%*%s0)
dP2[[k]]<-0*Irt
}
for (k in (sigphi+1):(sigphi+1))
{dP1[[k]]<- kronecker(Ir, s0%*%(2*lamt)%*%ginv.new(It-(lamt%*%lamt))$invx )
dP2[[k]]<-kronecker(-F1,It) + kronecker(E1,(2*lamt))
}
dsigi<-list()
dsig<-list()
for (k in 1:(sigphi+1))
{dsigi[[k]]<-dP1[[k]]%*%P2 + P1%*%dP2[[k]]
dsig[[k]]<- -sig%*%dsigi[[k]]%*%sig
}
list(sig=sig,sigi=sigi,dsig=dsig,phi=phi,partype=rep(1,length(phi)))
}
#
# end USI
##############################################################################
```

E R Code for implementing 2dMVAR MCAR models in Chapter 9

```
##############################################################################
#
# MCARUSgen
##############################################################################
MCARUSgen<-function(r,c,t,phi=c(rep(0.1,(t*(t+1)/2+2*t^2))), psi=NULL)
{
# forms the inverse sig and derivatives of sigma for the 2dMCAR model
# with general Lambda_r= US, Lambda_c=US (t times)and Sigma = US covariance matrix
# (symmetric)
# with constraints Lambda_rSigma, Lambda_cSigma, Lambda_rLambda_cSigma all symmetric)
# r=rows c=cols t=times
#  phi = sigma_11 sigma_12,  ...sigma_1t, sigma_22,
#, ... lambda_r11, lambda_r12,...lambda_r1t,lambda_r21, lambda_r22,... lambda_r2t,...
#, ... lambda_c11, lambda_c12,...lambda_c1t,lambda_c21, lambda_c22,... lambda_c2t,...
# psi = psi_r psi_c psi_rc?????
# Sept 2011
#JDF
##############################################################################
F1r<-mydiag(1,1,r-1,r-1)+mydiag(1,-1,r-1,r-1)
F1c<-mydiag(1,1,c-1,c-1)+mydiag(1,-1,c-1,c-1)
E1r<- diag(c(0,rep(1,r-2),0))
E1c<- diag(c(0,rep(1,c-2),0))
Ir<-diag(r)
Ic<-diag(c)
It<-diag(t)
Irt<-diag(r*t)
Irct<-diag(r*c*t)
Dc <- matrix(0, nrow=c, ncol=c)
Dc[1,2] <-1
Dc[c,c-1]<-1
Dr <- matrix(0, nrow=r, ncol=r)
Dr[1,2] <-1
Dr[r,r-1]<-1

Jt<-matrix(1,nrow=t,ncol=t)

sigphi<-t*(t+1)/2
lamrphi<-t*t
lamcphi<-t*t
totphi<-sigphi+lamrphi+lamcphi
#Sigma
Sigma<-matrix(0,t,t)
ei<-ej<-matrix(0,t,1)
dS<-dSo<-list()
k<-0
for (i in 1:t)
{for (j in i:t)
{k<-k+1
Sigma[i,j]<- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
```

```
dSo[[k]]<-ei%*%t(ej)
dS[[k]]<- dSo[[k]] + t(dSo[[k]]) - diag(diag(dSo[[k]]))
}
}
sigt<-Sigma+t(Sigma)-diag(diag(Sigma))
# Lambda_r
lamrt<-matrix(0,t,t)
dLr<-list()
k<-(sigphi)
for (i in 1:t)
{ for (j in 1:t)
{k <- k+1
lamrt[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dLr[[k]]<-ei%*%t(ej)
}
}
# Lambda_c
lamct<-matrix(0,t,t)
dLc<-list()
k<-(sigphi+lamrphi)
for (i in 1:t)
{ for (j in 1:t)
{k <- k+1
lamct[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dLc[[k]]<-ei%*%t(ej)
}
}
sti<-ginv.new(sigt)$invx
lamrc<- lamrt%*%lamct
s0r<-ginv.new(It-(lamrt%*%lamrt))$invx
s0c<-ginv.new(It-(lamct%*%lamct))$invx
s1r<-ginv.new(It+(lamrt%*%lamrt))$invx
s1c<-ginv.new(It+(lamct%*%lamct))$invx
gam3in<- sti%*%s0r%*%s0c
M11 <- kronecker(Ir, gam3in)
M12 <- Irt + kronecker(E1r,(lamrt%*%lamrt))
M1<- M11%*%M12
k1<-kronecker(Ir,lamct%*%lamct)
k2<-Irct + kronecker(E1c,k1)
gamin <- (kronecker(Ic, M1))%*%k2
C1<- kronecker((F1r-Dr), (lamrt%*%s1r)) + kronecker(Dr, lamrt)
C2<- (kronecker(Ir, lamct))%*%(Irt-C1)
C3<- C2%*%(kronecker(Ir, s1c))
BigLam<-kronecker(Ic, C1) + kronecker((F1c -Dc),C3) + kronecker(Dc, C2)
# full inverse covariance matrix
Rinv <-gamin%*%(Irct - BigLam)
sig<-ginv.new(Rinv)$invx
#
#Rinv = R1R2
```

```
R1<-gamin
R2<-(Irct-BigLam)
# diff wrt phi[k]
dR1<-list()
dR2<-list()
dM1<-dM11<-dM12<-dC1<-dC2<-dC3<-list()
for (k in 1:sigphi)
{dM1[[k]]<- kronecker(Ir, (-sti%*%(dS[[k]])%*%gam3in))%*%M12
dR1[[k]]<- kronecker(Ic, dM1[[k]])%*%k2
dR2[[k]]<-0*Irct
}
# der wrt lamr
for (k in (1+sigphi):(sigphi+lamrphi))
{dM11[[k]]<-kronecker(Ir, (sti%*%s0r%*%(lamrt%*%dLr[[k]]+
dLr[[k]]%*%lamrt)%*%s0r%*%s0c))
dM12[[k]]<-kronecker(E1r,(lamrt%*%dLr[[k]] + dLr[[k]]%*%lamrt))
dM1[[k]]<-dM11[[k]]%*%M12 + M11%*%dM12[[k]]
dR1[[k]]<-kronecker(Ic, dM1[[k]])%*%k2

dC1[[k]]<-kronecker((F1r-Dr), (dLr[[k]]%*%s1r - lamrt%*%s1r%*%(lamrt%*%dLr[[k]]+
 dLr[[k]]%*%lamrt)%*%s1r))
 + kronecker(Dr, dLr[[k]])
dC2[[k]]<- kronecker(-Ir,lamct)%*%dC1[[k]]
dC3[[k]]<-dC2[[k]]%*%(kronecker(Ir,s1c))
dR2[[k]]<-kronecker(-Ic,dC1[[k]]) -kronecker((F1c-Dc), dC3[[k]]) -
kronecker(Dc, dC2[[k]])
}
# der wrt lamc
for (k in (1+sigphi+lamrphi):totphi)
{dM11[[k]]<-kronecker(Ir, (sti%*%s0r%*%(s0c%*%(lamct%*%dLc[[k]]+
dLc[[k]]%*%lamct)%*%s0c)))
dM1[[k]]<-dM11[[k]]%*%M12
dR1[[k]]<-kronecker(Ic, dM1[[k]])%*%k2 +
 (kronecker(Ic,M1)%*%kronecker(E1c,(kronecker(Ir,(lamct%*%dLc[[k]]+
 dLc[[k]]%*%lamct)))))

dC2[[k]]<- kronecker(Ir,dLc[[k]])%*%(Irt-C1)
dC3[[k]]<-dC2[[k]]%*%(kronecker(Ir,s1c)) +
C2%*%(kronecker(Ir, (-s1c%*%(lamct%*%dLc[[k]]+dLc[[k]]%*%lamct)%*%s1c)))
dR2[[k]]<- -kronecker((F1c-Dc), dC3[[k]]) - kronecker(Dc, dC2[[k]])
}

dsigi<-list()
dsig<-list()
for (k in 1:totphi)
{dsigi[[k]]<-dR1[[k]]%*%R2 + R1%*%dR2[[k]]
dsig[[k]]<- -sig%*%dsigi[[k]]%*%sig
}
list(sig=sig,sigi=Rinv,dsig=dsig,phi=phi,partype=rep(1,length(phi)))
}
#
# end MCARUSgen
############################################################
#
# constraints and derivatives of constraints
```

```
################################################
MCARdconsgen<-function(t,tp,qran,rowmod)
{
phi<-rowmod$phi
psi<-rowmod$psi
sigphi<-t*(t+1)/2
lamrphi<-t*t
lamcphi<-t*t
totphi<-sigphi+lamrphi+lamcphi
# function to get constraints and differentials of constraints
#Sigma
Sigma<-matrix(0,t,t)
ei<-ej<-matrix(0,t,1)
dS<-dSo<-list()
k<-0
for (i in 1:t)
{for (j in i:t)
{k<-k+1
Sigma[i,j]<- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dSo[[k]]<-ei%*%t(ej)
dS[[k]]<- dSo[[k]] + t(dSo[[k]]) - diag(diag(dSo[[k]]))
}
}
sigt<-Sigma+t(Sigma)-diag(diag(Sigma))
# Lambda_r
lamrt<-matrix(0,t,t)
dLr<-list()
k<-(sigphi)
for (i in 1:t)
{ for (j in 1:t)
{k <- k+1
lamrt[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dLr[[k]]<-ei%*%t(ej)
}
}
# Lambda_c
lamct<-matrix(0,t,t)
dLc<-list()
k<-(sigphi+lamrphi)
for (i in 1:t)
{ for (j in 1:t)
{k <- k+1
lamct[i,j] <- phi[k]
ei<-ej<-matrix(0,t,1)
ei[i,1]<-1
ej[j,1]<-1
dLc[[k]]<-ei%*%t(ej)
}
}
```

```
sti<-ginv.new(sigt)$invx
lamrc<- lamrt%*%lamct

#stop(print (lamrc))
# now for constraints
dconsr<-matrix(0,t,t)
dconsc<-matrix(0,t,t)
dconsrc<-matrix(0,t,t)
psimatr<-matrix(0,t,t)
psimatc<-matrix(0,t,t)
psimatrc<-matrix(0,t,t)
dcon<-dconr<-dconc<-dconrc<-list()
dconpsir<-dconpsic<-dconpsirc<-list()
c<-cr<-cc<-crc<-list()

l<-0
# allocate psi[k] across top half of psimatr (then make as a symmetric matrix)
for (i in 1:(t-1))
{ for (j in (i+1):t)
  { l<-l+1
    psimatr[i,j]<-psi[l]
  }
}
# now create full matrices
psimatrf<-psimatr+t(psimatr)-diag(diag(psimatr))
psimatr<-psimatrf
# allocate psi[k] across top half of psimatc (then make as a symmetric matrix)
l<-(t*(t-1)/2)
for (i in 1:(t-1))
{ for (j in (i+1):t)
  { l<-l+1
    psimatc[i,j]<-psi[l]
  }
}
# now create full matrices
psimatcf<-psimatc+t(psimatc)-diag(diag(psimatc))
psimatc<-psimatcf

# allocate psi[k] across top half of psimatrc (then make as a symmetric matrix)
l<-2*(t*(t-1)/2)
for (i in 1:(t-1))
{ for (j in (i+1):t)
  { l<-l+1
    psimatrc[i,j]<-psi[l]
  }
}
# now create full matrices
psimatrcf<-psimatrc+t(psimatrc)-diag(diag(psimatrc))
psimatrc<-psimatrcf

# up to here
# now differentiate the constraints wrt phi[k] - from Sigma  - sigma_rs
# dC=dCr+dCc+dCrc
pvecr<-mvecr<-pvecc<-mvecc<-pvecrc<-mvecrc<-list()
Mr<-Mc<-Mrc<-list()
```

```
 for (k in 1:sigphi)
 {
   pvecr[[k]]<-matrix(0,t*(t+1)/2,1)
   mvecr[[k]]<-matrix(0,t*(t+1)/2,1)
   Mr[[k]] <- lamrt%*%dS[[k]] - dS[[k]]%*%t(lamrt)
   pvecr[[k]]<-as.vector(psimatr[lower.tri(psimatr)])
   mvecr[[k]]<-as.vector(Mr[[k]][lower.tri(Mr[[k]])])
   dconr[[k]] <-(t(pvecr[[k]])%*%mvecr[[k]])

   pvecc[[k]]<-matrix(0,t*(t+1)/2,1)
   mvecc[[k]]<-matrix(0,t*(t+1)/2,1)
   Mc[[k]] <- lamct%*%dS[[k]] - dS[[k]]%*%t(lamct)
   pvecc[[k]]<-as.vector(psimatc[lower.tri(psimatc)])
   mvecc[[k]]<-as.vector(Mc[[k]][lower.tri(Mc[[k]])])
   dconc[[k]] <-(t(pvecc[[k]])%*%mvecc[[k]])

   pvecrc[[k]]<-matrix(0,t*(t+1)/2,1)
   mvecrc[[k]]<-matrix(0,t*(t+1)/2,1)
   Mrc[[k]] <- lamrt%*%lamct%*%dS[[k]] - dS[[k]]%*%t(lamct)%*%t(lamrt)
   pvecrc[[k]]<-as.vector(psimatrc[lower.tri(psimatrc)])
   mvecrc[[k]]<-as.vector(Mrc[[k]][lower.tri(Mrc[[k]])])
   dconrc[[k]] <-(t(pvecrc[[k]])%*%mvecrc[[k]])

   dcon[[k]]<-dconc[[k]]+dconr[[k]]+dconrc[[k]]
  }

# now differentiate C wrt phi[k] - from Lambdar - lambdar_ij
# note dCc wrt lambdar = 0
for (k in (sigphi+1):(sigphi+lamrphi))
{
  pvecr[[k]]<-matrix(0,t*(t+1)/2,1)
  mvecr[[k]]<-matrix(0,t*(t+1)/2,1)
   Mr[[k]] <- dLr[[k]]%*%sigt - sigt%*%t(dLr[[k]])
  pvecr[[k]]<-as.vector(psimatr[lower.tri(psimatr)])
  mvecr[[k]]<-as.vector(Mr[[k]][lower.tri(Mr[[k]])])
  dconr[[k]] <-(t(pvecr[[k]])%*%mvecr[[k]])

  pvecrc[[k]]<-matrix(0,t*(t+1)/2,1)
  mvecrc[[k]]<-matrix(0,t*(t+1)/2,1)
   Mrc[[k]] <- dLr[[k]]%*%lamct%*%sigt - sigt%*%t(lamct)%*%t(dLr[[k]])
  pvecrc[[k]]<-as.vector(psimatrc[lower.tri(psimatrc)])
  mvecrc[[k]]<-as.vector(Mrc[[k]][lower.tri(Mrc[[k]])])
  dconrc[[k]] <-(t(pvecrc[[k]])%*%mvecrc[[k]])

  dcon[[k]]<-dconr[[k]]+dconrc[[k]]
  }
 # now differentiate C wrt phi[k] - from Lambdac - lambdar_ij
# note dCr wrt lambdac = 0
for (k in (sigphi+lamrphi+1):totphi)
{

  pvecc[[k]]<-matrix(0,t*(t+1)/2,1)
  mvecc[[k]]<-matrix(0,t*(t+1)/2,1)
   Mc[[k]] <- dLc[[k]]%*%sigt - sigt%*%t(dLc[[k]])
  pvecc[[k]]<-as.vector(psimatc[lower.tri(psimatc)])
```

```
  mvecc[[k]]<-as.vector(Mc[[k]][lower.tri(Mc[[k]])])
  dconc[[k]] <-(t(pvecc[[k]])%*%mvecc[[k]])

  pvecrc[[k]]<-matrix(0,t*(t+1)/2,1)
  mvecrc[[k]]<-matrix(0,t*(t+1)/2,1)
   Mrc[[k]] <- lamrt%*%dLc[[k]]%*%sigt - sigt%*%t(dLc[[k]])%*%t(lamrt)
  pvecrc[[k]]<-as.vector(psimatrc[lower.tri(psimatrc)])
  mvecrc[[k]]<-as.vector(Mrc[[k]][lower.tri(Mrc[[k]])])
  dconrc[[k]] <-(t(pvecrc[[k]])%*%mvecrc[[k]])

  dcon[[k]]<-dconc[[k]]+dconrc[[k]]
  }


# diff wrt psir
Pr<-list()
l<-0
for (s in 1:(t-1))
 {for (r in (s+1):t)
  {l<-l+1
  pvecr[[l]]<-mvecr[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  Pr[[l]]<-er%*%t(es)
  Mr[[l]]<-lamrt%*%sigt - sigt%*%t(lamrt)
  pvecr[[l]]<-as.vector(Pr[[l]][lower.tri(Pr[[l]])])
  mvecr[[l]]<-as.vector(Mr[[l]][lower.tri(Mr[[l]])])
  dconpsir[[l]]<-t(pvecr[[l]])%*%mvecr[[l]]
    # match up constraints with psis in correct order
  cr[[l]]<-dconpsir[[l]]
  }
}
  # diff wrt psic
Pc<-list()
l<-0
for (s in 1:(t-1))
 {for (r in (s+1):t)
  {l<-l+1
  pvecc[[l]]<-mvecc[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  Pc[[l]]<-er%*%t(es)
  Mc[[l]]<-lamct%*%sigt - sigt%*%t(lamct)
  pvecc[[l]]<-as.vector(Pc[[l]][lower.tri(Pc[[l]])])
  mvecc[[l]]<-as.vector(Mc[[l]][lower.tri(Mc[[l]])])
  dconpsic[[l]]<-t(pvecc[[l]])%*%mvecc[[l]]
  # match up constraints with psis in correct order
  cc[[l]]<-dconpsic[[l]]
  }
}
   # diff wrt psirc
Prc<-list()
l<-0
```

```
for (s in 1:(t-1))
 {for (r in (s+1):t)
  {l<-l+1
  pvecrc[[l]]<-mvecrc[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  Prc[[l]]<-er%*%t(es)
  Mrc[[l]]<-lamrt%*%lamct%*%sigt - sigt%*%t(lamct)%*%t(lamrt)
  pvecrc[[l]]<-as.vector(Prc[[l]][lower.tri(Prc[[l]])])
  mvecrc[[l]]<-as.vector(Mrc[[l]][lower.tri(Mrc[[l]])])
  dconpsirc[[l]]<-t(pvecrc[[l]])%*%mvecrc[[l]]
  # match up constraints with psis in correct order
  crc[[l]]<-dconpsirc[[l]]
  }
}



# now need to adjust AI matrix with second derivatives where relevant
# use matrix O to add extra bits to AI matrix
#(AI in order sigma2, gammas, phir[k], phic[k], phi_col, psir[l], psic[l],psirc[l])
# tp= total number of parameters - O is tpxtp

lpsir<-(t*(t-1)/2)
lpsic<-(t*(t-1)/2)
lpsirc<-(t*(t-1)/2)
# second differentials of constraints wrt psir_rs and sigma_uv
O<-matrix(0,tp,tp)
evec<-mvec<-list()
E<-M<-list()
l<-(tp-(lpsir +lpsic +lpsirc))
for (s in 1:(t-1))
{ for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for ( k in (1+qran+1):(1+qran+sigphi))
  {q<- k-(1+qran)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]]<-lamrt%*%dS[[q]] - dS[[q]]%*%t(lamrt)
   evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
   mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
   }
}
}
# second differentials of constraints wrt psic_rs and sigma_uv
l<-(tp-(lpsic+lpsirc))
for (s in 1:(t-1))
{ for (r in (s+1):t)
```

```
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for ( k in (1+qran+1):(1+qran+sigphi))
  {q<- k-(1+qran)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]]<-lamct%*%dS[[q]] - dS[[q]]%*%t(lamct)
   evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
   mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
   }
}
}

# second differentials of constraints wrt psirc_rs and sigma_uv
l<-(tp-(lpsirc))
for (s in 1:(t-1))
{ for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for ( k in (1+qran+1):(1+qran+sigphi))
  {q<- k-(1+qran)
   mvec[[k]]<-matrix(0,t*(t-1)/2,1)
   M[[k]]<-lamrt%*%lamct%*%dS[[q]] - dS[[q]]%*%t(lamct)%*%t(lamrt)
   evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
   mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
   }
}
}




#stop (print (mvec))
#

# second differentials wrt psir_rs and lambdar_uv
l<-(tp-(lpsir + lpsic+lpsirc))
for (s in 1:(t-1))
{ for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
```

```
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for (k in (1+qran+sigphi+1):(1+qran+sigphi+lamrphi))
   {q<- k-(1+qran)
      mvec[[k]]<-matrix(0,t*(t-1)/2,1)
      M[[k]]<-dLr[[q]]%*%sigt - sigt%*%t(dLr[[q]])
     evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
    }
    }
 }
# second differentials wrt psic_rs and lambdac_uv
l<-(tp-(lpsic+lpsirc))
for (s in 1:(t-1))
{ for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for (k in (1+qran+sigphi+lamrphi+1):(1+qran+totphi))
   {q<- k-(1+qran)
      mvec[[k]]<-matrix(0,t*(t-1)/2,1)
      M[[k]]<-dLc[[q]]%*%sigt - sigt%*%t(dLc[[q]])
     evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
    }
    }
 }

 # second differentials wrt psirc_rs and lambdar_uv
l<-(tp-(lpsirc))
for (s in 1:(t-1))
{ for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for (k in (1+qran+sigphi+1):(1+qran+sigphi+lamrphi))
   {q<- k-(1+qran)
      mvec[[k]]<-matrix(0,t*(t-1)/2,1)
      M[[k]]<-dLr[[q]]%*%lamct%*%sigt - sigt%*%t(lamct)%*%t(dLr[[q]])
     evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
    }
    }
```

```
 }
 # second differentials wrt psirc_rs and lambdac_uv
l<-(tp-(lpsirc))
for (s in 1:(t-1))
{ for (r in (s+1):t)
  {l<-l+1
  evec[[l]]<-matrix(0,t*(t-1)/2,1)
  er<-es<-matrix(0,t,1)
  er[r,1]<-1
  es[s,1]<-1
  E[[l]]<-er%*%t(es)
   for (k in (1+qran+sigphi+lamrphi+1):(1+qran+totphi))
   {q<- k-(1+qran)
      mvec[[k]]<-matrix(0,t*(t-1)/2,1)
      M[[k]]<-lamrt%*%dLc[[q]]%*%sigt - sigt%*%t(dLc[[q]])%*%t(lamrt)
     evec[[l]] <-as.vector(E[[l]][lower.tri(E[[l]])])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])
      O[l,k]<- -(t(evec[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
    }
    }
 }

# second differential wrt sigma_rs and lambdar_uv
Mr<-Mc<-Mrc<-list()
for (l in (1+qran+1):(1+qran+sigphi))
  { p<-l-(1+qran)
  pvecr[[l]]<-matrix(0,t*(t-1)/2,1)
  pvecrc[[l]]<-matrix(0,t*(t-1)/2,1)
   for (k in (1+qran+sigphi+1):(1+qran+sigphi+lamrphi))
    {q<-k - (1+qran)
     mvecr[[k]]<-matrix(0,t*(t-1)/2,1)
     Mr[[k]]<-dLr[[q]]%*%dS[[p]] - dS[[p]]%*%t(dLr[[q]])
     pvecr[[l]] <-as.vector(psimatr[lower.tri(psimatr)])
     mvecr[[k]] <-as.vector(Mr[[k]][lower.tri(Mr[[k]])])

     mvecrc[[k]]<-matrix(0,t*(t-1)/2,1)
     Mrc[[k]]<-dLr[[q]]%*%lamct%*%dS[[p]] - dS[[p]]%*%t(lamct)%*%t(dLr[[q]])
     pvecrc[[l]] <-as.vector(psimatrc[lower.tri(psimatrc)])
     mvecrc[[k]] <-as.vector(Mrc[[k]][lower.tri(Mrc[[k]])])
   #check all of this?
      O[l,k]<- -(t(pvecr[[l]])%*%mvecr[[k]])-(t(pvecrc[[l]])%*%mvecrc[[k]])
      O[k,l]<-O[l,k]
     }
    }
# second differential wrt sigma_rs and lambdac_uv
for (l in (1+qran+1):(1+qran+sigphi))
  { p<-l-(1+qran)
  pvecc[[l]]<-matrix(0,t*(t-1)/2,1)
  pvecrc[[l]]<-matrix(0,t*(t-1)/2,1)
   for (k in (1+qran+sigphi+lamrphi+1):(1+qran+totphi))
    {q<-k - (1+qran)
     mvecc[[k]]<-matrix(0,t*(t-1)/2,1)
     Mc[[k]]<-dLc[[q]]%*%dS[[p]] - dS[[p]]%*%t(dLc[[q]])
     pvecc[[l]] <-as.vector(psimatc[lower.tri(psimatc)])
```

```
     mvecc[[k]] <-as.vector(Mc[[k]][lower.tri(Mc[[k]])])

     mvecrc[[k]]<-matrix(0,t*(t-1)/2,1)
     Mrc[[k]]<-lamrt%*%dLc[[q]]%*%dS[[p]] - dS[[p]]%*%t(dLc[[q]])%*%t(lamrt)
     pvecrc[[l]] <-as.vector(psimatrc[lower.tri(psimatrc)])
     mvecrc[[k]] <-as.vector(Mrc[[k]][lower.tri(Mrc[[k]])])
     O[l,k]<- -(t(pvecc[[l]])%*%mvecc[[k]]) -(t(pvecrc[[l]])%*%mvecrc[[k]])
     O[k,l]<-O[l,k]
     }
    }



# second differential wrt lambdar_rs and lambdac_uv
for (l in (1+qran+sigphi+1):(1+qran+sigphi+lamrphi))
  { p<-l-(1+qran)
  pvecrc[[l]]<-matrix(0,t*(t-1)/2,1)
     for (k in (1+qran+sigphi+lamrphi+1):(1+qran+totphi))
    {q<-k - (1+qran)
     mvec[[k]]<-matrix(0,t*(t-1)/2,1)
     M[[k]]<-dLr[[p]]%*%dLc[[q]]%*%sigt - sigt%*%t(dLc[[q]])%*%t(dLr[[p]])
     pvecrc[[l]] <-as.vector(psimatrc[lower.tri(psimatrc)])
     mvec[[k]] <-as.vector(M[[k]][lower.tri(M[[k]])])

      O[l,k]<- -(t(pvecrc[[l]])%*%mvec[[k]])
      O[k,l]<-O[l,k]
     }
    }
    # need to joing all elements of lists cr cc crc into list c
    c<-list(cr,cc,crc)

#check all done
list(phi=phi,constraint=c,dcons=dcon,
 #dconspsir=dconpsir, dconspsic=dconpsic,dconspsirc=dconpsirc,
 psi=psi,Omat=O)
}
#
#end MCARdconsgen
#
################################################################################
```

Bibliography

Assuncao, R. & Krainski, E. (2009). Neighborhood dependence in Bayesian spatial models. *Biometrical Journal* **51**, 851–869.

Banerjee, S., Carlin, B., & Gelfand, A. (2004). *Hierarchical Modelling and Analysis for Spatial Data*. Boca Raton, Florida: Chapman and Hall/CRC Press.

Bartlett, M. (1978). Nearest neighbour models in the analysis of field experiments (with discussion). *Journal of the Royal Statistical Society, Series B* **40**, 147–174.

Beeck, C., Cowling, W., Smith, A., & Cullis, B. (2010). Analysis of yield and oil from a series of canola breeding trials. part 1: Fitting factor analytic models with pedigree information. *Genome* **53**, 992–1001.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with disucussion). *Journal of the Royal Statistical Society, Series B* **35**, 192–236.

Besag, J. & Kempton, R. A. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics* **42**, 231–251.

Billheimer, D., Cardoso, T., Freeman, E., Guttorp, P., Ko, H., & Ilkey, M. (1997). Natural variability of benthic species composition in the Delaware Bay. *Environmental and Ecological Statistics* **4**, 95–115.

Bjornsson, H. (1978). Analysis of a series of long-term grassland experiments with autocorrelated errors. *Biometrics* **34**, 645–651.

Box, G. & Jenkins, G. (1970). *Time Series Analysis forecasting and control*. Holden-Day.

Brien, C. J. & Bailey, R. A. (2006). Multiple randomizations (with discussion). *Journal of the Royal Statistical Society, Series B* **68**, 571–609.

Busbice, T. (1969). Inbreeding in synthetic varieties. *Crop Science* **9**, 601–604.

Butler, D., Cullis, B., Gilmour, A., & Gogel, B. (2009). *ASReml-R, Reference Manual*. Release 3. Queensland Department of Primary Industries and Fisheries, Toowoomba.

Carlin, B. & Banerjee, S. (2003). Hierachical multivariate CAR models for spatio temporally correlated survival data. In Bernardo, J., editor, *Bayesian Statistics*, volume 7, pages 45–63. Oxford University Press.

Costa e Silva, J., Dutkowski, G., & A.R, G. (2001). Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. *Canadian Journal of Forest Research* **31**, 1887–1893.

Cressie, N. (1993). *Statistics for Spatial Data*. New York: Wiley.

Crowder, M. & Hand, D. (1990). *Analysis of Repeated Measures*. CRC Press.

Cullis, B. & Gleeson, A. (1991). Spatial analysis of field experiments - an extension to two dimensions. *Biometrics* **47**, 1449–1460.

Cullis, B., Gogel, B., Verbyla, A., & Thompson, R. (1998). Spatial analysis of multi-environment early generation trials. *Biometrics* **54**, 1–18.

Cullis, B. & McGilchrist, C. (1990). A model for the analysis of growth data from designed experiments. *Biometrics* **46**, 131–142.

Cullis, B., Smith, A., Beeck, C., & Cowling, W. (2010). Analysis of yield and oil from a series of canola breeding trials. part 11: Exploring variety by environment interaction using factor analysis. *Genome* **53**, 1002–1016.

Cullis, B., Smith, A., & Coombes, N. (2006). On the design of early generation variety field trials with correlated data. *Journal of Agricultural, Biological and Environmental Statistics* **11**, 381–393.

Cullis, B., Thomson, F., Fisher, J., Gilmour, A., & Thompson, R. (1996a). The analysis of the NSW wheat variety database. I. Modelling trial error variance. *Theoretical and Applied Genetics* **92**, 21–27.

Cullis, B., Thomson, F., Fisher, J., Gilmour, A., & Thompson, R. (1996b). The analysis of the NSW wheat variety database. II. Variance component estimation. *Theoretical and Applied Genetics* **92**, 28–39.

Davik, J. & Honne, B. (2005). Genetic variance and breeding values for resistance to a wind-borne disease [*sphaerotheca macularis (wallr. ex fr.)*] in strawberry (*fragaria ananassa duch.*) estimated by exploring mixed and spatial models and pedigree information. *Theoretical and Applied Genetics* **111**, 256–264.

Diggle, P. (1988). An approach to the analysis of repeated measurements. *Biometrics* **44**, 959–971.

Diggle, P., Heagerty, P., Liang, K., & Zegar, S. (2002). *Analysis of Longitudinal Data.* Oxford.

Dutkowski, G., Costa e Silva, J., Gilmour, A., & Lopez, G. (2002). Spatial analysis methods for forest genetic trials. *Canadian Journal of Forest Research* **32**, 2201–2214.

Evans, J. & Roberts, E. (1979). Analysis of sequential observations with applications to experiments on grazing animals and perennial plants. *Biometrics* **35**, 687–693.

Falconer, D. & Mackay, T. (1996). *Introduction to Quantitative Genetics.* 4^{th} *edition.* Harlow, UK:Longman.

Finkenstadt, B., Isham, V., & Held, L. (2006). *Statistical Methods of Spatio Temporal Systems.* Chapman and Hall.

Finlay, K. & Wilkinson, G. (1963). The analysis of adaptation in a plant breeding programme. *Australian Journal of Agricultural Research* **14**, 742–754.

Fisher, R. (1935). *The design of experiments.* Edinburgh: Oliver and Boyd.

Frensham, A. B., Cullis, B., & Verbyla, A. (1997). Genotype by environment variance heterogeneity in a two stage analysis. *Biometrics* **53**, 1373–1383.

Gabriel, K. (1962). Ante-dependence analysis of an ordered set of variables. *The Annals of Mathematical Statistics* **33**, 201–212.

Galecki, A. (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Communications in Statistics - Theory and Methods* **23**, 3105–3119.

Gauch, H. (1992). *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*. Elsevier, Amsterdam.

Gelfand, A. & Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* **4**, 11–15.

Gilmour, A., Thompson, R., & Cullis, B. (1995). Average information REML, an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51**, 1440–1450.

Gilmour, A. R., Cullis, B. R., & Verbyla, A. P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics* **2**, 269–273.

Gleeson, A. & Cullis, B. (1987). Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. *Biometrics* **43**, 277–288.

Green, P., Jennison, C., & Scheult, A. (1985). Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society, Series B* **47**, 299–315.

Green, P. & Silverman, B. (1994). *Nonparametric Regression and Generalied Linear Models*. London: Chapman and Hall.

Haddad, J. (1998). A simple method for computing the covariance matrix and it's inverse of a stationary autoregressive process. *Communications in Statistics - Simulation and Computation* **27**, 617–623.

Hardner, C., Dieters, M., Dale, G., DeLacy, I., & Basford, K. (2010). Patterns of genotype-by-environment interaction in diameter at breast height at age 3 for eucalypt hybrid clones grown for reafforestation of lands affected by salinity. *Tree Genetics and Genomes* **6**, 833–851.

Harison, L., Penny, W., & Friston, K. (2003). Multivariate autoregressive modelling of fMRI time series. *Neuroimage* **19**, 1477–1491.

Haskard, K., Cullis, B., & Verbyla, A. (2007). Anisotropic Matérn correlation and spatial prediction using REML. *Journal of Agricultural, Biological, and Environmental Statistics* **12**, 147–160.

Henderson, C. (1950). Estimation of genetic parameters. *The Annals of Mathematical Statistics* **21**, 309–310.

Henderson, C. (1963). *Statistical Genetics and Plant Breeding*, chapter Selection index and expected gain, pages 141–163. National Academy of Sciences.

Hrafnkelsson, B. & Cressie, N. (2003). Hierarchical modelling of count data with application to nuclear fall-out. *Environmental and Ecological Statistics* **10**, 179–200.

Humphries, A. & Auricht, G. (2001). Breeding lucerne for Australia's southern dryland cropping environments. *Australian Journal of Agricultural Research* **52**, 153–169.

Hytti, H., Takalo, R., & Ihalainen, H. (2006). Tutorial on multivariate autoregressive modelling. *Journal of Clinical Monitoring and Computing* **20**, 101–108.

Irwin, J., Lloyd, D., & Lowe, K. (2001). Lucerne biology and genetic improvement - an analysis of past activities and future goals in Australia. *Australian Journal of Agricultural Research* **52**, 699–712.

Jaffrezic, F., Thompson, R., & Hill, W. (2003). Structured antedependence models for genetic analysis of repeated measures on multiple quantitative traits. *Genetics Research, Cambridge* **82**, 55–65.

Jamrozik, J. & Shaeffer, L. (1997). Estimates of genetic parameters for a test day model with random regression for yield traits of first lactation Holsteins. *Journal of Dairy Science* **80**, 762–770.

Jensen, J. (2001). Genetic evaluation of dairy cattle using test-day models. *Journal of Dairy Science* **84**, 2803–2812.

Jin, X., Carlin, B., & Banerjee, S. (2005). Generalized hierarachical multivariate CAR models for areal data. *Biometrics* **61**, 950–961.

Jones, R. (1993). *Longitudinal Data with Serial Correlation: A State Space Approach.* Chapman and Hall.

Jones, T., Cullis, B., Clingeleffer, P., & Ruhl, E. (2009). Effects of novel hybrid and traditional rootstocks on vigour and yield components of shiraz grapevines. *Australian Journal of Grape and Wine Research* **15**, 284–292.

Julier, B., Huyghe, C., & Ecalle, C. (2000). Within- and among-cultivar genetic variation in alfalfa: Forage quality, morphology and yield. *Crop Science* **40**, 365–369.

Kaufman, L. & Rousseeuw, P. (1990). *Finding groups in Data: An Introduction to Cluster Analysis.* Wiley, New York.

Kelly, A., Smith, A., Eccleston, J., & Cullis, B. (2007). The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. *Crop Science* **47**, 1063–1070.

Kenward, M. & Roger, J. (1997). The precision of fixed effects estimates from restricted maximum likelihood. *Biometrics* **53**, 983–997.

Knight, R. (1970). The measurement and interpretation of genotype-environment interactions. *Euphytica* **19**, 225–235.

Kyriakidis, P. & Journel, A. (1999). Geostatistical space-time models: A review. *Mathematical Geology* **31**, 651–684.

Laird, N. & Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

Lodge, G. (1985). Effects of grazing and hay cutting on the yield and persistence of dryland aphid-resistant lucerne cultivars at Tamworth, NSW. *Australian Journal of Experimental Agriculture and Animal Husbandry* **25**, 138–148.

Lodge, G. & Gleeson, A. (1984). A comparison of methods of estimating lucerne population for monitoring persistence. *Australian Journal of Experimental Agriculture and Animal Husbandry* **24**, 174–177.

Lowe, K., Gramshaw, D., Bowdler, T., & Ludke, D. (1985). Performance of North American and Australian lucernes in the Queensland subtropics 2. yield and plant survival in irrigated stands. *Australian Journal of Experimental Agriculture* **25**,

82–90.

Mardia, K. (1988). Multi-dimensional multivariate gaussian markov random fields with application to image processing. *Journal of Multivariate Analysis* **24**, 265–284.

Meyer, K. (1998). Estimating covariance functions for longitudinal data using a random regression model. *Genetics Selection Evolution* **30**, 221–240.

Meyer, K. (1999). Estimates of genetic and phenotypic covariance functions for post-weaning growth and mature weight of beef cows. *Journal of Animal Breeding and Genetics* **116**, 181–205.

Meyer, K. (2005). Advances in methodology for random regression analyses. *Australian Journal of Experimental Agriculture* **45**, 847–858.

Meyer, K. & Kirkpatrick, M. (2005). Up hill, down dale: quantitative genetics of curvacoeus traits. *Philosophical Transactions of The Royal Society B* **360**, 1443–1455.

Morley, F., Daday, H., & Peak, J. (1957). Quantative inheritance in lucerne, medicago sativa l. i. inheritance and selection for winter yield. *Australian Journal of Agricultural Research* **8**, 635–651.

Murison, B., Ayres, J., Lane, L., & Woodfield, D. (2006). Statistical methods to address spatial variation in pasture evaluation trials. *Proceedings of the 13th Australasian Plant Breeding Conference* pages 339–346.

Musial, J., Basford, K., & Irwin, J. (2002). Analysis of genetic diversity within Australian lucerne cultivars and implications for future genetic improvement. *Australian Journal of Experimental Agriculture* **53**, 629–636.

Naik, D. & Rao, S. (2001). Analysis of multivariate repeated measures data with a kronecker product structured covariance matrix. *Journal of Applied Statistics* **28**, 91–105.

Neyman, J. & Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1–32.

Núñez-Antón, V. & Zimmerman, D. L. (2000). Modeling nonstationary longitudinal data. *Biometrics* **56**, 699–705.

Oakey, H., Verbyla, A., Pitchford, W., Cullis, B., & Kuchel, H. (2006). Joint modelling of additive and non-additive genetic line effects in single field trials. *Theoretical and Applied Genetics* **113**, 809–819.

Pace, R., Barry, R., Gilley, O., & Sirmans, C. (2000). A method for spatio-temporal forecasting with application to real estate prices. *International Journal of Forecasting* **16**, 229–246.

Patterson, H. & Lowe, B. I. (1970). The errors of long-term experiments. *The Journal of Agricultural Science* **74**, 53–60.

Patterson, H. D. & Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **31**, 100–109.

Pettit, A., Weir, I., & Hart, A. (2002). A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. *Statistics and Computing* **12**, 353–367.

Piepho, H. (1997). Analyzing genotype-environment data by mixed models with multi-

plicative terms. *Biometrics* **53**, 761–767.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN.

Resende, M., Thompson, R., & Welham, S. (2006). Multivariate spatial statistical analysis of longitudinal data in perennial crops. *Brazilian Journal of Mathematics and Statistics* **24**, 147–169.

Robinson, G. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science* **6**, 15–51.

Sahu, S. (2005). *Workshop on Recent Advances in Modelling Spatio Temporal Data: A book of abstracts*. maths.soton.ac.uk/s3riwshop.

Sahu, S. & Mardia, K. (2005). Recent trends in modelling spatio temporal data. *Proceedings of the Italian Statistical Society Meeting, Messina, Italy Sep 2005* pages 69–83.

Sain, S. & Cressie, N. (2007). A spatial model for multivariate lattice data. *Journal of Econometrics* **140**, 226–259.

Schaeffer, L. (2004). Application of random regression models in animal breeding. *Livestock Production Science* **86**, 35–45.

Siddiqui, M. (1958). On the inversion of the sample covariance matrix in a stationary autoregressive process. *The Annals of Mathematical Statistics* **29**, 585–588.

Smith, A., Cullis, B., & Thompson, R. (2001). Analysing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* **57**, 1138–1147.

Smith, A. B., Cullis, B., & Thompson, R. (2005). The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *Journal of Agricultural Science, Cambridge* **143**, 1–14.

Smith, A. B., Stringer, J. K., Wei, X., & Cullis, B. R. (2007). Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. *Euphytica* **157**, 253–266.

Smith, K. & Casler, M. (2004). Spatial analysis of forage grass trials across locations, years, and harvests. *Crop Science* **44**, 56–62.

Smith, K. & Kearney, G. (2002). Improving the power of pasture cultivar trials to discriminate cultivars on the basis of differences in herbage yield. *Australian Journal of Agricultural Research* **53**, 191–199.

Stefanova, K., Smith, A., & Cullis, B. (2009). Enhanced diagnostics for the spatial analysis of field trials. *Journal of Agricultural, Biological, and Environmental Statistics* **14**, 392–410.

Strabel, T., Szyda, J., Ptak, E., & Jamrozik, J. (2005). Comparison of random regression test-day models for Polish black and white cattle. *Journal of Dairy Science* **88**, 3688–3699.

Stram, D. O. & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171–1177.

Stringer, J. (2006). *Joint Modelling of spatial variability and interplot competition to*

improve the efficiency of plant improvement. PhD thesis, The University of Queensland.

Stringer, J. & Cullis, B. (2002). Application of spatial analysis techniques to adjust for fertility trends and identify interplot competition in early stage sugarcane selection trials. *Australian Journal of Agricultural Research* **53**, 911–918.

Thompson, R. (2008). Estimation of quantitative genetic parameters. *Proc. Biol. Sci.* **275**, 679–686.

Thompson, R., Cullis, B., Smith, A., & Gilmour, A. (2003). A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. *Australian and New Zealand Journal of Statistics* **45**, 445–460.

Vaida, F. & Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika* **92**, 351–370.

Verbyla, A. (1985). A note on the inverse covariance matrix of the autoregressive process. *Australian Journal of Statistics* **2**, 221–224.

Verbyla, A. (1990). A conditional derivation of residual maximum likelihood. *Australian Journal of Statistics* **32**, 227–230.

Verbyla, A. & Cullis, B. (1990). Modelling in repeated measures experiments. *Journal of the Royal Statistical Society, Series C* **39**, 341–356.

Verbyla, A., Cullis, B., Kenward, M., & Welham, S. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). *Applied Statistics* **48**, 269–312.

Wall, M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference* **121**, 311–324.

Waller, L., Carlin, B., Xia, H., & Gelfand, A. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* **92**, 607–617.

White, I., Thompson, R., & Brotherstone, S. (1999). Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy Science* **82**, 632–638.

White, M., Cullis, B., Gilmour, A., & Thompson, R. (1998). Smoothing biological data with splines. *Invited Pap. 19th Int. Biometric Conf.* pages 57–66.

Wilkinson, G., Eckert, S., Hancock, T., & Mayo, O. (1983). Nearest neighbour (NN) analysis of field experiments (with discussion). *Journal of the Royal Statistical Society, Series B* **45**, 152–212.

Zimmerman, D. & Harville, D. (1991). A random field approach to the analysis of field plot experiments. *Biometrics* **47**, 223–239.