Geostatistical Modelling and Simulation of Karst Systems

A thesis by

Razieh Ghobadi

Faculty of Engineering, Computer and Mathematical Sciences

Submitted in partial fulfilment of the requirements for the degree of

Master of Philosophy

The University of Adelaide

June 2016
Contents

Contents .................................................................................................................. ii
List of Figures ........................................................................................................ iv
List of Tables ........................................................................................................... ix
List of Abbreviations ............................................................................................. x
Declaration of Originality .................................................................................. xiv
Acknowledgements ............................................................................................. xv

Chapter 1: Introduction ......................................................................................... 1
  1.1 Overview ........................................................................................................ 1
  1.2 Research Objective ...................................................................................... 5
  1.3 Thesis Outline ............................................................................................. 6

Chapter 2: Literature Review ............................................................................... 8
  2.1 Overview of Karst Aquifers ....................................................................... 8
  2.2 Geometry of Karst Networks ..................................................................... 9
  2.3 Existing Techniques for Modelling Meandering Channels and Karst Systems...10
    2.3.1 Traditional geostatistical techniques ............................................... 11
    2.3.2 Multiple-point geostatistics ............................................................... 16
  2.4 Current Approaches in Evaluating Geostatistical Methods ....................... 21
  2.5 Gap Statement ............................................................................................ 23
    2.5.1 Descriptive list of gaps .................................................................... 24

Chapter 3: Methodology ....................................................................................... 26
  3.1 Introduction to Geostatistical Methods ....................................................... 26
  3.2 Potentially Useful Methods ....................................................................... 27
    3.2.1 Two-point statistics .......................................................................... 28
    3.2.2 Multiple-point statistics ................................................................... 29
      3.2.2.1 Gridded MPS method ................................................................. 31
      3.2.2.2 Non-gridded MPS method ....................................................... 35
      3.2.2.3 Modified non-gridded MPS method ....................................... 38
  3.3 Evaluation of Modified Non-gridded MPS .............................................. 42
    3.3.1 Measuring variability among realisations ....................................... 43
    3.3.2 Measuring variability between training image and realisations ........... 45

Chapter 4: Results ............................................................................................... 46
  4.1 Gridded MPS ............................................................................................. 46
    4.1.1 Gridded MPS scanning algorithm ............................................... 46
    4.1.2 Gridded MPS simulation algorithm .............................................. 50
  4.2 Modified Non-gridded MPS .................................................................... 51
    4.2.1 Modified non-gridded MPS pre-processing ................................... 51
    4.2.2 Modified non-gridded MPS analysis .......................................... 55
    4.2.3 Modified non-gridded MPS simulation ....................................... 57
  4.3 Distance Measures ..................................................................................... 59
    4.3.1 The within-realisation variability ............................................... 60
    4.3.2 The between-realisation variability ............................................. 62
    4.3.3 Sensitivity of non-gridded template size ...................................... 64
  4.4 Case Study (Demonstration Examples) ................................................... 68
    4.4.1 Olwolgin Cave ................................................................................. 69
      4.4.1.1 Gridded MPS of Olwolgin Cave ............................................ 70
      4.4.1.2 Modified non-gridded MPS for Olwolgin Cave ................... 72
    4.4.2 Tank Cave ....................................................................................... 74
List of Figures

Figure 1.1. Schematic diagram of a heterogeneous karst aquifer (adapted from Goldscheider & Drew 2007) .................................................................2

Figure 1.2. Turbulent water flow in a cave (photo: Wenger)..............................2

Figure 1.3. Formation of sinkholes in Florida (photo: Yunji De Nies & Daisha Riley 2012) .................................................................................................3

Figure 1.4. Surface stream sinking underground via a sinkhole (photo: Goldscheider) .................................................................................................4

Figure 2.1. Schematic model of karst aquifer illustrating various zones (Geyer 2008) ........................................................................................................9

Figure 2.2. 3D finite element model generated by coupling the karst conduit simulator with flow simulation (after Jaquet et al. 2004) .......................12

Figure 2.3. 2D map of conduit network realisations; black lines: inlets, which are known as the starting points of the conduits; green points: outlet or karst spring (Borghi et al. 2012) ........................................................................13

Figure 2.4. (a) 3D map of the non-connected conduits; (b) 3D map of a realisation of a karst network with connectivity using a modified diffusion-limited aggregation technique (Pardo-Iguzquiza et al. 2011) ..........15

Figure 2.5. 3D map of karst network realisations generated by Henrion et al. (2008) ...16

Figure 2.6. Top view of pattern simulation of large-scale passages; (a) original data, (b) simulated network, circle: conditioning data, lines: simulated connection (Erzeybek Balan 2012) ..........................................................20

Figure 3.1. General geostatistical simulation workflow (Desbarats 1996) ...............27

Figure 3.2. A geological model using a training image (Honarkhah & Caers 2010).....30

Figure 3.3. Example of gridded training image (Honarkhah & Caers 2010) ..........31

Figure 3.4. Example of gridded template; a 3x3 template on an 11x11 grid (Arpat & Caers 2007) .........................................................................................31

Figure 3.5. (a) Sampling of a gridded training image using a gridded template, (b) observed patterns (after Arpat & Caers 2007) ........................................32

Figure 3.6. General algorithm of distance-based pattern modelling .....................34

Figure 3.7. Generic template (adapted from Erzeybek Balan 2012) .......................36
Figure 3.8. Example of a two-node template, a centre node (red circle) and two template nodes (blue circles) (adapted from Erzeybek Balan 2012) .......... 37

Figure 3.9. Left: non-gridded training image; middle and right: two realisations generated with non-gridded MPS by Erzeybek Balan (2012) using a C++ code ......................................................................................................................... 39

Figure 3.10. General algorithm of the modified non-gridded MPS .............................................. 41

Figure 3.11. Template used to scan the non-gridded training image by modified non-gridded MPS .................................................................................................................................................. 42

Figure 3.12. General framework of evaluation of the modified non-gridded MPS ...... 43

Figure 4.1. Conceptual model of a channel system 101×101×1 used as a gridded training image for the gridded MPS scanning analysis ....................................................... 47

Figure 4.2. Entropy plot for determining the optimum template size ............................................. 48

Figure 4.3. MDS dimension selection; upper: scree plot; lower: correlation coefficient vs dimension ................................................................................................................................................ 49

Figure 4.4. MDS result using the 15×15 pattern template .................................................................. 50

Figure 4.5. Two realisations using the gridded MPS simulation algorithm ............................... 51

Figure 4.6. (a) Conceptual model of a channel system with resolution 101×101×1, (b) digitised model used as a non-gridded training image for the modified non-gridded MPS scanning analysis (generated by Digitize it 2010) ........................................................................................................................................ 52

Figure 4.7. Digitised model used as training data for the modified non-gridded MPS scanning algorithm; arrow indicates north ................................................................. 53

Figure 4.8. Frequency histogram of the observed patterns as obtained by scanning the training image with the template ....................................................................................... 55

Figure 4.9. Two realisations using the modified non-gridded MPS simulation algorithm; black points denote simulated nodes ........................................................................... 57

Figure 4.10. Two realisations using the non-gridded MPS simulation algorithm; black points denote simulated nodes ...................................................................................... 58

Figure 4.11. Distances between training image and 50 realisations for gridded MPS; blue dots are the distance values between each realisation and the training image; red dots represent the average of the distances ................. 60

Figure 4.12. Distances between training image and 50 realisations for modified non-gridded MPS; blue dots are the distance values between each realisation and training image; red dots represent the average of the distances ............ 61
Figure 4.13. Distance plot between realisations and training image for gridded and modified non-gridded MPS methods; red dots represent the average of the distances .................................................................61

Figure 4.14. Distances between 50 realisations in gridded MPS; blue dots are the distance values between pairs of realisations; red dots represent the average of the distances.................................................................62

Figure 4.15. Distances between 50 realisations from modified non-gridded MPS; blue dots are the distance value between pairs of realisations; red dots represent the average of the distances ..................................................................63

Figure 4.16. Distance plot between realisations for gridded and modified non-gridded MPS methods; red dots represent the average of the distances .....63

Figure 4.17. Distance plot for the modified non-gridded MPS with a small template size (i.e., distances between training image and 50 realisations using a template with a smaller tolerance window and lag distance) .........................64

Figure 4.18. Distance plot for the modified non-gridded MPS with a small template size (i.e., distances between 50 realisations using a template with a smaller tolerance window and lag distance) ...........................................65

Figure 4.19. Distance plot for the modified non-gridded MPS with a large template (i.e., distances between training image and 50 realisations using a template with a larger tolerance window and lag distance) ..................65

Figure 4.20. Distance plot for the modified non-gridded MPS with a large template (i.e., distances between 50 realisations using a template with a larger tolerance window and lag distance) ...........................................66

Figure 4.21. Distance plot between realisations and training image for modified non-gridded MPS method using different template sizes .......................66

Figure 4.22. Distance plot between realisations for modified non-gridded MPS method using different template sizes ........................................67

Figure 4.23. Olwolgin Cave downstream section ...........................................68

Figure 4.24. Map of Tank Cave, the longest phreatic cave in the Mt Gambier area, with spatial analysis of passage directions (Harris 2008) .....................69

Figure 4.25. Olwolgin Cave grey-scale intensity image ....................................70

Figure 4.26. Olwolgin Cave digitised binary image at resolution of 170×100×1 (training image) ..............................................................................70
Figure 4.27. (a) Entropy plot; (b) profile log-likelihood; for optimum template size selection ............................................................... 71
Figure 4.28. Variance of entropy plot ................................................................................................................................. 71
Figure 4.29. Three realisations generated using gridded MPS simulation algorithm for Olwolgin Cave ................................................................. 72
Figure 4.30. Digitised model of Olwolgin Cave used as training data for the modified non-gridded MPS scanning algorithm; arrow indicates north ..... 73
Figure 4.31. Conditioning data used for the modified non-gridded MPS simulation; arrow indicates north .......................................................... 73
Figure 4.32. Two different realisations of Olwolgin Cave generated by modified non-gridded MPS ................................................................. 74
Figure 4.33. Tank Cave binary image at resolution of 100×100×1 (training image) ............................................................. 75
Figure 4.34. (a) Entropy plot; (b) profile log-likelihood; presenting the application of optimum template size selection ................................................................. 75
Figure 4.35. Variance of entropy plot ................................................................................................................................. 76
Figure 4.36. Three realisations resulted using gridded MPS simulation algorithm for Tank Cave ................................................................. 77
Figure 4.37. Digitised model of Tank Cave used as training data for the modified non-gridded MPS scanning algorithm; arrow indicates north .......... 78
Figure 4.38. Conditioning data used for the modified non-gridded MPS simulation algorithm; arrow indicates north ................................................................. 78
Figure 4.39. Two different realisations of Tank Cave generated by modified non-gridded MPS ................................................................. 79
Figure 4.40. Distances between training image and 50 realisations for gridded MPS for Olwolgin Cave; blue dots are the distance values between each realisation and the training image; red dots represent the average of the distances ............................................................................................ 80
Figure 4.41. Distances between training image and 50 realisations for modified non-gridded MPS for Olwolgin Cave; blue dots are the distance values between each realisation and the training image; red dots represent the average of the distances ............................................................................................ 80
Figure 4.42. Distances between 50 realisations for gridded MPS for Olwolgin Cave; blue dots are the distance values between pairs of realisations; red dots represent the average of the distances .................................................................................. 81
Figure 4.43. Distances between 50 realisations for modified non-gridded MPS for Olwolgin Cave; blue dots are the distance values between pairs of realisations; red dots represent the average of the distances ........................................81

Figure 4.44. Distances between training image and 50 realisations for gridded MPS for Tank Cave; blue dots are the distance values between each realisation and the training image; red dots represent the average of the distances ........................................................................82

Figure 4.45. Distances between training image and 50 realisations for modified non-gridded MPS for Tank Cave; blue dots are the distance values between each realisation and the training image; red dots represent the average of the distances ........................................................................82

Figure 4.46. Distances between 50 realisations for gridded MPS for Tank Cave; blue dots are the distance values between pairs of realisations; red dots represent the average of the distances ........................................................................83

Figure 4.47. Distances between 50 realisations for modified non-gridded MPS for Tank Cave; blue dots are the distance values between pairs of realisations; red dots represent the average of the distances ........................................................................83

Figure 4.48. Distance plot between 50 realisations and training image for gridded and modified non-gridded MPS methods for Olwolgin Cave .........................................................84

Figure 4.49. Distance plot between 50 realisations for gridded and modified non-gridded MPS methods for Olwolgin Cave .....................................................................................84

Figure 4.50. Distance plot between 50 realisations and training image for gridded and modified non-gridded MPS methods for Tank Cave .........................................................85

Figure 4.51. Distance plot between realisations for gridded and modified non-gridded MPS methods for Tank Cave ..............................................................
List of Tables

Table 4.1. Template properties ................................................................. 54
Table 4.2. First template properties ............................................................ 56
Table 4.3. Second template properties ......................................................... 56
Table 4.4. Third template properties ............................................................. 56
List of Abbreviations

2D two-dimensional
3D three-dimensional
CPU central processing unit
FILTERSIM filter-based pattern simulation
FMA fast marching algorithm
JSD jensen shannon divergence
MDS multi-dimensional scaling
MPS multiple-point statistics
RAM random-access memory
SGS Sequential Gaussian Simulation
SIMPAT Simulation of Patterns
SIS Sequential Indicator Simulation
SNESIM single normal equation simulation
Abstract

Groundwater is a significant water resource and in many parts of the world it occurs in karst aquifers. The modelling of karst systems is a critical component of groundwater resource assessment and flow. Geostatistical techniques have shown useful applications in the area of groundwater research because of their ability to quantify spatial variability, uncertainty and risk. Traditional geostatistical methods, based on variogram models, use only two-point statistics and thus are not capable of modelling the complex and, high-connectivity structures of karst networks. This has led to an increasing focus on spatial multiple-point statistics (MPS) to model these complex systems.

In this approach, a training image is used instead of a variogram. Patterns are obtained by scanning and sampling the training image and during the simulation they are reproduced using MPS. There are two implementations of MPS: (i) gridded and (ii) non-gridded. In gridded MPS, the training image, templates and simulations are based on rigid grids, whereas the spatially flexible non-gridded approach does not depend on rigidly specified grids. The non-gridded approach is relatively new (Erzeybek Balan 2012), and applications, especially in hydrogeology are few; however, the method has been used to simulate paleokarsts in petroleum applications. Non-gridded MPS has potential to improve the modelling of karst systems by replacing the fixed gridding procedure, used in the original form of MPS, by a more flexible grid adapted to each specific application. However, there are some weaknesses in the non-gridded approach reported in the literature. For example, the proposed template cannot properly represent the tortuous nature of a network, and the variation of the passage widths is not taken into account. In the case of a simple channelised system with a constant width, sampling
the central line of the passages is sufficient; however, most karst systems have networks with significantly varying widths. In addition, the variability among the realisations generated by non-gridded MPS is relatively small, indicating that the realisations do not cover the full space of uncertainty. In practical applications, it is not possible to know the exact extent of the full space of uncertainty, but the observed variability of the geology and geomorphology of similar structures would tell us when the variability among the simulations is too small (or too large). A lack of significant variability among simulated realisations makes the method inapplicable. This thesis presents a modified non-gridded MPS method that increases the variability among realisations and adequately captures the tortuosities of karst networks. To do this, it includes the width and constructs an optimal template based on a representative variety of directions adapted to each network instead of considering only a few major directions using a generic template as applied by Erzeybek Balan (2012).

The performance of Erzeybek Balan’s (2012) non-gridded MPS method has only been visually demonstrated, which is not a sufficiently robust measure of performance. In this thesis, a systematic measure is developed to evaluate the variability among the realisations. This provides an objective way of comparing an important feature of the simulations generated by gridded MPS and the proposed modified non-gridded MPS.

The research starts with an investigation and modification of non-gridded MPS. A widely used demonstration image, which is based on a channelised system, is used to compare the performances of the original non-gridded MPS (Erzeybek Balan 2012) and the modified version proposed in this thesis. A distance-based measure is used to evaluate and compare pattern reproduction and the variability of the realisations generated by the modified non-gridded MPS and standard gridded MPS methods. This
distance measure can be used to compare the multiple-point histograms of the realisations and training images. Gridded MPS and modified non-gridded MPS are then applied to two different karst systems—Olwolgin Cave and Tank Cave—and the realisations generated by each method are evaluated in terms of pattern reproduction and the extent of the uncertainty space.

The comparison examples demonstrate that the proposed modified non-gridded MPS generates a larger uncertainty space than that generated by gridded MPS. The results also confirm that modified non-gridded MPS performs significantly better than the original version of non-gridded MPS in terms of a larger (and more realistic) space of uncertainty and pattern reproduction when applied to a complex karst system.
Declaration of Originality

I certify that this work contains no material which has been accepted for the award of any other degree or diploma 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 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 catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.
I express my gratitude and appreciation to my supervisor, Professor Peter Dowd, and I am grateful for his contributions to my academic career. He introduced me to the great area of Geostatistics and I am thankful to him for sharing his knowledge and his expertise with me. His consistent support has allowed me to develop and has provided an optimal environment for research. I would like to thank Elite Editing for providing a professional editor to assist me in preparing my thesis. It was funded in part by the Walter and Dorothy Duncan trust. Last, I would like to thank Mohsen Hasanzadeh, my husband, Zohre Mazaheri, my dear mother-in-law, who passed away last year, Hooshang Hasanzadeh, my father-in-law, and Dariush and Mali Ghobadi, my parents, for their continued support for my goals.
Chapter 1: Introduction

1.1 Overview

Groundwater is one of the most important water resources in some areas around the world where there is a lack of fluvial water sources. Increasing demand for water for domestic, agricultural and industrial use has resulted in calls for efficient and suitable management of this resource. Groundwater management measures require an extensive understanding of the spatial and temporal behaviour of groundwater (Kumar & Remadevi 2006).

This study focuses on karst aquifers, which comprise an important groundwater resource in many fragile environments throughout the world, including South Australia. A karst is a geological feature formed by the dissolution of soluble rocks. Karst areas are mostly composed of chemically soluble rock such as limestone, dolomite, gypsum and anhydrite. A karst is a complex heterogeneous hydrogeological environment. Generally, a karst aquifer is a network of channels or flow pathways that are often referred to as conduits and caves. They are created by the preferential dissolution of calcite, providing for quick and frequently turbulent water flow, as shown in Figures 1.1 and 1.2 (Goldscheider & Drew 2007). Karst environments have complex features that distinguish them from other aquifers. Given their high heterogeneity and significant anisotropy (due to complex geometry), it is difficult to model them and estimate their characteristics (Cherubini & Pastore 2010).

The study of the geometry and connectivity of karst conduits and caves is of considerable interest because the characterisation of these complex features is essential
to understanding the nature of fluid flows in karst aquifers. The modelling and prediction of the flow of fluid and solutes through karst aquifers is difficult because of the heterogeneity and anisotropies of these systems (Peterson & Wicks 2006).

Figure 1.1. Schematic diagram of a heterogeneous karst aquifer (adapted from Goldscheider & Drew 2007)

Figure 1.2. Turbulent water flow in a cave (photo: Wenger)
A significant number of people throughout the world live in regions that have been formed by karsts, and they rely on these aquifers for drinking water. However, there are many problems associated with these aquifers. Given their specific hydrogeological properties, it is difficult and costly to exploit and collect data from these aquifers, and there is often insufficient data to provide reliable models and resource assessments. In addition, karst aquifers are vulnerable to contamination, and they can absorb contaminants from soil or sinkholes (called inlets of karst aquifers) and then quickly distribute them over large distances as a result of fast and turbulent flow in the conduit networks (i.e., flow pathways that are created by the preferential dissolution of calcite) (Drew & Hötzl 1999). Other problems include: leakages from channels and reservoirs, collapse of underground cavities leading to the occurrence of sinkholes, flooding (as illustrated in Figures 1.3 and 1.4) and stability problems in geotechnical applications such as tunnelling and dam construction (Drew & Hötzl 1999). Thus, there is a need for reliable modelling and simulation of karst systems.

*Figure 1.3. Formation of sinkhole in Florida (photo: Yunji De Nies & Daisha Riley 2012)*
The simulation of karst networks on the basis of spatial variability and hydrological and hydrogeological properties is a challenging exercise. Attempts to simulate karst networks using stochastic methods have been reported in the literature. Depending on the nature of the problem, some methods have been more effective than others (Dassargues 1997). Some of the existing methods have been applied only in areas related to karsts; however, their wider applicability to complex channels in karst networks is of significant interest. In this research, a stochastic method that has been recently applied in paleokarst (fossilised karst) reservoirs to model fracture networks (Erzeybek Balan 2012) is investigated and modified to simulate highly complex meandering channels in karst networks. As these methods are being widely developed, there is an increasing need to evaluate them objectively on the basis of their performance rather than solely on the more common basis of visual validation. Accordingly, in this research, a distance-based measure is implemented to provide a quantitative comparison of methods. The next chapter provides an overview of karst
aquifers and outlines different techniques that have been applied to simulate and model karst systems and complex geological structures similar to karsts.

1.2 Research Objective

The primary aim of this research is to develop and validate geostatistical methods for simulating channel networks—particularly in karst systems—and thereby provide the basis for modelling flow in porous and permeable media such as karst conduits and caves. The simulated systems also provide the basis for predicting the extent of the intrusion of contaminants in groundwater.

Research has developed and modified methods using the relatively new geostatistical area of multiple-point statistics (MPS) to simulate channel networks such as those found in karst systems. In the MPS approach, spatial pattern statistics are inferred from a training image representing the spatial distribution of the networks, and these statistics are then combined with specific conditioning information and prior geological knowledge to simulate the spatial distribution of the network. In reservoir characterisation studies, MPS have been widely applied in modelling complex geological structures and fracture networks, as well as integrating different datasets. By analogy, it should also be possible to use MPS to characterise and simulate the curvilinear geological features of karst systems, which are similar to fracture networks and facies distributions in hydrocarbon reservoirs. However, in traditional MPS implementations, a gridded training image is required for inferring the statistics. Some network features are surveyed by recording only the XYZ coordinates of a central line along the network; in these cases, it is not practical to perform gridding by applying a rigidly gridded template to capture statistics using such ‘point’ data surveys. To address
these issues, a relatively new MPS algorithm that works on a non-gridded basis has been used to characterise ‘point-set data’ karst networks. Compared with traditional MPS algorithms, the non-gridded method does not require any gridding of the ‘point-set’ data, which might be challenging for network surveys with more than several thousand measurements. The algorithm basically consists of MPS analysis using spatially flexible templates that are capable of handling complex orientations and subsequent pattern simulation using the statistics. The approach captures spatial distributions of connected pathways and simulates the networks while honouring the conditioning data. However, this method has several shortcomings, such as the ability to capture sufficient spatial variability and the consequent inability to represent the highly complex, tortuous nature of many channel networks. One objective of this research is to overcome these deficiencies through appropriate modification of non-gridded MPS.

To the author’s knowledge, the performance of non-gridded MPS (in terms of pattern reproduction) has never been compared with that of gridded MPS methods. For this purpose, pattern reproduction has been carried out using a fixed two-dimensional (2D) channelised training image for both the modified non-gridded MPS and gridded MPS methods. A distance-based measure is applied to calculate the dissimilarity between the realisations and the training image for both methods. A distance plot is derived from the resulting measures to provide a reliable comparison between these methods.

1.3 Thesis Outline

Chapter 2 presents a literature review of karst aquifers, the geometry of karst networks and existing stochastic simulation methods for modelling karst systems or
complex structures similar to karst networks. It also discusses karst formation mechanisms, factors affecting karst formation and general information on karst aquifers. In addition, the chapter provides a general overview of geostatistical modelling techniques. In Chapter 3, the modified non-gridded MPS analysis algorithm, pattern simulation and corresponding properties are explained using an example of a 2D channelised system. The algorithms are demonstrated using the point-set data, and the corresponding results are presented. Chapter 4 presents an approach for evaluating the method by calculating the distance between the realisations and the training image. This approach is applied in both the modified non-gridded MPS and gridded MPS methods, and the ability of the methods to reproduce spatial patterns is compared. Chapter 5 concludes this thesis with a discussion of the main findings and suggested directions for future research.
Chapter 2: Literature Review

2.1 Overview of Karst Aquifers

Karst aquifers are made heterogeneous through the presence of conduit networks with high permeability inside a fractured matrix with low permeability (Bakalowicz 2005). Practically, these aquifers are a kind of porous media formed by carbonate rocks. When rainwater or streams enter fractures and bedding planes from the inlets, under the influence of dissolution, the fractures are enlarged and form a hierarchical conduit structure that redirects water towards outlets known as springs (Bakalowicz 2005).

It is generally assumed that karst aquifers consist of three zones (as shown in Figure 2.1). First, the epikarstic zone, which has little depth and relatively high permeability, prevents surface water runoff and conducts water. This zone recharges into neighbouring zones by diffusion (Geyer 2008). Second, there is an unsaturated zone, which is also known as the vadose zone. This zone is deeper than the epikarst layer, and the direction of conduits is generally vertical (Rooij 2008). Third, the saturated zone (or phreatic zone) is located in the deepest layer with a mainly horizontal direction. The latter zone is of considerable importance due to the concentration of water flow in the conduits, which can transmit a highly turbulent flow towards the discharge points or springs (Geyer 2008).
2.2 Geometry of Karst Networks

The preferential dissolution of carbonate rocks develops conduit networks. The length of conduits may reach hundreds of kilometres, and their diameter varies from a few centimetres to tens of metres. These characteristics are the cause of significant inhomogeneity in the geometry of karst systems. As mentioned in the previous section, conduit networks are considered a recharge component in karst aquifers and therefore have a critical role in the flow of water towards springs; thus, it is important to understand their location and shape. These features of karst conduits rely on two main factors: the location of outlets and the existing inception horizons (Filipponi & Jeannin 2010). Moreover, the conduit geometry is locally characterised by: (1) the geometry of the conduit cross-sections (i.e., 2D profiles defined by size and shape); (2) the type of flow and the texture and structure of the rocks; (3) the geometry of passages in the
direction of flow, which may be straight (in the direction of major fractures), angular (under the influence of high gradients) or curvilinear (in the direction of bedding planes); and (4) the geometry of the network and its hydraulic properties (Rooij 2008).

2.3 Existing Techniques for Modelling Meandering Channels and Karst Systems

The modelling of karst networks is difficult and challenging because of the lack of knowledge of the detailed geometry of the networks—specifically, the locations, sizes and shape of the conduits (Rooij 2008). In practice, very little is invested in collecting accurate data, and most data are obtained from amateur speleologists. Various techniques have been developed to simulate conduit geometry over the past three decades (e.g., modified lattice-gas automaton, sequential Gaussian simulation, sequential indicator simulation,); among them, geostatistical methods have provided the best results. Geostatistical methods provide probabilistic models that can be used to quantify the uncertainty of the geometry of karst conduits. They can also be conditioned to field observations; see explanation below and Jaquet et al. (2004), Henrion et al. (2008), Borghi et al. (2012), Pardo-Igúzquiza et al. (2011), Fournillon et al. (2010), Strebeille (2002), Hu and Chungunova (2008), Huysmans and Dassargues (2009), Arpat and Caers (2007), Zhang et al. (2006), Honarkhah and Caers (2010) and Erzeybek Balan (2012). These methods are generally divided into two groups: traditional geostatistical methods, which are based on the variogram; and multiple-point geostatistics, which rely on MPS instead of two-point statistics (as applied in variogram-based methods), as explained below.
2.3.1 Traditional geostatistical techniques

Jaquet et al. (2004) outlined an important application of geostatistical methods to simulate karst systems by using a modified lattice-gas cellular automaton method that relies on a random-walk technique applied to particle displacement to simulate the geometry of karst conduits. This method assumes that the fluid contains particles that represent dissolution processes. When the fluid enters the medium, particles spread over the areas with a higher density of fractures due to the dissolution processes. Particles preferentially erode the rock and create conduits within these areas. The size of the conduits depends on the number of particles flowing through them. The presence of a simulated conduit influences the shape of the next conduit, thereby creating a hierarchical structure in the karst conduit network. Jaquet et al. (2004) then combined their model with flow simulation using a finite element method (see Figure 2.2). However, as this method is based on a random walk, it is difficult to condition the model to field observations. In addition, their study was limited to the simulation of 2D conduit network geometries, and they did not take into account variable velocity fields in fractured domains.
Other authors have emphasised the necessity of modelling the conduits for flow and transport simulation. For example, Borghi et al. (2012) developed a new method to model the geometry of conduits that honoured the modelled geology and then used the resulting model for the flow and transport simulation. Their approach was based on a pseudo-genetic algorithm. Borghi et al. (2012) stated that the shape and position of conduits are influenced by two main factors: the geology and the hydrological and hydrogeological conditions. They first constructed a three-dimensional (3D) geological model of the study domain and then simulated heterogeneous hydrogeological features including fractures, bedding planes and inception horizons using stochastic methods. Stochastic methods are promising because of their flexibility in defining recharge and discharge areas. That is, the potential inlets and outlets of the karst systems, as well as the base level and different phases of karstification, can be predicted. Finally, conduit networks are extracted via an efficient shortest path technique called the fast marching algorithm (FMA). Consequently, the recharge area is connected to the main discharge area (see Figure 2.3; Borghi et al. 2012). The FMA method is a subclass of level-set...
methods and relies on a random-walk technique (similar to Jaquet et al. 2004) to compute the least possible travel time between deterministic outlets (springs) and any point in the area. As a result, a network of karst conduits is generated iteratively.

Figure 2.3. 2D map of conduit network realisations; black lines: inlets, which are the starting points of the conduits; green points: outlet or karst spring (Borghi et al. 2012)

As mentioned above, Borghi et al. (2012) modelled flow and transport through their simulated karst systems. This subsequent modelling, when compared with actual flows and transport, can be used to assess the validity of simulated karst structures. Their approach employed a comprehensive range of information, such as inception horizons, which significantly influence the genesis of karst networks. The models were based on a probabilistic framework, thus allowing the evaluation of uncertainty. However, there were several drawbacks in their proposal. For example, the apertures of fractures were not taken into account, although apertures significantly influence the shape and direction of conduits. Additionally, their study only focused on a type of karst reservoir known as an epigenetic karst.
In addition to Borghi et al. (2012), other authors, such as Fournillon et al. (2010), have relied on 3D geological models to describe karst networks. The authors applied one of the most common geostatistical methods—variogram-based Sequential Indicator Simulation (SIS; Deutsch & Journel 1998)—to model the spatial distribution of conduit networks. In their approach, the presence/absence of a karst conduit was quantified by an indicator function equal to 0 (non-conduit) or 1 (conduit). Several groups of karst networks were created, and the variogram and proportion of each group were estimated (Fournillon et al. 2010a). The method can be used to generate 3D probability maps of karst networks. Simulations can be conditioned to any field observation data, including existing conduits. However, the method cannot reproduce the observed connectivity of karst conduits; real karst conduits are very complex, and the determination of their connectivity is very important. Fournillon et al. (2010b, 2012) later achieved more promising results by using an algorithm that extracts the topology and geometry of observed karst networks and compares them statistically with simulated networks.

To simulate the topology and geometry of karst networks, Pardo-Igúzquiza et al. (2011) suggested a simulation technique that addresses the complexity of karst networks and honours any existing observations. They initially generated non-connected karst conduits by means of a re-sampling method and then combined them using a modified diffusion-limited aggregation method to create connected conduit networks (as shown in Figure 2.4). In their work, both conditional and non-conditional karst conduit networks were simulated, and statistical comparisons were made between the two groups and real karst networks. The results showed that their approach was useful for modelling the geometry and connectivity of karst networks. However, they did not incorporate fractures (and thereby fracture networks) in the model. The
incorporation of fractures in conduit networks could improve the reliability of the resulting simulations.

![Figure 2.4](image.png)

*Figure 2.4. (a) 3D map of the non-connected conduits; (b) 3D map of a realisation of a karst network with connectivity using a modified diffusion-limited aggregation technique (Pardo-Igúzquiza et al. 2011)*

Most authors have focused on generating karst conduit networks and not cave geometry, although there are usually large caves in real karst aquifers. Several authors have implemented geostatistical approaches based on the formation of complex caves (e.g., Henrion et al. 2008; Erzeybek Balan 2012). Henrion et al. (2008) first presented a stochastic discrete method to identify preferential pathways in a generated fracture network, and then used a variogram-based geostatistical simulation to reproduce 3D cave networks surrounding these paths. In this study, Henrion et al. (2008) used a graph search algorithm to extract preferential flow pathways and eventually modelled the geometry of caves using Sequential Gaussian Simulation (SGS; as illustrated in Figure 2.5). Although their study demonstrated the efficiency of geostatistical approaches for producing realistic cave models, there were some disadvantages. First, the study lacked a deeper understanding of the fracture network. Second, they did not consider the influence of cave geometry on fluid flow.
2.3.2 Multiple-point geostatistics

The methods described above are, more or less, traditional geostatistical methods and are variogram-based. They characterise the correlation between only two points in space. Hence, it is difficult (some would say impossible) to model the complex geometry of reservoirs with high connectivity using these approaches. Multiple-point geostatistics were developed to overcome the inability of traditional geostatistical simulation techniques to adequately model complex structures that have high heterogeneity, such as meandering channels in hydrocarbon reservoirs (Strebelle 2002; Arpat & Caers 2007; Zhang et al. 2006; Hu & Chugunova 2008; Honarkhah & Caers 2010). Although this method was first applied in the field of petroleum engineering, it has recently been applied to karst systems (Erzeybek Balan 2012).

MPS simulation was first introduced by Guardiano and Srivastava (1993). They claimed that using just some informed nodes of the simulation grid is not enough for simulation, and that there is a need for a probabilistic tool to depict the behaviour of the simulation space. Accordingly, a training image was introduced for the first time to represent the probability distribution function of the domain under simulation. By using the training image instead of the variogram, multiple-point geostatistics simulation
methods have been able to reproduce more complex patterns than two-point statistics simulation methods such as those explained above.

The first practical multiple-point geostatistics technique, called single normal equation simulation (SNESIM), was developed by Strebelle (2002). The algorithm is probabilistic and replicates patterns by reproducing statistics or conditional probabilities by scanning the training image. After scanning the training image, the conditional probabilities of different patterns are stored in a dynamic data structure, the search tree. As the method uses a large search tree, it is computationally expensive.

Hu and Chungunova (2008) and Huysmans and Dassargues (2010) successfully applied the method originally proposed by Strebelle (2002) to model subsurface heterogeneity. However, Huysmans and Dassargues (2010) applied the method only to hydrogeology, and applications are rare in this field. Although the method was successfully implemented in their work, its limitations should not be ignored. The original method required large amounts of random-access memory (RAM) and long computing times. Moreover, non-stationarity was not considered in the algorithm, and the method was limited to categorical variables. The authors stressed the importance of building an appropriate training image as a critical step of MPS simulation. The training image should be representative of the geological heterogeneity and must be large enough to characterise the essential features by the implicit statistics defined on a limited point configuration. Thus, an improvement was needed in terms of implementation (Hu & Chungunova 2008; Huysmans & Dassargues 2010). In 2011, Huysmans and Dassargues developed SNESIM further to make it significantly faster. Their method was named Direct MPS Simulation of Edge Properties and used edge properties instead of pixel values. It led to a larger cell size for MPS simulations and consequently decreased central processing unit (CPU) times and memory requirements.
Moreover, both continuous and categorical variables can be used in this methodology. However, the realisations generated were not entirely convincing reproductions of the features in the training images.

Arpat and Caers (2007) proposed a pattern-based algorithm called Simulation of Patterns (SIMPAT) to enhance the reproduction of the training image patterns. The approach can use both categorical and continuous variables. In this approach, a pattern database is first built by scanning the training image. The sequential simulation then selects a pattern from the database and pastes it onto the simulation grid (Arpat & Caers 2007). As patterns are extracted from the training image and directly pasted onto the simulation grid, any pattern, including a non-stationary element of the training image, will appear in the simulation. In addition, this algorithm is less demanding than SNESIM in terms of computer memory and computation time (Arpat & Caers 2007). However, the method is still time-consuming because all of the training image patterns are searched to find the match.

Zhang et al. (2006) presented an MPS method that is similar to SIMPAT, but that does not search all of the training patterns to find the match. The method, called filter-based pattern simulation (FILTERSIM), clusters the training image into groups to reduce the dimensionality of the space of patterns. In this approach, a few features are extracted from each training image pattern using some given filters in the feature space of regularly spaced cells. Simulation data events are compared with the representatives of the clusters, and a random pattern from the most similar cluster is pasted to the simulation grid. FILTERSIM is a fast method, but there are many user-set parameters. Further, reproduced patterns in the realisations are not satisfactorily realistic. Thus, there is still a need to improve these pattern-based methods.
Honarkhah and Caers (2010) developed a modified clustering-based method, called distance-based pattern modelling, to reduce the number of clusters and enhance the quality of the realisations using kernel space mapping. The methodology is fast because it uses cluster representatives instead of all available patterns (similar to FILTERSIM). Patterns are successfully reproduced using this method, and the realisations are more realistic than those generated by the method proposed by Zhang et al. (2006), as they show more variability.

The multiple-point simulation methods described above are applied to gridded fields (Strebelle 2002; Hu & Chungunova 2008; Huysmans & Dassargues 2009; Arpat & Caers 2007; Zhang et al. 2006; Honarkhah & Caers 2010). Erzeybek Balan (2012) reported a recent development of multiple-point geostatistics in which a training image was sampled on a non-gridded field to reproduce the statistics of connected geological structures; the example application was in paleokarst reservoirs using sparse data. The MPS were first extracted from real cave information, and afterwards the patterns were reproduced to simulate the spatial distribution of the paleokarst systems (Erzeybek Balan 2012). The author demonstrated the potential of geostatistical models to simulate realistic geometries. Erzeybek Balan (2012) implemented this approach for the first time in a paleokarst oil reservoir using a point dataset. The work showed significant potential for the use of karst network modelling in petroleum applications. Further, as this approach is pattern-based and relies on MPS, it can simulate complex structures, whereas conventional approaches are variogram-based and therefore can only use two-point statistics. Another advantage of this method is that using non-gridded flexible templates instead of conventional gridded templates makes it more applicable to complex systems when using point-set information. However, karst networks are represented by surveys reporting XYZ coordinates of karst networks on a central line.
This method was only applied on an epigenic karst, whereas modelling hypogenic karst networks is more appropriate for modelling karst reservoirs. Moreover, uncertainty was not considered in the simulated networks, but it could be evaluated by applying different spatial templates and different tolerance windows. A further difficulty is that the method considers point locations, but not the widths of structures at those locations. This is of no consequence for the simple demonstration example in which the width of the system is constant, but it is a problem for the general case of varying widths. Additionally, the algorithm for selecting the template could be improved, and the proposed template appeared to be unable to adequately represent the tortuous nature of the network. In karst systems, the training images usually contain more complex geological structures than those used by Erzeybek Balan (2012), and it appears that there is insufficient variability in her model (as shown in Figure 2.6). Therefore, there is still a need to improve the performance of pattern reproduction in this approach.

Figure 2.6. Top view of pattern simulation of large-scale passages; (a) original data, (b) simulated network, red dots are conditioning data, lines are simulated connections (Erzeybek Balan 2012)
2.4 Current Approaches in Evaluating Geostatistical Methods

It is important to recognise the differences among the simulation methods in order to choose the most appropriate one for both the geological structure to be modelled and the type of predictions the model will be used for. Geostatistical techniques have previously been evaluated on the basis of visual comparisons of the images generated by the various methods and/or the statistics of the images. However, visual comparisons could result in significant errors. For example, Soleng et al. (2007) presented a simple metric that relied more on visual comparisons. Using their algorithm, a number of realisations were simply scanned, and the global volume fractions of each facies and the number of facies bodies of each type were computed. Then, in each direction, the algorithm computed the surface areas, volumes and extensions. A simple statistical analysis of the realisations from three different simulation methods—variogram-based methods, multiple-point methods and sequential Markov random fields—was implemented, and the summary statistics were checked against the properties of a training image. The results were only considered in terms of similarity or dissimilarity obtained through the simple visual inspection, although the algorithm was fast and applicable in 2D and 3D.

De Iaco and Maggio (2011) applied the algorithm proposed by Soleng et al. (2007) to compare a traditional variogram-based simulation algorithm (e.g., SIS) with an MPS algorithm (e.g., SNESIM) to reproduce the realistic geological continuity. The simulation was carried out for a spatial distribution of limestone with meandering channels. The similarity or dissimilarity obtained through merely visual inspections of the simulated images confirmed that the spatial distribution of limestone with meandering channels and other statistics were better reproduced using an MPS
algorithm (SNESIM) rather than traditional variogram-based methods (SISIM). This study could be more comprehensive by using other spatial simulation methods for comparisons, such as the most recent training-image-based simulation algorithms.

dell’Arciprete et al. (2011) reported a connectivity measure for comparing three geostatistical simulation techniques (SIS, transition probability geostatistical simulation and multiple-point simulation) at different scales for a well-exposed aquifer analogue. Using these simulation techniques, several realisations were computed for a test volume of about 400 m$^3$ and for the entire volume. The results were compared by visual inspection and connectivity analysis of the very, or poorly, permeable structures when characterising aquifers and reservoirs by their flow properties. The comparison showed that the geological model was best reproduced when the simulations were realised separately for each highest-rank depositional element and subsequently merged. Further, the three simulation methods gave different images of the volume. MPS was especially efficient in reproducing the geometries of the most represented hydro-facies, whereas SIS and transition probability geostatistical simulation were better at reproducing the distribution of the less-represented facies.

The studies reported so far have relied more on a visual comparison or simple metrics such as connectivity analysis; however, they have only addressed one issue—that is, the reproduction of statistics without taking into account the space of uncertainty. Tan et al. (2014) reported a numerical distance measure that provides an accurate comparison of the geostatistical methods. Their practical methodology measured performance and ranked training-image-based geostatistical algorithms without considering the nature of the training image (e.g., whether it was discrete or continuous, on grids of small or large dimensions, or 2D or 3D). They applied a statistical measure of distance, called the Jensen–Shannon divergence (Cover &
Thomas 1991) to compare two frequency distributions. Their paper focused on two concepts: the between-realisation variability (i.e., the space of uncertainty) and the within-realisation variability (i.e., reproduction performance). Although the methodology was efficient, it was restricted to only the standard MPS methods, such as SNESIM, which rely on the effective extraction of statistics from the training images. Moreover, their work was limited to only unconditional simulations; however, an investigation of conditional simulation could be interesting, as conditioning data would affect the space of uncertainty.

### 2.5 Gap Statement

From the literature review, it can be concluded that geostatistical methods can provide realistic and accurate models of karst network geometries. As they are mostly variogram-based (e.g., SGS, SIS), they have a limited ability to be conditioned to field data and to quantify the full range of uncertainty, as they are based on two-point statistics (Strebelle 2002). Thus, such conventional methods cannot reproduce the very high heterogeneity (complexity) of karst structures. Multiple-point geostatistics were developed to address this problem (Strebelle 2002; Hu & Chungunova 2008; Huysmans & Dassargues 2009; Arpat & Caers 2007; Zhang et al. 2006; Honarkhah & Caers 2010; Erzeybek Balan 2012). However, in the gridded version of MPS, the training images, templates and simulations are based on rigid grids that, for very complex shapes and connectivity, may not be able to extract the complete range of patterns from a training image. The spatially flexible non-gridded approach is not limited in this respect, and it can be adapted to the complexity in the training image, especially for point datasets. The non-gridded approach is relatively new (Erzeybek Balan 2012), and there have
been few applications, especially in hydrogeology; however, the method has been implemented to simulate paleokarsts in petroleum applications. A significant gap in this algorithm is the selection of the template, which requires modification to ensure that it adequately represents the network tortuosities, especially when using a channelised training image, and captures the full range of spatial variability. Moreover, the method considers point locations, but not the widths of structures at those locations. This is of no consequence for the simple demonstration example in which the width of the channelised system is constant, but it is a problem for the general case of varying widths. Therefore, there is a need to improve the performance of pattern reproduction in non-gridded MPS and provide an objective means of comparing the applicability of this method with the gridded MPS method. The specific goal of this comparison is to improve the understanding of the strengths and weaknesses of the MPS method. This requires a numerical measure to provide more accurate results (Tan et al. 2014). Finally, recent work on the comparative evaluation of geostatistical methods has been restricted to unconditional simulations; thus, it would be useful to extend this to the investigation of conditional simulations.

2.5.1 Descriptive list of gaps

The following list outlines the gaps that have been addressed in this research:

- Develop and modify the non-gridded MPS method, especially in terms of selecting a suitable template, so it has a critical effect on pattern reproduction for modelling complex structures such as karst systems, and include the width in order to adequately capture the tortuosities of karst networks.
• Use a channelised training image with more complexity instead of a training image that represents simple fractures as a line without considering the tortuosities.

• Improve the spatial variability of the realisations obtained by non-gridded MPS simulation.

• Quantify the performance of pattern reproduction in the non-gridded MPS method.

• Develop a method to evaluate the spatial variability between realisations generated by the non-gridded MPS method so as to evaluate the space of uncertainty.

• Compare non-gridded MPS results with gridded MPS results in the case of conditional simulations using a distance measure to evaluate the ability to reproduce the full range of variability.
Chapter 3: Methodology

3.1 Introduction to Geostatistical Methods

It is difficult and costly to collect data from karst networks, and most detailed information is obtained from mapping by amateur speleologists. Geostatistical methods can be used to estimate or simulate karst topology from the sparse data. These methods rely on models of spatial variability inferred from the data.

Geostatistical simulation produces a series of realisations (e.g., images) that represent the spatial distribution of the variable of interest (Vann et al. 2002). These images differ from each other, and the differences among them provide a means of quantifying the uncertainty caused by insufficient data. Analysis of uncertainty is crucial for risk assessment and decision-making (Deutsch 2002). The general concept of geostatistical simulation is illustrated in Figure 3.1.
3.2 Potentially Useful Methods

Geostatistical simulation generates a series of images, or realisations, with each representing a possible reality based on the spatial variability and distribution of the variable as inferred from the data. The believability of these possible realisations depends on the assumptions and methodology applied in the simulation process (Vann et al. 2002). Geostatistical simulation methods can be classified into two broad categories: two-point statistics and MPS.
3.2.1 Two-point statistics

Methods in this category are also known as variogram-based techniques, as they describe spatial continuity in terms of a variogram. A variogram is defined as the variance of the difference between the values of a variable at two locations—that is, it characterises the spatial dependence (or correlation) of a variable between two points in space. Thus, this type of geostatistical simulation reproduces the spatial variability of data as quantified by a variogram model (Vann et al. 2002).

Two of the most common methods of two-point geostatistical simulation are SGS and SIS. SGS is an efficient variogram-based method that is widely used in hydrogeology applications. This technique is suitable for continuous variables such as porosity. Data are transformed to a Gaussian distribution, and a variogram model is established for the transformed values. A random path is defined through all simulation grid nodes. The data and all previously simulated values are used to krige the value of the variable at a grid node. The kriged value and the associated kriging variance are interpreted as the estimated mean and variance respectively of the Gaussian distribution of possible values at the grid node. A value is drawn at random from this distribution and assigned to the grid node as the simulated value of the variable. The process is repeated at the next grid node in the random path and continued until simulated values have been assigned to all nodes.

Although this method is flexible, some disadvantages decrease its applicability. In this method, the size of the field being simulated must be much larger than the ranges of the variogram model. Moreover, the search neighbourhood selection is a critical stage, and the selection of small neighbourhoods can lead to poor conditioning and poor reproduction of the variogram (Vann et al. 2002). The biggest problem in this method
(and all others based on the variogram) is that it is difficult to capture the spatial variability of complex structures in the variogram—for example, curvilinear geological features such as tortuous fractures and meandering channels (as typically observed in karst structures).

SIS is based on indicator (co)kriging and was originally developed for reservoir modelling in the petroleum industry, and especially for structures with high connectivity in spaces that could not be adequately modelled and simulated using SGS. SIS can be used for both continuous and categorical variables such as facies. Indicator simulations can generate complex heterogeneity patterns and different spatial correlation structures for each facies. At a given location, the facies is represented by an indicator variable that takes the value 1 if the facies is present and 0 if it is absent. Using an individual variogram to represent each indicator variable makes the method more flexible when handling complex structures compared with other common traditional geostatistical methods (dell’Arciprete et al. 2011). However, the possibility of inconsistent indicator variograms can cause probability constraints to be violated. Further, computation time increases significantly as the number of indicator variables increases. Additionally, as the method is limited to only two-point statistics, it prevents an accurate reproduction of highly complex reservoir structures, which would be a major problem (Strebelle 2002).

3.2.2 Multiple-point statistics

Multiple-point geostatistics were developed to address the inability of traditional geostatistical simulation techniques to model complex structures with high heterogeneity, such as meandering channels (which are similar to karst networks). Multiple-point geostatistics use implicit cross-correlation moments at three or more
locations to reproduce more realistic spatial patterns. These MPS are obtained by scanning and sampling a conceptual model or training image (Guardiano & Srivastava 1993). The training image effectively replaces the variogram as a model geological spatial heterogeneity. A training image depicts the expected spatial distribution of features of interest. In fact, the training image contains the patterns or MPS related to the spatial features under study (Strebelle 2002). Digitised photographs of outcrops and geological maps may be used as training images (Strebelle 2012). The idea is to capture the patterns of variability from the training images and incorporate these patterns into the simulation. Therefore, building a suitable training image is one of the most critical stages of MPS.

![Figure 3.2. A geological model using a training image (Honarkhah & Caers 2010)](image)

There are two ways of implementing MPS: gridded and non-gridded. In the gridded version of MPS, the training images, templates and simulations are based on rigid grids, whereas the spatially flexible non-gridded approach does not depend on rigidly specified grids.
3.2.2.1 Gridded MPS method

In this approach, the training images, templates and simulation results are in the form of grids (Strebelle 2002; Arpat & Caers 2007; Zhang et al. 2006; Honarkhah & Caers 2010), and the training images are usually in binary format (see Figures 3.3 and 3.4). In general, the training image is sampled using the gridded template to obtain patterns of the training image (see Figure 3.5). These patterns are then applied to simulate processes in different ways according to various algorithms.

![Figure 3.3. Example of gridded training image (Honarkhah & Caers 2010)](image)

![Figure 3.4. Example of gridded template; a 3×3 template on an 11×11 grid (Arpat & Caers 2007)](image)
The various algorithms include SNESIM (Strebelle 2002), SIMPAT (Arpat & Caers 2007), FILTERSIM (Zhang et al. 2006) and the distance-based method (Honarkhah & Caers 2010).

The SNESIM algorithm is a probabilistic method that builds a database of patterns detected in the training image, and this database is searched to find the best match to the current data event to generate a conditional simulation (Strebelle 2002). A data event is the set of observed values in the current simulation grid window, and it contains previously simulated values as well as hard conditioning data. Hard conditioning data are the data sampled directly from the space under simulation (Strebelle 2002). The method uses a large search tree; hence, it requires extensive computational effort. Further, this algorithm is only suitable for categorical variables. Moreover, non-stationarity is not considered in the algorithm.

Figure 3.5. (a) Sampling of a gridded training image using a gridded template, (b) observed patterns (after Arpat & Caers 2007)
In SIMPAT, patterns are extracted from the training image by scanning the training image using a template (see Figure 3.5), and the patterns are directly pasted to the simulation grid in the sequential simulation process. If a pattern contains a non-stationary element of the training image, this behaviour will appear in the simulation. All training image patterns are searched to find the match (Arpat & Caers 2007).

FILTERSIM is similar to SIMPAT, but it does not search all training image patterns to find the match. Instead, patterns are clustered into groups to reduce the dimensionality of the pattern space (Zhang et al. 2006). In this approach, a few features are extracted from each training image pattern using specified filters in the feature space. Simulation data events are compared with the representatives of the clusters, and a random pattern from the most similar cluster is pasted to the simulation grid (Zhang et al. 2006). FILTERSIM is a fast method, but there are many user-set parameters. Further, reproduced patterns in the realisations are not always satisfactorily realistic.

Distance-based pattern modelling (Honarkhah & Caers 2010) is a modified clustering-based method that applies distance-based methods to represent patterns in the training image as points in a binary space. The advantages of this method are the reduced number of clusters required and the enhanced quality of the realisations by using kernel space mapping. The methodology is fast because it uses cluster representatives instead of all available patterns (as in FILTERSIM). Patterns are successfully reproduced with this method, and the realisations are more realistic than those generated by FILTERSIM, as there is greater variability among the realisations (Honarkhah & Caers 2010). As this is a comprehensive method within a unifying framework, it has been chosen as a gridded MPS method for comparison with a non-gridded method. The algorithm depicted in Figure 3.6 provides a general description of the gridded MPS simulation within the context of distance-based pattern modelling.
Obtain the optimal template by analysing the training image

Scan the training image using the template

Store all patterns in the pattern database

Map the patterns to points in MDS space using distance-based method

Map points in MDS space to kernel space

Apply k-means clustering on dataset in kernel space for pattern classification

Define a random path on the grid of realisation

At every node along the random path:
- Extract the data event from realisation
- Find the cluster prototype closest to the data event
- Sample a pattern from the prototype class
- Paste the pattern to the realisation

Move to the next node of the random path and repeat the above steps until all the grids nodes along the random path are exhausted

Figure 3.6. General algorithm of distance-based pattern modelling
3.2.2.2 Non-gridded MPS method

Non-gridded MPS is a relatively new technique proposed by Erzeybek Balan (2012). It has potential to simulate realistic cave network geometries described in the form of point-sets (i.e., locations or coordinates of cave networks), and thus provide a more realistic assessment of flow mechanisms. In this approach, a non-gridded training image and a non-gridded spatially flexible template are used to reproduce statistics of the structures from sparse data. MPS are first extracted from the non-gridded training image (real information), and afterwards the patterns are reproduced to simulate spatial distributions over the area under study (Erzeybek Balan 2012). The use of a non-gridded training image and non-gridded flexible templates instead of gridded ones (as applied in conventional MPS methods) provides the flexibility required to capture and simulate complex structures using point datasets. For example, karst networks are represented by data comprising the spatial coordinates of a central line through the karst conduits (Erzeybek Balan 2012). This method is implemented in three steps, which are described in the following sections.

3.2.2.2.1 Non-gridded MPS pre-processing

Non-gridded MPS pre-processing can be summarised as (Erzeybek Balan 2012):

1. Non-gridded training image

A training image is digitised to produce a non-gridded training image. The training image is essentially a set of point training data.
2. Non-gridded template

An optimal template is determined by considering a generic template with 19 nodes surrounding a central node. These 19 nodes are separated by an angle of 5° from each other over a range of 0° to 90°, as shown in Figure 3.7. The lag distance is the distance between the centre node and each of the surrounding nodes; it is based on the length of connected passages in the network. The non-gridded training image is scanned using the generic template to identify network orientation and average passage length before calculating MPS. The number of connections captured by each node of the generic template is compared, and configurations that honour the overall geology and have a high number of connections are determined in order to generate the optimal template. A tolerance window is constructed around each template node using user-specified distance and angle tolerances (see Figure 3.8). Distance tolerance values are specified as a fraction of the lag spacing between the nodes of the template, while angle tolerances are defined in a manner similar to variogram computation using irregularly spaced data.

![Figure 3.7. Generic template (adapted from Erzeybek Balan 2012)](image-url)
3.2.2.2.2 Non-gridded MPS analysis

Non-gridded MPS analysis of the point-set training image is implemented using the spatial template obtained from the generic template. A pattern histogram that records the frequency of the observed patterns is constructed by scanning the training image. The non-gridded MPS scanning algorithm is described as follows (Erzeybek Balan 2012):

- Define a template with angle and distance tolerance.
- Select a point location from the point-set training image.
- Centre the template on the selected point and scan the current pattern of surrounding data. If a data point falls within the tolerance window, the corresponding template node is activated and the resultant pattern is stored.
- Store the pattern of data observed at the current location.
- Update the pattern frequency by translating the template over the entire dataset and computing the resulting pattern histogram.
3.2.2.3 Non-gridded MPS simulation

The pattern simulation is implemented after scanning the training image and obtaining the frequency pattern histogram. Based on this algorithm, the simulated network starts growing at the conditioning data locations that are randomly visited (Erzeybek Balan 2012). The algorithm comprises the following steps (Erzeybek Balan 2012):

- Randomly select one of the conditioning data points.
- Place the pattern with the highest frequency at the selected point.
- Consider the node, including the conditioning data in its vicinity, a simulatable node.
- Visit all conditioning data locations at random and add simulatable nodes to a list.
- Randomly select the simulatable nodes from the list, centre the template on them and scan the data configuration. The nodes, including the data in their vicinity, are considered new simulation nodes and are added to the list.
- Repeat the previous step until the specified maximum simulation iteration is reached.

3.2.2.3 Modified non-gridded MPS method

The non-gridded MPS technique has the potential to simulate realistic cave networks described in the form of point-sets. The methodology has only been implemented in the modelling of paleokarst networks in petroleum applications; it has not been investigated in hydrology. The published work is limited to modelling and
simulating simple fractures, as shown in Figure 3.9. There appears to be little spatial variability in the realisations, as can be seen by the simulated patterns following only the limited configuration of the template.

As the template configuration considerably influences the pattern simulation, the selection of an optimal template is a critical step in the non-gridded MPS approach. Although this technique is flexible, the template selection algorithm requires modification to adequately represent the highly complex, tortuous nature of many karst networks. Moreover, for the general case of varying widths, the central line is digitised, but in addition the width is included by the resolution of the pixels used in digitisation. There is, however, still an increasing need to improve the performance of pattern reproduction in non-gridded MPS. Accordingly, this research presents a modified non-gridded MPS that focuses more on the spatial variability and tortuosities of karst networks.

Figure 3.9. Left: non-gridded training image; middle and right: two realisations generated with non-gridded MPS by Erzeybek Balan (2012) using a C++ code
In modified non-gridded MPS, the optimal template is constructed from a variety of possible directions of the network instead of only considering limited major directions using a generic template. The optimal template is a template that has several nodes honouring major directions, as well as the possible directions around them. Moreover, the template size must be small enough to represent the complex tortuous nature of the karst network. To improve spatial variability, several templates with different angles are considered and selected at random in the simulation process. The algorithm in Figure 3.10 provides a general description of the MPS simulation within the context of the modified non-gridded approach. As shown, there is no need to use the generic template proposed by Erzeybek Balan (2012). The point-set training image is directly scanned using a template composed of 37 nodes with a range of angles between $0^\circ$ and $180^\circ$ that can capture all possible connections of the passages in the network (see Figure 3.11). For simulation, patterns with high frequencies are considered templates to be randomly centred on the conditioning data to generate simulatable nodes, whereas only one pattern is used in original non-gridded MPS, and this results in less variability. Moreover, the size of the templates is determined by the tortuosity of the karst networks.
Digitise the training image to have a non-gridded training image

Scan the point-set training image using a template with 37 nodes in a range of angles between 0° and 180°

Store all observed patterns

Obtain observed pattern frequency histogram

Determine patterns with high frequencies

Determine simulatable nodes using patterns with highest frequencies
At every point location:
- Centre the patterns with highest frequency at random
- Find the patterns included any conditioning data point within their tolerance window
- Store configuration
- Add observed node to a simulatable node list

Determine simulation nodes using patterns with high frequencies at random (templates are randomly centred on simulatable nodes)

Repeat the above steps until the maximum simulation iteration is reached

Figure 3.10. General algorithm for the modified non-gridded MPS
3.3 Evaluation of Modified Non-gridded MPS

A general overview of the work undertaken in this research is illustrated in Figure 3.12. One component of this framework is measuring the variability among the realisations resulting from the simulations generated by different methods in order to assess the methodologies. For this reason, several settings should be defined for the simulations. In particular, the realisations generated by the distance-based (gridded) MPS method are compared with those generated by the modified non-gridded MPS. In addition to visual comparisons, they will be compared systematically using a distance-based measure (Tan et al. 2014) that ranks the algorithms for the purpose of assessing their capability in pattern reproduction and evaluating the space of uncertainty.
As shown in the framework, this work is explored with two purposes:

1. Evaluate the variability among the realisations to evaluate the space of uncertainty (or between-realisation variability; Tan et al. 2014).
2. Evaluate the variability between the training image and the realisations to evaluate pattern reproduction (or within-realisation variability; Tan et al. 2014).

3.3.1 Measuring variability among realisations

The algorithm comprises the following steps:

- Obtain some statistics of each realisation. In this case, the statistics are frequency histograms of simulated patterns for each realisation.
Calculate a statistical measure of distance called the Jensen–Shannon divergence (Lin 1991).

In statistics, the Jensen–Shannon divergence is a popular method of measuring the similarity between two distributions. It is based on the Kullback–Leibler divergence (i.e., a measure of the difference between two probability distributions), with some notable (and useful) differences, including that it is symmetric and it is always a non-negative finite value. The extension of this measure has been derived based on Jensen’s inequality and the Shannon entropy (Endres & Schindelin 2003):

\[ \text{JSD}(p, q) = I(p_i, q_i) + I(q_i, p_i) \]  

\[ \text{JSD}(p, q) = \frac{1}{2} \sum_i p_i \log \left( \frac{p_i}{M} \right) + \frac{1}{2} \sum_i q_i \log \left( \frac{q_i}{M} \right) \]  \hspace{1cm} (2)

Where \( I \) is the Kullback–Leibler divergence between two frequency distributions \( p_i \) and \( q_i \). In this application, \( p \) is the frequency distribution of the simulated patterns in a realisation (MP histogram of the realisation) and \( q \) is the frequency distribution of the simulated patterns in another realisation.

The square root of the Jensen–Shannon divergence is a real distance metric that is often referred to as the Jensen–Shannon distance (Endres & Schindelin 2003). Although the Jensen–Shannon divergence is a positive definite measure and is also symmetric, it does not satisfy the triangle inequality condition for a real metric. For this reason, the square root of the JSD is used rather than the JSD itself. Equation (2) is calculated as (Endres & Schindelin 2003):

\[ \text{JSD}(p, q) = \frac{1}{2} \sum_i p_i \log \left( \frac{p_i}{M} \right) + \frac{1}{2} \sum_i q_i \log \left( \frac{q_i}{M} \right) \]  \hspace{1cm} (3)

\[ M = \frac{p_i + q_i}{2} \]
\[ D(p, q) = \sqrt{JSD(p, q)} \quad (4) \]

Equation (4) is considered a useful distance formula in this research. A smaller distance represents a smaller space of uncertainty.

3.3.2 Measuring variability between training image and realisations

The algorithm comprises the following steps:

- Obtain some statistics of the training image and the realisations. In this case, the statistics are the frequency histogram of the observed patterns in the training image and the frequency histogram of simulated patterns of the realisation.

- Calculate Equation (4). In this case, \( p \) is the frequency distribution of the observed patterns in the training image (MP histogram of the training image) and \( q \) is the frequency distribution of the simulated patterns in the realisation (MP histogram of the realisation).
Chapter 4: Results

This chapter presents the results of the application of geostatistical simulation methods to karst systems or maze-like systems and the evaluation of simulation outputs using a distance measure. This chapter is divided into five major sections: (1) gridded MPS, (2) modified non-gridded MPS, (3) distance measures, (4) sensitivity of the non-gridded template size and (5) case study. A detailed explanation of the methods used to obtain the results is presented in each of these sections. The outputs have been obtained using algorithms implemented using MATLAB codes.

4.1 Gridded MPS

This study uses a recent gridded MPS method developed by Honarkhah and Caers (2010), called distance-based pattern modelling. A description of the method’s implementation and outputs is presented in the following sections.

4.1.1 Gridded MPS scanning algorithm

First, a simple 2D binary image representing a simplified channelised depositional system is used as a training image of size 101×101 (see Figure 4.1). This image is widely used as a demonstration example and as a means of assessing the performance of MPS methods. The first step is to select an optimal template. Although the size of the template should be small enough to reproduce small-scale features, it should also be large enough to reproduce actual structures occurring in the training
image. An optimum template size is determined using a single-point entropy measure—that is, a statistical measure of randomness—to describe the texture of the training image (Honarkhah & Caers 2010). The entropy of a pattern is calculated as:

\[ H = - \sum_{i=1}^{K} p_i \log(p_i) \]  \hspace{1cm} (4.1)

where \( K \) is equal to the number of possible outcomes of the random variables, and \( p_i \) represents the probability mass function (i.e., histogram).

The training image is first scanned using different sizes of the template. Then, for whole patterns, the mean entropy is calculated and plotted against the corresponding template size. From the start, the mean entropy generally increases. An optimum size of the template is obtained when the mean entropy reaches an approximately constant value. The entropy plot is illustrated in Figure 4.2.

---

*Figure 4.1. Conceptual model of a channel system 101×101×1 used as a gridded training image for the gridded MPS scanning analysis*
The mean entropy reaches an approximately constant value of 0.86 for a template size of 15×15, which is then taken as the optimum size of the template.

Once the optimal template is chosen, it is used to scan the training image to construct a pattern database. All patterns are stored in the pattern database and then mapped as a set of points in the multi-dimensional scaling (MDS) space using a distance-based method. That is, the patterns are represented as points in a multi-dimensional space in which the dissimilarity between the patterns is related to the distances of the corresponding points in that space. The dissimilarity between the patterns is calculated using a distance function; in this case, the Euclidean distance is used. Selecting the number of dimensions in the MDS mapping is a critical issue. To select the best dimension for the MDS space, a scree plot is generated from the eigenvalues of the covariance matrix (see Figure 4.3). The slope of the scree plot in Figure 4.3 changes significantly at a dimension value of around 12, which is chosen as the MDS dimension. The correlation coefficient is plotted against the dimension in Figure 4.3, from which it appears that the curve reaches a relatively fixed value at a dimension of around 20.
The MDS results are shown in Figure 4.4. With respect to the dissimilarities between the patterns, the points that are closer to each other in the Euclidean space correspond to similar patterns. The points in MDS space are mapped to a kernel space, and k-means clustering is applied to the dataset in the kernel space for pattern classification.
4.1.2 Gridded MPS simulation algorithm

Several realisations were generated by the gridded MPS simulation algorithm, two of which are shown in Figure 4.5. As shown, both images reproduce the complex spatial patterns in the training image. Further, the realisations represent significant connectivity as well as more variability in pattern reproduction. The MATLAB implementation of the algorithm generates a realisation in 3.78 seconds and 4.42 seconds respectively, which demonstrates the computational efficiency of the algorithm.
4.2 Modified Non-gridded MPS

The purpose here is to assess the performance of the non-gridded MPS method when applied to a channelised training image. From the literature search, it appears that such an assessment has never been done. One of the objectives of this research is to evaluate and modify the non-gridded MPS method to ensure its applicability to channelised training images.

The following sections summarise the modifications made to the non-gridded MPS simulation and the outputs obtained.

4.2.1 Modified non-gridded MPS pre-processing

1. Non-gridded training image

In this stage, a non-gridded training image is first built, consisting of points in the form of locations or XY coordinates of the networks. The
training image was first digitised using *Digitize it 2010* software to obtain a point-set training image. The same simple two-dimensional binary image representing the channel networks used in the gridded MPS simulation (see Figure 4.1) is used with the same resolution of 101×101×1 (see Figures 4.6a and 4.6b). Application of the same training image and the same resolution provides a fair comparison of the images reproduced by non-gridded and gridded MPS. The digitised model represents information of the channel locations along a central line of the channel. In this case, as the width of the passages is approximately constant, only the central line of the channel (not the thickness) is considered. A MATLAB code was written to plot the point-set training image shown in Figure 4.7. The point-sets represent the channel locations.

*Figure 4.6. (a) Conceptual model of a channel system with resolution 101×101×1, (b) digitised model used as a non-gridded training image for the modified non-gridded MPS scanning analysis (generated by Digitize it 2010)*
2. **Non-gridded template**

The procedure of template selection in modified non-gridded MPS is different from that used for non-gridded MPS, in which a generic template can be used to construct an optimal template. A non-gridded template adapted to each point-set training image must be selected. To do so, a template comprising a centre node and 37 nodes over a range of angles between 0° and 180° is used (as in Figure 3.9). This template includes the information shown in Table 4.1. The proposed template is used to capture MPS by scanning the training image.
### Table 4.1. Template properties

<table>
<thead>
<tr>
<th>#Node</th>
<th>Angle (°)</th>
<th>Angle Tolerance (°)</th>
<th>Lag Distance (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>35</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>40</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>45</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>50</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>55</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>60</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>65</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>70</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>75</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>80</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>85</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>90</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>95</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>100</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>105</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>110</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>115</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>120</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>26</td>
<td>125</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>27</td>
<td>130</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>28</td>
<td>135</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>29</td>
<td>140</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>30</td>
<td>145</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>31</td>
<td>150</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>32</td>
<td>155</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>33</td>
<td>160</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>34</td>
<td>165</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>35</td>
<td>170</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>36</td>
<td>175</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>37</td>
<td>180</td>
<td>2.5</td>
<td>2</td>
</tr>
</tbody>
</table>
4.2.2 Modified non-gridded MPS analysis

The proposed template is used to scan the point-set training image to capture MPS with the objective of determining the major pathway directions. The output of the scanning analysis is a pattern frequency histogram that summarises the numbers of patterns in different directions. Patterns with higher frequencies of occurrence will identify major channel directions and be applied in the next stage of MPS simulation.

The algorithm is implemented using a MATLAB code (see Appendix B). The output from the algorithm is shown in Figure 4.8; the X-axis is the ID of the observed patterns, and the Y-axis gives the corresponding frequency of the patterns.

![Histogram of observed Patterns in TI](image)

*Figure 4.8. Frequency histogram of the observed patterns as obtained by scanning the training image with the template*

Given the large number of observed patterns in Figure 4.8, it is visually difficult to recognise the patterns with the highest frequency. Nodes with angles 0° (node 1) and
180° (node 37) have the maximum number of patterns. Most patterns are within ranges 0° to 35° and 155° to 180°. Consequently, three templates are chosen for the next stage. The first template comprises two nodes with angles of 0° and 180°. The second template comprises eight nodes with angles between 0° and 35°. The third template comprises six nodes with angles between 155° and 180°. These three templates (see Tables 4.2–4.4) will be used to generate the simulations.

Table 4.2. First template properties

<table>
<thead>
<tr>
<th># Node</th>
<th>Angle (°)</th>
<th>Tolerance Angle</th>
<th>Distance</th>
<th>Tolerance Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>45</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>180</td>
<td>45</td>
<td>2</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 4.3. Second template properties

<table>
<thead>
<tr>
<th># Node</th>
<th>Angle (°)</th>
<th>Tolerance Angle</th>
<th>Distance</th>
<th>Tolerance Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>35</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 4.4. Third template properties

<table>
<thead>
<tr>
<th># Node</th>
<th>Angle (°)</th>
<th>Tolerance Angle</th>
<th>Distance</th>
<th>Tolerance Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>155</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>160</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>165</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>170</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>175</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>180</td>
<td>2.5</td>
<td>2</td>
<td>30</td>
</tr>
</tbody>
</table>
4.2.3 Modified non-gridded MPS simulation

The pattern simulation process is completed by scanning the training image and obtaining the frequency pattern histogram. In this case, as there are no specific conditioning data, a set is generated by sampling the training image. Conditioning data are chosen at random from the training image, with no more than 10% of the total pixels being selected.

Simulation starts by applying the first template, which represents the major trends of the channels. The process continues by successively applying each of the other templates. The templates are randomly applied at each step of the simulation to increase spatial variability. Two different realisations using modified non-gridded MPS are shown in Figure 4.9, and two different realisations using non-gridded MPS are shown in Figure 4.10. These realisations are generated using the same conditioning data.

Figure 4.9. Two realisations using the modified non-gridded MPS simulation algorithm; black points denote simulated nodes
As shown in Figure 4.9, the general trends of the channels have been reproduced, and the modified method has significantly improved the connectivity over the non-gridded MPS shown in Figure 4.10. Although it cannot be observed from two realisations, the random selection of the templates in modified non-gridded MPS has generated significantly more variability among the realisations than those generated by non-gridded MPS. This was an identified weakness in the method proposed by Erzeybek Balan (2012).

The modified method has clearly enhanced pattern reproduction in comparison with non-gridded MPS (see Figure 4.10). As observed in Figure 4.10, the realisations derived from the non-gridded MPS technique do not sufficiently reproduce the spatial patterns, especially in terms of connectivity. The modified non-gridded MPS realisations were generated in MATLAB in 5.42 and 6.23 seconds respectively. Therefore, it appears that the technique is reasonable in terms of computational cost.

Figure 4.10. Two realisations using the non-gridded MPS simulation algorithm; black points denote simulated nodes
4.3 Distance Measures

The purpose of the distance-based measure is to quantify the variability among the realisations generated by the simulation methods. In particular, it is used here to assess the relative performances of gridded MPS and modified non-gridded MPS. The process for doing so is as follows:

- Standardise the frequency distributions.
- Calculate Equation (4).

In Equation (1), \( I(p_i, q_i) \) (i.e., Kullback–Leibler divergence) is undefined if \( p_i = 0 \) and \( q_i \neq 0 \). This means that distribution \( p_i \) must be absolutely continuous with respect to distribution \( q_i \) for \( I(p_i, q_i) \) to be defined. Similarly, \( JSD(p_i, q_i) \) requires that \( p_i \) and \( q_i \) be absolutely continuous with respect to each other. This is one of the problems with these divergence measures (Lin 1991).

Two ways have been suggested to overcome these difficulties, although the results are similar for both:

1. A pseudo-count of 0.000001 is added to the frequencies to avoid zero in the numerator and/or denominator of the equation (Raes et al. 2011).
2. The values of \( \log(p/q) \) or \( \log(q/p) \) are set to zero if they are undefined.

In this study, the dissimilarity between the 50 realisations and the training image, and the dissimilarity among the realisations from both methods, have been calculated using this distance measure, which can compare the multiple-point histograms of the realisations and the training image. To provide valid comparisons of the images generated by modified non-gridded and gridded MPS, the same conditioning data are used in both methods. The results are presented in the following sections.
4.3.1 The within-realisation variability

The dissimilarities are calculated between the 50 realisations and the training image in both gridded MPS and modified non-gridded MPS using the Jensen–Shannon divergence method. To make for an easier comparison, the numerical results are presented as graphs, and the distances for both methods are plotted in Figures 4.11 and 4.12; the X-axis is the ID of the realisations, and the Y-axis gives the distance (i.e., dissimilarity) between the training image and each realisation.

The dissimilarity plots of the two methods have been drawn in one diagram (see Figure 4.13).

![Figure 4.11. Distances between training image and 50 realisations for gridded MPS; blue dots are the distance values between each realisation and the training image; red dots represent the average of the distances](image-url)
Figure 4.12. Distances between training image and 50 realisations for modified non-gridded MPS; blue dots are the distance values between each realisation and training image; red dots represent the average of the distances.

Figure 4.13. Distance plot between realisations and training image for gridded and modified non-gridded MPS methods; red dots represent the average of the distances.

From the graphs, it appears that there is more fluctuation in the modified non-gridded MPS values than in the gridded MPS values. Further, there is a larger range of
dissimilarity between the training image and the realisations. The distance range for gridded MPS is approximately from 0.03 to 0.07, with an average value of 0.048. For modified non-gridded MPS, it ranges from 0.05 to 0.16, with an average of 0.09. A larger value of the distance indicates a greater dissimilarity between the realisations and the training image.

4.3.2 The between-realisation variability

As well as the within-realisation variability, the dissimilarities between the 50 realisations (i.e., the between-realisation variability) in both gridded MPS and modified non-gridded MPS are obtained using the Jensen–Shannon divergence method. The distances for both methods are plotted in Figures 4.14 and 4.15; the X-axis is the ID of pairs of the realisations, and the Y-axis gives the distance (i.e., dissimilarity) between the pairs of the realisations.

Figure 4.14. Distances between 50 realisations in gridded MPS; blue dots are the distance values between pairs of realisations; red dots represent the average of the distances
Figure 4.15. Distances between 50 realisations from modified non-gridded MPS; blue dots are the distance value between pairs of realisations; red dots represent the average of the distances.

Figure 4.16. Distance plot between realisations for gridded and modified non-gridded MPS methods; red dots represent the average of the distances.

Figure 4.16 shows the dissimilarity between each pair of realisations from gridded MPS and modified non-gridded MPS. As shown, modified non-gridded MPS
generates a significantly larger space of uncertainty than gridded MPS, with a difference of about 0.035 in average distance.

4.3.3 Sensitivity of non-gridded template size

To evaluate the sensitivity of the non-gridded template size on pattern reproduction and uncertainty space, the distances are calculated for different template sizes in modified non-gridded MPS. The templates previously used as the optimal templates for the modified non-gridded MPS simulation in the channelised training image are used to build the templates; small and large size templates are obtained by changing the lag distance and tolerance window. Figures 4.17–4.20 show dissimilarity plots for different template parameters (e.g., the lag distance and tolerance window) in modified non-gridded MPS.

Figure 4.17. Distance plot for the modified non-gridded MPS with a small template size (i.e., distances between training image and 50 realisations using a template with a smaller tolerance window and lag distance)
Figure 4.18. Distance plot for the modified non-gridded MPS with a small template size (i.e., distances between 50 realisations using a template with a smaller tolerance window and lag distance).

Figure 4.19. Distance plot for the modified non-gridded MPS with a large template (i.e., distances between training image and 50 realisations using a template with a larger tolerance window and lag distance).
The dissimilarity between the 50 realisations and the training image using different template sizes (i.e., small, large and optimal) are shown in a single plot to enable a better comparison (see Figures 4.21 and 4.22).

Figure 4.20. Distance plot for the modified non-gridded MPS with a large template (i.e., distances between 50 realisations using a template with a larger tolerance window and lag distance)

Figure 4.21. Distance plot between realisations and training image for modified non-gridded MPS method using different template sizes
Generally, there appears to be a significant increase in the average distance between the training image and the realisations as the size of the template increases, whereas the distances among the pairs of realisations show a decrease in the space of uncertainty regardless of whether the template size increases or decreases. However, the distance between the realisations for the small template size presents almost the same values. The significant dissimilarity between the training image and the realisations for large size of the template, as well as a smaller space of uncertainty for both large and small template sizes, demonstrated that a template with a size of very small or large cannot give reasonable results that are consistent with reality. Moreover, given that using large templates might generate spurious connections, it is important to verify whether the captured patterns are consistent with the known geology. Spurious connections can be eliminated by using a template with a sufficiently tighter tolerance window. Although the size of the template should be small enough to reproduce small-scale features, it should also be large enough to reproduce actual structures occurring in the training image.
4.4 Case Study (Demonstration Examples)

Thus far, the methods have been applied to a meandering channel similar in geometry to a karst system. As this thesis does not have a case study with sparse data and a training image, some examples are created to assess the applicability of modified non-gridded MPS against gridded MPS. The methods are applied to two different karst systems, Olwolgin Cave and Tank Cave, respectively (see Figures 4.23 and 4.24). In each application, one cave is used as a training image and the other is used to select conditioning data in order to replicate what would happen in practice. Relatively few data would be available from the site to be simulated, and a training image with the general (assumed or inferred) characteristics of the site would be used.

![Figure 4.23. Olwolgin Cave downstream section](image)
4.4.1 Olwolgin Cave

Olwolgin Cave is located on the Nullarbor Plain near Roe Plains, Western Australia. It is a fragile cave with numerous interesting features, including rocks covered in black bacteria, hanging roots and green tannic water. Many of the passages are very tight. A total length of 2700 m of the cave passage has been surveyed. The maximum system depth is approximately 21 m. The system is completely water-filled and mapped by cave divers. Part of the map is depicted in Figure 4.23 (http://www.cavediving.net.au/index.php/olwolgin-cave).
4.4.1.1 Gridded MPS of Olwolgin Cave

To generate gridded MPS realisations of Olwolgin Cave, the map is first changed to a grey-scale image (see Figure 4.25) and then digitised at a resolution of 170×100×1 to give the training image shown in Figure 4.26.

An optimum template size is determined using the single-point entropy measure. The profile log-likelihood plot gives a maximum value of 21 (see Figure 4.27), which gives an optimum template size of 21×21.

Figure 4.25. Olwolgin Cave grey-scale intensity image

Figure 4.26. Olwolgin Cave digitised binary image at resolution of 170×100×1 (training image)
Figure 4.27. (a) Entropy plot; (b) profile log-likelihood; for optimum template size selection

Figure 4.28. Variance of entropy plot

In this case, Tank Cave is used to obtain the conditioning data applied for the simulation process. The conditioning data comprised approximately five percent of the Tank Cave data. The Tank Cave training image is first rotated to align it with the major network direction of Olwolgin Cave. Then, the rotated training image is digitised to provide the point training data. Conditioning data are sampled from these training data. Fifty realisations are generated by the gridded MPS simulation algorithm. Three of
these are shown in Figure 4.29, which reproduces the complex spatial patterns in the training image. Moreover, the realisations show connectivity similar to that of the training images, and there is reasonable variation in pattern reproduction among the simulated images.

![Training Image](image)

*Figure 4.29. Three realisations generated using gridded MPS simulation algorithm for Olwolgin Cave*

### 4.4.1.2 Modified non-gridded MPS for Olwolgin Cave

First, the Olwolgin Cave map is digitised at the same resolution as that of the gridded training image (i.e., 170×100×1) to build a non-gridded training image that is used for scanning analysis (see Figure 4.30). The digitised model represents passage locations. In this application, as the width of the passages varies, the central line of the passages is digitised, but in addition the width. The passages’ width is included by the
adjacent pixels of the central line of the passages. The modified non-gridded MPS simulation algorithm is then implemented by means of the same conditioning data used in gridded MPS (see Figure 4.31).

![Digitised Training Image (Olwalgin)](image)

**Figure 4.30.** Digitised model of Olwalgin Cave used as training data for the modified non-gridded MPS scanning algorithm; arrow indicates north

![Conditioning Data](image)

**Figure 4.31.** Conditioning data used for the modified non-gridded MPS simulation; arrow indicates north
Fifty simulated realisations of Olwolgin Cave are generated, two of which are shown in Figure 4.32.

![Figure 4.32. Two different realisations of Olwolgin Cave generated by modified non-gridded MPS](image)

The realisations show the same degree of connectivity as well as depicting the changes in the width of passages.

### 4.4.2 Tank Cave

Tank Cave is located in the Tantanoola area of lower south-eastern South Australia. Tank Cave is an extensive, maze-like phreatic system with more than 7 km of mostly underwater passages, which makes it one of the longest underwater caves in Australia. Spatial analysis of Tank Cave shows that passages are manifestly oriented along a trend aligned at 320° (see Figure 4.25), which reflects the regional joint structure (Webb et al. 2010). The maximum depth of the system is approximately 22 m. The system is completely water-filled and mapped by cave divers.
4.4.2.1 Gridded MPS of Tank Cave

A binary digital image at a resolution of $100 \times 100 \times 1$ is first built from the original Tank Cave map and then used as a training image (see Figure 4.33). An optimum template size of $23 \times 23$ is determined using the single-point entropy measure (see Figure 4.34).

![Tank Cave binary image at resolution of 100×100×1 (training image)](image1.png)

*Figure 4.33. Tank Cave binary image at resolution of 100×100×1 (training image)*

![Entropy plot; (b) profile log-likelihood; presenting the application of optimum template size selection](image2.png)

*Figure 4.34. (a) Entropy plot; (b) profile log-likelihood; presenting the application of optimum template size selection*
In the case of Tank Cave, conditioning data are obtained from Olwolgin Cave in the same manner as applied in the Olwolgin Cave case and afterwards used for the gridded MPS simulation process. Fifty realisations of Tank Cave are generated using the gridded MPS simulation algorithm. As shown in Figure 4.36, the realisations reproduce the connectivity and spatial variability of the training image.
4.4.2.2 Modified non-gridded MPS for Tank Cave

A non-gridded training image is built by digitising the Tank Cave map at a resolution of 100×100×1. In the Tank Cave application, given the variation in the width of the passages the central line of the passages is not digitised, and width is included by the resolution of the pixels used in digitisation. Thus, for example, for wider passages, the pixels adjacent to the central line are included so as to capture the passage width. This point-set training image is then used for the non-gridded scanning analysis (see

Figure 4.36. Three realisations resulted using gridded MPS simulation algorithm for Tank Cave
Figure 4.37). The same conditioning data used in gridded MPS of Tank Cave are applied for the implementation of the modified non-gridded MPS simulation algorithm (see Figure 4.38).

Figure 4.37. Digitised model of Tank Cave used as training data for the modified non-gridded MPS scanning algorithm; arrow indicates north

Figure 4.38. Conditioning data used for the modified non-gridded MPS simulation algorithm; arrow indicates north
Fifty simulated realisations of Tank Cave are generated, two of which are shown in Figure 4.39.

Figure 4.39. Two different realisations of Tank Cave generated by modified non-gridded MPS

The realisations reproduce the connectivity and spatial variability, as well as the varying widths of the passages.

4.4.3 Distance measures

The dissimilarities between the training image and the 50 realisations in both gridded MPS and modified non-gridded MPS are calculated using the Jensen–Shannon divergence method for both Olwolgin Cave and Tank Cave. Similarly, the dissimilarities between the 50 realisations are obtained for both cave systems. All realisations are generated by conditional simulation. The outcomes are depicted as graphs and plotted in figures. The dissimilarity plots of the two methods are presented in one diagram for each cave (see Figures 4.48–4.51).
Figure 4.40. Distances between training image and 50 realisations for gridded MPS for Olwolgin Cave; blue dots are the distance values between each realisation and the training image; red dots represent the average of the distances.

Figure 4.41. Distances between training image and 50 realisations for modified non-gridded MPS for Olwolgin Cave; blue dots are the distance values between each realisation and the training image; red dots represent the average of the distances.
Figure 4.42. Distances between 50 realisations for gridded MPS for Olwolgin Cave; blue dots are the distance values between pairs of realisations; red dots represent the average of the distances.

Figure 4.43. Distances between 50 realisations for modified non-gridded MPS for Olwolgin Cave; blue dots are the distance values between pairs of realisations; red dots represent the average of the distances.
Figure 4.4. Distances between training image and 50 realisations for gridded MPS for Tank Cave; blue dots are the distance values between each realisation and the training image; red dots represent the average of the distances.

Figure 4.45. Distances between training image and 50 realisations for modified non-gridded MPS for Tank Cave; blue dots are the distance values between each realisation and the training image; red dots represent the average of the distances.
Figure 4.46. Distances between 50 realisations for gridded MPS for Tank Cave; blue dots are the distance values between pairs of realisations; red dots represent the average of the distances.

Figure 4.47. Distances between 50 realisations for modified non-gridded MPS for Tank Cave; blue dots are the distance values between pairs of realisations; red dots represent the average of the distances.
Figure 4.48. Distance plot between 50 realisations and training image for gridded and modified non-gridded MPS methods for Olwolgin Cave

Figure 4.49. Distance plot between 50 realisations for gridded and modified non-gridded MPS methods for Olwolgin Cave
Figure 4.50. Distance plot between 50 realisations and training image for gridded and modified non-gridded MPS methods for Tank Cave

Figure 4.51. Distance plot between realisations for gridded and modified non-gridded MPS methods for Tank Cave
The Olwolgin Cave and Tank Cave simulations confirm the findings for the comparison between gridded MPS and modified non-gridded MPS on the widely used meandering channel demonstration example.

Clearly, modified non-gridded MPS generates a relatively larger space of uncertainty than gridded MPS, especially in the case of Tank Cave. Moreover, there is considerably more fluctuation in modified non-gridded MPS values than in gridded MPS values. That is, there is a larger range of dissimilarity between the training image and the realisations. In general, the graphs demonstrate better performance of modified non-gridded MPS in terms of spatial uncertainty and pattern reproduction when applied to a complex karst system; however, this technique is only applicable when a point dataset is available.
Chapter 5: Conclusions and Recommendations for Future Work

5.1 Conclusion

In summary, the modified non-gridded MPS method provides an improved approach to the characterisation and modelling of complex networks such as karst systems. The method, as applied in this research, uses a much more adaptable template to infer the multiple statistics from point-set data, and it eliminates gridding that might not be suitable for complex networks. Gridded MPS and modified non-gridded MPS have been evaluated in terms of pattern reproduction and the extent of the uncertainty space provide. The methods used for comparing the performances of gridded MPS and modified non-gridded MPS provide a sound basis choosing a methodology for any given application.

The major conclusions drawn from this research are:

- Gridded MPS was applied to a demonstration example of a channelised depositional system. The complex spatial patterns in the training image were captured using a 15×15 template. The simulated realisations adequately reproduced the connectivity and spatial variability of the training image.

- Non-gridded MPS was applied to the same channelised system. The simulated realisations clearly showed that the method cannot reproduce the level of connectivity of the training image. In particular, at small scales, there were many obvious discontinuities in the simulated images that conflicted with the reality of the training image. In addition, large-scale spatial features were not captured to any extent. A further difficulty was that the method only considered
point locations and not widths of structures at those locations. This was of no consequence for the simple demonstration example in which the width of the channelised system was constant. However, it was a problem for the general case of varying widths. From a general simulation point of view, the most serious issue was the lack of sufficient variability among the realisations generated by the method—that is, the uncertainty space represented by a set of realisations was significantly smaller than would be expected.

- The major reason for using non-gridded MPS methods was to enhance pattern reproduction for complex patterns that could not be adequately sampled (and represented) by regular grids—that is, templates comprising a regular array of grid rectangles could not adequately capture complex patterns. Spatially flexible non-gridded templates were used to obtain MPS, and network orientation and data configurations were transferred by pattern histograms. The problems identified with the current non-gridded MPS were addressed by modifying the method. In the modified non-gridded MPS, the optimal template was constructed from a variety of possible directions of the network, whereas the original non-gridded MPS only used a limited number of major directions based on a generic template.

- Modified non-gridded MPS implicitly included the width of structures (e.g., channels or caves) by specifying the appropriate resolution of the pixels used in digitisation.

- Modified non-gridded MPS derives network characteristics by analysing point-set training images using spatially flexible templates. The requirement was to construct templates that were capable of capturing the overall (large-scale) network direction and the (local or small-scale) connectivity. As with all MPS
approaches, some level of stationarity must be assumed to be able to use a template to extract MPS from the training image.

- MP-histograms were obtained by scanning the point-set training image. The resultant pattern histograms transferred the network orientation information and the connectivity information.

- Pattern simulation of the network was conditioned to sparse data in the form of locations or coordinates of the data. The performance of the modified non-gridded MPS was first tested on the demonstration example of a channelised training image. The simulated realisations demonstrated that the general trends of the channels had been reproduced, and the modified method had significantly improved the connectivity over that generated by the original non-gridded method. In addition, the modified non-gridded MPS realisations had significantly higher variability than those generated by the non-gridded MPS. More generally, the modified method had clearly enhanced the pattern reproduction in comparison with non-gridded MPS.

- Instead of relying on only a visual inspection of the realisations, a distance-based measure—the Jensen–Shannon divergence method—was used to quantify the variability among the realisations. In this study, the dissimilarity between 50 realisations and the training image, and among the realisations from both methods, i.e., gridded and modified non-gridded MPS, were calculated using this distance measure. The measure compared the multiple-point histograms of the realisations and the training image. The same conditioning data were used in both methods to make valid comparisons of the images generated by modified non-gridded and gridded MPS. The numerical results were presented as graphs.
The graphs showed that there was more variation in the modified non-gridded MPS realisations than in the gridded MPS realisations, which meant that there was a larger range of dissimilarity between the training image and the realisations. A larger value of the distance demonstrated a greater dissimilarity between the realisation and the training image. In terms of spatial uncertainty, the modified non-gridded MPS generated a relatively larger space of uncertainty than the gridded MPS.

The sensitivity of the size of the non-gridded template was assessed by using different templates with various configurations, and by varying the tolerance window used in the pattern simulation. For a clearer demonstration, the distances were calculated for different template parameters (e.g., the lag distance and tolerance window in modified non-gridded MPS) and then plotted for different template sizes (i.e., small, large and optimal size of the template).

In general, there was a significant increase in the average distance between the training image and the realisations as the size of the template increased. The distance values between the pairs of the realisations showed a decrease in the space of uncertainty, regardless of whether the template size increased or decreased. However, the distance between the realisations for the small template size presented almost the same values with a small average distance. The significant dissimilarity between the training image and the realisations for large size of the template, as well as a smaller space of uncertainty for both large and small template sizes, demonstrated that a template with a size of very small or large could not give reasonable results that were consistent with reality. Moreover, given that using large templates might generate spurious connections, it was important to verify whether the captured patterns were consistent with the
known geology. Using a template with a sufficiently tighter tolerance window could eliminate spurious connections. Although the size of the template should be small enough to reproduce small-scale features, it should also be large enough to reproduce actual structures occurring in the training image.

- Generally, using different templates resulted in simulated networks with different passage configurations. Therefore, it was important to assign templates that represented the target network structure.

- An application of gridded MPS and modified non-gridded MPS was presented for two different karst systems: Olwolgin Cave and Tank Cave. In each application, one cave was considered a training image and the other was used to obtain conditioning data.

- For Olwolgin Cave, an optimum gridded template size of 21×21 was obtained using the single-point entropy measure. Tank Cave was used to obtain the conditioning data. The gridded MPS simulation algorithm was then implemented, and 50 realisations were generated that adequately reproduced the complex spatial patterns in the training image. The simulated images showed connectivity similar to that of the training images; further, there was reasonable variation in pattern reproduction among the simulated images.

- Modified non-gridded MPS was applied to Olwolgin Cave using the non-gridded training image. In this case, as the width of the passages varied, width was included by the resolution of the pixels used in digitisation. Pattern simulation of the network was implemented with the same conditioning data used in gridded MPS. The 50 realisations that were generated displayed the same degree of connectivity as the training image, as well as the changing width of the passages.
• The gridded MPS simulation was implemented for Tank Cave using an optimum template size of 23×23 and conditioning data obtained from Olwolgin Cave, and 50 realisations are generated. The simulated realisations reproduced the connectivity and spatial variability of the training image.

• Modified non-gridded MPS was implemented for Tank Cave using the non-gridded point-set training image. The same conditioning data used in the gridded MPS of Tank Cave were used in the modified non-gridded MPS simulation algorithm, and 50 realisations were generated. The simulated images reproduced the connectivity and spatial variability of the training image, as well as the varying widths of the passages.

• The distance-based measure was used to compare the realisations generated by the two methods applied to the two different karst systems, and the outcomes were plotted.

• In general, the graphs obtained from both methods on both cave systems (i.e., Olwolgin Cave and Tank Cave) displayed similar results. There was significantly more variation in the modified non-gridded MPS realisations than in the gridded MPS realisations, especially for Tank Cave, whereas there was a larger range of dissimilarity between the training image and the realisations. Consequently, the modified non-gridded MPS realisations had more spatial variability, and they generated a relatively larger space of uncertainty than gridded MPS. In general, the graphs demonstrated better performance of modified non-gridded MPS in terms of spatial uncertainty and pattern reproduction when applied to a complex karst system.

• In conclusion, the proposed algorithm yielded accurate descriptions of karst networks. The modified non-gridded MPS technique can be applied to any kind
of network data that are defined by point-sets. Pattern simulation using performance statistics is a good practical tool for modelling and simulating the spatial distribution of karst networks. Moreover, this technique eliminates the gridding procedure, which may not be able to adequately capture the variability and connectivity of complex networks such as karst systems. More generally, the modified non-gridded algorithm adds value to reservoir simulation models by providing an accurate prediction of spatial distributions of the structures. Therefore, complex reservoirs such as karst systems are better represented in reservoir engineering studies, and more reliable results in future predictions can be obtained.

5.2 Recommendations for Future Work

Suggestions for further applications of the improved algorithms are:

- For complex aquifers, hydrogeological models depend critically on the connectivity of high- and low-permeability features. The modified non-gridded MPS method has significant potential in the spatial quantification of connectivity for these models, and for models of flow and ingression of contaminants (e.g., seawater intrusion into coastal aquifers). The method is not limited to karst systems; it could be applied in other areas of hydrogeology.

- Some recent studies (e.g., Mohammadmoradi & Rasaei 2014; Honarkhah & Caers 2012) have introduced non-stationarity in the conventional grid-based MPS method. Thus, further research could adapt the modified non-gridded MPS method to include non-stationary fields. In principle, the adaptive template used
in modified non-gridded MPS should make it more amenable to capturing and dealing with non-stationarity.

- This thesis was limited to quasi 2D applications, and much of the general published applications of MPS are 2D. One of the most important challenges for multiple-point geostatistics is in 3D applications, which require either the construction of 3D training images or generating 3D multiple-point simulations with 2D training images. This remains an interesting research topic for further research. Such 3D simulations, if theoretically feasible, would incur large computational costs (in hardware and time), and research into feasible computation would also be an interesting research area.
Appendices

Modified non-gridded MPS algorithm is developed in MATLAB. A detailed explanation of the methods used to obtain the results is presented in Chapters 3 and 4.

Appendix A: Gridded MPS Simulation Code

```matlab
%% Distance-based Pattern Modelling Method on Channelised training image
clear all
close all
clc;
m = 1;
% find the root path of the method's files
fprintf('Adding Paths to Matlab ......................
');
fp = fullfile('fullpath');
dirName = fileparts(fp);
slash = strfind(dirName, '\');
dirName = dirName(1:slash(end)-1);
% add the paths to Matlab
addpath(fullfile(dirName, 'dist_pat'));
addpath(fullfile(dirName, 'Training Images'));
addpath(fullfile(dirName, 'HardData'));
addpath(fullfile(dirName, 'Slice 3D'));
addpath(fullfile(dirName, 'scmdscale'));
addpath(fullfile(dirName, 'kdtree'));
addpath(fullfile(dirName, 'Fast Kmeans'));
addpath(fullfile(dirName, 'sltoolbox'));
install_paths(fullfile(dirName, 'sltoolbox'));
adapt(fullfile(dirName, 'stprtool_small'));
adapts(fullfile(dirName, 'stprtool_small\preimage'));
fprintf(' Done!n');
% ---------------------------------------------------------------
% Parameter Initialization
% ---------------------------------------------------------------
% possible choices for Training Image
% ---------------------------------------------------------------
% channel.dat 101x101

tiName = 'channel.dat';
par.Dimension = 101;
par.Dimensionz = 1;
par.Pat = 15;
par.Patz = 1;
par.innerPatch = 11;
par.innerPatchz = 1;
par.multipleGrid = 2;
par.m = 4;
par.way = 8;
par.MDS = 12;
par.clus = 100;
loop = par.multipleGrid;
```
% Read Training Image
fprintf('Loading the Training Image ....................');
file_location = [dirName, '\Training Images\', tiName];
[datain, colnames, line1] = loadgeoeas(file_location);
out = geoeas2matlab(datain, [par.Dim par.Dim par.Dimz]);
fprintf(' Done!\n');

clusterModel = []; Z = []; radonX = [];
patternIdx = struct;

% Change the resolution for initial multiple-resolution setting
if par.newMG
    out_c = out;
    m1 = 1;
    out = imresize(out_c, 1/loop);
    level = graythresh(out);
    out = im2bw(out, level);
    par.Dim = size(out, 1);
    par.multipleGrid = 1;
end

%% Multi Simulations
nsim = 2;
xmph = 4;
zmph = 1;

dd = 

hists = zeros(1+bin2dec(num2str(ones(1,xmph^2*zmph))), nsim);
for ii = 1:nsim
    close all;
    % construct the realization grid
    % -----------------------------------------------
    par.szRealization = par.Dim + (par.Pat - 1)*2^(par.multipleGrid-1);
    par.szRealizationz = par.Dimz + (par.Patz - 1)*2^(par.multipleGrid-1);
    realization = 0.5*ones(par.szRealization, par.szRealizationz);
    % -----------------------------------------------
    % assigning hard data
    % --------------------------------------------------
    hardData = NaN(size(realization));
    if par.hardData
        hardData = readHardData('hardData_Channel.dat', hardData, par);
    end
    % For each Coarse Grid Do:
    % -----------------------------------------------
    for m1 = loop:-1:1
        % change resolutions in multiple-resolution option
        if par.newMG
            out = imresize(out_c, 1/m1);
            par.Dim = size(out, 1);
            level = graythresh(out);
            out = im2bw(out, level);
        if m1 ~= loop
            realization = imresize(realization, (m1+1)/m1);
        end
    end
end
realization = +realization;
level = graythresh(realization);
realization = im2bw(realization, level);
par.szRealization = par.Dim + (par.Pat - 1)*2^(par.multipleGrid-1);
par.szRealizationz = par.Dimz + (par.Patz - 1)*2^(par.multipleGrid-1);
didf = (size(realization,1) - par.szRealization)/2;
realization = realization(didf+1:end-didf,didf+1:end-didf);
end
m1 = 1;
end
par.m1 = 2^(m1-1);

% Store the frozen nodes in each coarse simulation
frozenRealiz = zeros(par.szRealization, par.szRealization, par.szRealizationz);

if par.bLoadVariables
    fprintf('

Loading variables from savedVar%d.mat ....
','par.m1);
load([dirName '\dist_pat\savedVar' num2str(par.m1) 'mat'],'par.m1),'
    fprintf('Done!

','spaceString,spaceString,spaceString);
else
% Classify Patterns by dissimilarity Matrix and MDS and & Kernel
[X, Y, K, idx, prototype, sigma, par.MDS] = classifyPatterns(out,par);

    % [clusterModel, Z] = initializeKernelModel(Y, K, sigma, par);

    clusterIdx = cell(par.clus,1);
    clusterIdx = fillClusterIndex(clusterIdx, idx);
    % radonX = calculateRadonX(X, par.Pat); %not
adapted to 3D case
    fprintf('s%s
','spaceString,spaceString,spaceString);
end

if par.bSaveVariables
    save([dirName '\dist_pat\savedVar' num2str(par.m1) 'mat'],'X', 'Y', 'K', 'idx', 'prototype', 'sigma', 'clusterModel', 'Z', 'clusterIdx', 'radonX');
end
% Define a random path through the grid nodes
par.szCoarseGrid = fix((par.Dim - 1)/par.m1)+1;
par.szCoarseGridz = fix((par.Dimz - 1)/par.m1)+1;
lengthRandomPath = par.szCoarseGrid^2*par.szCoarseGridz;
randomPath = randperm(lengthRandomPath);

% Change to subscripts
[nodeI, nodeJ, nodeK] = ind2sub([par.szCoarseGrid, par.szCoarseGrid, par.szCoarseGridz],
randomPath);
node = vertcat(nodeI, nodeJ, nodeK);
[wx, wy, wz] = getPatternShape(node, par);

% fix the locations of hardData
hardDataMoved = moveHardData(hardData, par);
if par.m1 == 2, hardDataMoved(67, 65) = 0; hardDataMoved(69, 65) = 1; end
realization (~isnan(hardDataMoved)) = hardDataMoved(~isnan(hardDataMoved));

frozenRealiz(~isnan(hardDataMoved)) = 1;
% Perform simulation
% for each node in the random path Do:
% -------------------------------------------------------------
for i = 1:lengthRandomPath

% fprintf('node: %5d Percentage Completed: %3.0f%%', i, 100*i/lengthRandomPath);

if par.bSkipPreviousFrozenNodes
    if frozenRealiz(wx(i, (par.Pat+1)/2), wy(i, (par.Pat+1)/2), wz(i, (par.Patz+1)/2)) == 1
        continue
    end
end
[dataEvent, status] = getDataEvent(realization, wx(i,:), wy(i,:), wz(i,:));
if par.hardData
    weightEvent = findWeight(hardDataMoved, frozenRealiz, wx(i,:), wy(i,:), wz(i,:));
else
    weightEvent = ones(1,par.Pat^2*par.Patz);
end
% Check if there is any data conditioning event or not and find the
% pattern to be pasted on the simulation grid
rand('twister', sum(100*clock));
switch status
    case 'empty'
        randIdx = ceil(size(X,1).*rand(1,1));
        if par.bUseDualTemplate
            patternIdx = findRangeDualTemplate(randIdx, par);
        end
        Pattern = X(randIdx,:);
    case 'some'
        idxNumber = findClosestCluster(dataEvent, prototype, weightEvent);
        [Pattern, patternIdx] = findClosestInCluster(dataEvent, X, clusterIdx{idxNumber}, par, radonX, weightEvent);
%        idxNumber = findClosestCluster(dataEvent, prototype);
%        [Pattern, cluster] = createPattern(idxNumber, clusterModel, Z, X, Y, idx, Pat, cluster, radonX);
%        Pattern = reshape(Pattern, 1, Pat^2);

    end
end
case 'full'
    if existNonFrozenNodes(frozenRealiz, wx(i,:), wy(i,:), wz(i,:))
        idxNumber = findClosestCluster(dataEvent, prototype, wieghtEvent);
        [Pattern, patternIdx] = findClosestInCluster(dataEvent, X, clusterIdx{idxNumber}, par, radonX, wieghtEvent);
    else
        continue
    end
% Paste the pattern on simulation grid and updates frozen nodes
    if par.bUseDualTemplate
        [realization, frozenRealiz] = pastePattern(Pattern, wx(i,:), wy(i,:), wz(i,:), realization, frozenRealiz, par, out(patternIdx.x,patternIdx.y, patternIdx.z));
    else
        [realization, frozenRealiz] = pastePattern(Pattern, wx(i,:), wy(i,:), wz(i,:), realization, frozenRealiz, par, []);
    end
end
% show results at each multiple-grid
if par.bShowMultiGrids
    if par.Dimz > 1 %3D case
        slice3d(realization((par.szRealization-par.Dim)/2+1:(par.szRealization-par.Dim)/2+par.Dim,(par.szRealization-par.Dim)/2+1:(par.szRealization-par.Dim)/2+par.Dim));
        pause(0.1);
    else
        colormap(gray); %2D
        imagesc(realization((par.szRealization-par.Dim)/2+1:(par.szRealization-par.Dim)/2+par.Dim,(par.szRealization-par.Dim)/2+1:(par.szRealization-par.Dim)/2+par.Dim));
        axis square off;
        drawnow expose;
        pause(0.1);
    end
end
% to do TRANSCAT in the penultimate multigrids
if par.bTransCat || par.bTransCon
    if par.m1 == 2
        limits = (par.szRealization -par.Dim )/2;
        limitsz = (par.szRealizationz -par.Dimz)/2;
        if ~par.bUseDualTemplate
            real = realization(limits+1:limits+par.Dim , limits+1:limits+par.Dim , limitsz+1:limitsz+par.Dimz);
            harddata = hardData (limits+1:limits+par.Dim , limits+1:limits+par.Dim , limitsz+1:limitsz+par.Dimz);
        end
        if par.bTransCon
            real = histeq(real(1:par.m1:end,1:par.m1:end,1:par.m1:end),hist(out(:),par.Dim^2*par.Dimz));
        else
            real = transcat(real(1:par.m1:end,1:par.m1:end,1:par.m1:end), out, harddata(1:par.m1:end,1:par.m1:end,1:par.m1:end), par, 1);
        end
realization(limits+1:par.m1:limits+par.Dim , limits+1:par.m1:limits+par.Dim , limitsz+1:par.m1:limitsz+par.Dimz) = real;
else
real = realization(limits+1:limits+par.Dim , limits+1:limits+par.Dim , limitsz+1:limitsz+par.Dimz); harddata = hardData (limits+1:limits+par.Dim , limits+1:limits+par.Dim , limitsz+1:limitsz+par.Dimz);
if par.bTransCon
real = histeq(real,hist(out(:),par.Dim^2*par.Dimz));
else
real = transcat(real, out, harddata, par, 1);
end
realization(limits+1:limits+par.Dim , limits+1:limits+par.Dim , limitsz+1:limitsz+par.Dimz) = real;
else
realization(limits+1:limits+par.Dim , limits+1:limits+par.Dim , limitsz+1:limitsz+par.Dimz) = real;
end
end
end
end
% crop the realization to its true dimensions
limits = (par.szRealization - par.Dim )/2;
limitsz = (par.szRealizationz - par.Dimz)/2;
realization = realization(limits+1:limits+par.Dim , limits+1:limits+par.Dim , limitsz+1:limitsz+par.Dimz);
% hardData = hardData (limits+1:limits+par.Dim , limits+1:limits+par.Dim , limitsz+1:limitsz+par.Dimz);
print('-r200','-dpng',sprintf('Outputs\Images\sim_%0.0f.png',ii));
title(sprintf('realisation %0.0f',ii));
disp('>> figure saved.');</Democrats
%% TRANSCAT (Transformation of categorical proportions of realization to TI proportions)
if par.bTransCat
realization_C = transcat(realization, out, hardData, par, 1);
% show the transformed version of the realization too
figure; imagesc(realization_C);
axis square; axis xy off; colormap gray;
end
fprintf('\n\nFinish!\n');
%% multiple point histogram of TI variable
for m = 1
xmph = 4;
zmph = 1;
par.Dim = size(out,1);
par.Dimz = size(out,3);
disDim  = par.Dim - xmph +1;
disDimz = par.Dimz - zmph +1;
hist1 = zeros(1,1+bin2dec(num2str(ones(1,xmph^2*zmph))));
for i=1:disDim
for j=1:disDim
for k=1:disDimz
temp = 1+bin2dec(num2str(reshape(out(i:i+xmph-1,j:j+zmph-1,k:k+zmph-1),1,[])));
hist1(temp) = hist1(temp) + 1;
end
end
end
%% MP histogram of Re variable
for m = 1

xmph = 4;
zmph = 1;
par.Dimre = size(realization,1);
par.Dimzre = size(realization,3);
disDimre = par.Dimre-xmph +1;
disDimzre = par.Dimzre-zmph +1;
hist11 = zeros(1,1+bin2dec(num2str(ones(1,xmph^2*zmph))));
for i=1:disDimre
    for j=1:disDimre
        for k=1:disDimzre
            temp = 1+bin2dec(num2str(reshape(realization(i:i+xmph-1,j:j+xmph-1,k:k+zmph-1),1,1)));
            hist11(temp) = hist11(temp) + 1;
        end
    end
end
histRe = hist11(:)/sum(hist11(:));
hists(:,ii) = histRe;
end
end
%% Save MPH of Ti and Re
filename = 'Outputs\MPH_Ti.xlsx';
xlswrite(filename,hist1)    %save distance table in excel format
filename = 'Outputs\MPH_Realisations.xlsx';
xlswrite(filename,hists)    %save distance table in excel format
end

Appendix B: Modified Non-gridded MPS Simulation Code

%% Modified non-gridded MPS Method on Channelised training image

close all
clear all
clc;
m = 1;
%% Pre-processing
for m = 1
    figure(1);
    pts = xlsread('Channel_data.xlsx');
    pts = Scale2(pts);
    npt = size(pts,1);
    plot(pts(:,1),pts(:,2),'b.');
    axis image; axis([0 101 0 101])
end

%% Parameters
for m = 1
    nsim = 2;
atol = (2.5)*(pi/180);
distance = 1;
distances1 = 1;
distances1ti = distances1;
dtols1 = 15;
dtols1ti = 2;
dtols10 = 30;
nnodes1 = 10;
nnodes2 = 10;
nnodesti = 37;
anglesti = [0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180]*(pi/180);
angles2 = [0 5 10 15 20 25 30 35 40 45]*(pi/180);
angles1 = [135 140 145 150 155 160 165 170 175 180]*(pi/180);
npat1 = 2^nnodes1;
npat2 = 2^nnodes2;
Dist = [];
hists1 = zeros(npat1,nsim);
hists2 = zeros(npat2,nsim);
a = 500;
end

%% Scanning Analysis Ti
for m = 1
    %% template definition
    for m = 1
        xnodest1 = distances1ti.*cos(anglesti);
ynodest1 = distances1ti.*sin(anglesti);
cnt = [0,0];
    end
    %% Scanning TI
    for m = 1
        disp('Scanning TI ...');
mch1 = cell(npt,1);
nmch1 = [];
rndm1 = randperm(npt);
        for j=1:npt
            rn11 = rndm1(j);
            for i=1:nnodest1
                cnt = [pts(rn11,1),pts(rn11,2)];
xn = cnt(1)+xnodest1(i);
yn = cnt(2)+ynodest1(i);
                for k=1:npt
                    if j==k continue; end
                    dx = pts(k,1)-cnt(1);
dy = pts(k,2)-cnt(2);
dd = (dx^2+dy^2)^0.5;
if abs(dd-distances1ti)<=dtols1ti
            alpha = atan2(dy,dx);
if abs(anglesti(i)-alpha)<=atol
                nmch1 = [nmch1 ; [xn,yn]];
mch1{j} = [mch1{j};i];
            end
        end
    end
end

%% Calculate frequencies of matched patterns
for m = 1
    disp('Calculate frequencies of matched patterns ...');
npat1i = 2^nnodes1;
 freq1 = zeros(1,npat1i);
%number of possible patterns
    for j=1:npat1i
        if isempty(mch1{j})
            continue
        end
        nds1 = unique(mch1{j});
end
mask1 = uint32(0);
for i=1:length(nds1)
    mask1 = bitset(mask1,nds1(i));
end
freq1(mask1) = freq1(mask1)+1;
end
Hists1 = freq1(:)/sum(freq1(:));

%% Save MPH of Ti
for m = 1
    filename = 'new MPH Ti1.xlsx';
xlswrite(filename,Hists1) %save frequency histogram of Ti in excel format
end
cts = xlsread('cd.xlsx'); %conditioning data
cts = Scale2(cts);
ncd = size(cts,1); %number of conditioning data
disp(ncd);
figure(2);
plot(cts(:,1),cts(:,2),'r.');
axis image; axis([0 101 0 101])

%% Multi Simulation
for ii = 1:nsim
    disp(ii);
    close all
% Finding simulatable nodes 1
for m = 1
    disp('Determining Simulatable Nodes ...');
simpth = randperm(ncd);
pot1 = [];
potcts = [];
for w=1:ncd
    u = simpth(w);
nnodes = 2;
    angles = [0 -180]*(pi/180);
    atols = (80)*(pi/180);
    xnodes = distance.*cos(angles);
    ynodes = distance.*sin(angles);
    for i=1:nnodes
        cnt = [cts(u,1),cts(u,2)];
        xn = cnt(1)+xnodes(i);
        yn = cnt(2)+ynodes(i);
        for k=1:ncd
            if k==w; continue; end
            dx = cts(k,1)-cnt(1);
            dy = cts(k,2)-cnt(2);
            dd = (dx^2+dy^2)^0.5;
            if abs(dd-distance)<dtols10
                alpha = atan2(dy,dx);
                if abs(angles(i)-alpha)<atols
                    %pot = [pot;[xn,yn]];
                    potcts = [cts;pot1;[xn,yn]];
                    pot1 = [pot1;[xn,yn]];
                    %line = [line;[cnt,xn,yn]];
                    %plot([cnt(1),xn],[cnt(2),yn],'k-');
                    %plot(xn,yn,'k.');
                end
            end
        end
    end
end
end
end

% [~,idx] = unique(strcat(potcts(:,1),potcts(:,2)),'rows');
potcts = potcts(idx,:);
npotcts = length(potcts);

% [~,idx] = unique(strcat(pot1(:,1),pot1(:,2)),'rows');
pot1 = pot1(idx,:);
npot1 = length(pot1);
end

end

%% Finding simulatable nodes 2
for m = 1
    disp('Determining Simulatable Nodes ...');
    simpth = randperm(ncd);
    pot2 = [];
    potcts = [];
    for w=1:ncd
        u2 = simpth(w);
        nnodes = 2;
        angles = [0 -180]*(pi/180);
        atols = (80)*(pi/180);
        xnodes2 = distance.*cos(angles);
        ynodes2 = distance.*sin(angles);
        for i=1:nnodes
            cnt = [cts(u2,1),cts(u2,2)];
            xn = cnt(1)+xnodes2(i);
            yn = cnt(2)+ynodes2(i);
            for k=1:ncd
                if k==w; continue; end
                dx = cts(k,1)-cnt(1);
                dy = cts(k,2)-cnt(2);
                dd = (dx^2+dy^2)^0.5;
                if abs(dd-distance)<dtols10
                    alpha = atan2(dy,dx);
                    if abs(angles(i)-alpha)<atols
                        %pot = [pot;[xn,yn]];
                        potcts = [cts;pot2;[xn,yn]];
                        pot2 = [pot2;[xn,yn]];
                        %line = [line;[cnt,xn,yn]];
                        %plot([cnt(1),xn],[cnt(2),yn],'k-');
                        %plot(xn,yn,'k.');
                    end
                end
            end
        end
    end
end

% [~,idx] = unique(strcat(potcts(:,1),potcts(:,2)),'rows');
potcts = potcts(idx,:);
npotcts = length(potcts);

% [~,idx] = unique(strcat(pot2(:,1),pot2(:,2)),'rows');
pot2 = pot2(idx,:);
npot2 = length(pot2);

end

%% Simulation 1
for m = 1
    disp('To Simulate 1...');

end
jj = a;
mchsim1 = cell(jj,1);
jj = 0;
while jj<a
    jj = jj+1;
    cnt = [pot1(jj,1),pot1(jj,2)];
xnodes1 = distances1.*cos(angles1);
ynodes1 = distances1.*sin(angles1);
    for i=1:nnodes1
        xn = cnt(1)+xnodes1(i);
        yn = cnt(2)+ynodes1(i);
        for k=1:npotcts
            if jj==k; continue; end
            dx = potcts(k,1)-cnt(1);
            dy = potcts(k,2)-cnt(2);
            dd = (dx^2+dy^2)^0.5;
            if abs(dd-distances1)<dtols1
                alpha = atan2(dy,dx);
                if abs(angles1(i)-alpha)<atol
                    %pot = [pot(:,:,1),pot(:,:,2)];
                    potcts = [potcts(:,:,1),potcts(:,:,2)];
                    pot1 = [pot1(:,:,1),pot1(:,:,2)];
                    %line = [line(:,:,1),line(:,:,2)];
                    mchsim1(jj) = [mchsim1(jj);i];
                    %plot([cnt(1),xn],[cnt(2),yn],'k-');
                    %plot(xn,yn,'b.');
                end
            end
        end
    end
    if jj==size(pot1,1)
        break
    end
end
%%
[~,idx] = unique(strcat(pot1(:,1),pot1(:,2)),'rows');
pot1 = pot1(idx,:);
npot1 = length(pot1);
%%
[~,idx] = unique(strcat(potcts(:,1),potcts(:,2)),'rows');
potcts = potcts(idx,:);
npotcts = length(potcts);
end
%% Simulation 2
for m = 1
    disp('To Simulate 2...');
    jj2 = a;
mchsim2 = cell(jj2,1);
jj2 = 0;
while jj2<a
    jj2 = jj2+1;
    cnt = [pot2(jj2,1),pot2(jj2,2)];
xnodes2 = distances1.*cos(angles2);
ynodes2 = distances1.*sin(angles2);
    for i=1:nnodes2
        xn = cnt(1)+xnodes2(i);
        yn = cnt(2)+ynodes2(i);
        for k=1:npotcts
            if jj2==k; continue; end
            dx = potcts(k,1)-cnt(1);
dy = potcts(k,2)-cnt(2);
 dd = (dx^2+dy^2)^0.5;
 if abs(dd-distances1)<dtols1
   alpha = atan2(dy,dx);
   if abs(angles2(i)-alpha)<atol
     %pot = [pot;[xn,yn]];
     potcts = [potcts;[xn,yn]];
     pot2 = [pot2;[xn,yn]];
     %line = [line;[cnt,xn,yn]];
     mchsim2{jj2} = [mchaim2{jj2};i];
     %plot([cnt1,xn],[cnt2,yn],'k-');
     %plot(xn,yn,'b.');
   end
 end
 end

if jj2==size(pot2,1)
   break
end

%%
[~,idx] = unique(strcat(pot2(:,1),pot2(:,2)),'rows');
pot2 = pot2(idx,:);
npot2 = length(pot2);

[~,idx] = unique(strcat(potcts(:,1),potcts(:,2)),'rows');
potcts = potcts(idx,:);
npotcts = length(potcts);
end
%

print('-r200','-dpng',sprintf('Outputs\Images\Sim_%0.0f.png',ii));
end

%% Calculate frequencies of matched patterns 1
for m = 1
  disp('Calculate frequencies of matched patterns 1 ...');
  nppat1 = 2^nnodes1;
  freqRe1 = zeros(1,nppat1);
  for j=1:npotcts
    if isempty(mchsim1{j})
      continue
    end
    ndsRe1 = unique(mchsim1{j});
    maskRe1 = uint32(0);
    for i=1:length(ndsRe1)
      maskRe1 = bitset(maskRe1,ndsRe1(i));
    end
    freqRe1(maskRe1) = freqRe1(maskRe1)+1;
  end
  histRe1 = freqRe1(:)/sum(freqRe1(:));
  hists1(:,ii) = histRe1;
end

%% Calculate frequencies of matched patterns 2
for m = 1
  disp('Calculate frequencies of matched patterns 2 ...');
  nppat2 = 2^nnodes2;
  freqRe2 = zeros(1,nppat2);
  for j=1:npotcts
    if isempty(mchsim2{j})
      continue
    end
ndsRe2 = unique(mchsim2{j});
maskRe2 = uint32(0);
for i=1:length(ndsRe2)
    maskRe2 = bitset(maskRe2,ndsRe2(i));
end
freqRe2(maskRe2) = freqRe2(maskRe2)+1;
end
histRe2 = freqRe2(:,i)/sum(freqRe2(:,));
hists2(:,ii) = histRe2;

%% Save MPH of Realisations
for m = 1
    filename = 'Outputs\MPH_Realisations1.xlsx';
xlswrite(filename,hists1)
    filename = 'Outputs\MPH_Realisations2.xlsx';
xlswrite(filename,hists2)
end

Appendix C: Distance-based Measure Code

%% Calculating distance measure for gridded and modified non-gridded
MPS on Channelised training image

%% Calculating distances between 50 realisations & Ti
for m = 1
    Hists1 = xlsread('MPH_Ti_1.xlsx');
    Hists2 = xlsread('MPH_Ti_2.xlsx');
    hist1 = xlsread('MPH_Re_1.xlsx');
    hist2 = xlsread('MPH_Re_2.xlsx');
    DistTi = [];
    JSDTi = [];
    for i = 1:nsim
        pTi1 = Hists1;
        pTi2 = Hists2;
        qRe1 = hist1(:,i);
        qRe2 = hist2(:,i);
        %R = 0.5*(p+q);
        KLDpTi1 = log(pTi1./qRe1);
        KLDpTi2 = log(pTi2./qRe2);
        KLDqRe1 = log(qRe1./pTi1);
        KLDqRe2 = log(qRe2./pTi2);
        KLDpTi1(isnan(KLDpTi1)) = 0;
        KLDpTi2(isnan(KLDpTi2)) = 0;
        KLDqRe1(isnan(KLDqRe1)) = 0;
        KLDqRe2(isnan(KLDqRe2)) = 0;
        KLDpTi1(isinf(KLDpTi1)) = 0;
        KLDpTi2(isinf(KLDpTi2)) = 0;
        KLDqRe1(isinf(KLDqRe1)) = 0;
        KLDqRe2(isinf(KLDqRe2)) = 0;
        jsdT1 = 
            ((0.5).*sum(pTi1.*KLDpTi1)) + ((0.5).*sum(qRe1.*KLDqRe1));
        jsdT2 = 
            ((0.5).*sum(pTi2.*KLDpTi2)) + ((0.5).*sum(qRe2.*KLDqRe2));
        jsdT = (jsdT1+jsdT2)/2;
        jsdT = jsdT.^(0.5);
for m = 1
    filename = 'Outputs\jsd_between_Realisations_&_Ti.xlsx';
    xlsxwrite(filename,DistTi)
end

for m = 1
    X = (1:length(JSDTi));
    Y = JSDTi;
    figure(1);
    plot(X,Y,'.b');
    hold on
    plot(X,Y,'-b');
    Z = mean(Y);
    hold on
    plot(X,Z,'.r');
    ylabel('Distances');
    xlabel('Realisations');
end

%% Calculating distances between realisations
for m = 1
    Dist = [];
    JSD = [];
    for i = 1:nsim
        p1 = hists1(:,i);
        p2 = hists2(:,i);
        for j = (i+1):nsim
            q1 = hists1(:,j);
            q2 = hists2(:,j);
            if i==j
                djs1 = 0;
                djs2 = 0;
            else
                %R = 0.5*(p+q);
                KLDp1 = log(p1./q1);
                KLDp2 = log(p2./q2);
                KLDq1 = log(q1./p1);
                KLDq2 = log(q2./p2);
                KLDp1(isnan(KLDp1)) = 0;
                KLDp2(isnan(KLDp2)) = 0;
                KLDq1(isnan(KLDq1)) = 0;
                KLDq2(isnan(KLDq2)) = 0;
                KLDp1(isinf(KLDp1)) = 0;
                KLDp2(isinf(KLDp2)) = 0;
                KLDq1(isinf(KLDq1)) = 0;
                KLDq2(isinf(KLDq2)) = 0;
                jsd1 = ((0.5).*sum(p1.*KLDp1))+((0.5).*sum(q1.*KLDq1));
                jsd2 = ((0.5).*sum(p2.*KLDp2))+((0.5).*sum(q2.*KLDq2));
                jsd = (jsd1+jsd2)/2;
                jsd = jsd.^0.5;
                %disp([i,j,jsd]);
                JSD = [JSD;[i,j,jsd]];
                Dist = [Dist;[i,j,jsd]];
            end
        end
    end
end
end
end
end
end
end

%% Save distance table between Realisations
for m = 1
    filename = 'Outputs\jsd_between_50_Realisations.xlsx';
    xlswrite(filename,Dist)
end

%% Plot JSD between Realisations
for m = 1
    X1 = (1:length(JSD));
    Y1 = JSD;
    figure(2);
    x1 = (1:30:length(JSD));
    Z1 = sum(JSD)/(length(JSD));
    plot(x1,Z1,'.r');
    ylabel('Distances');
    xlabel('Realisations');
    hold on
    plot(X1,Y1,'.b');
end
References


