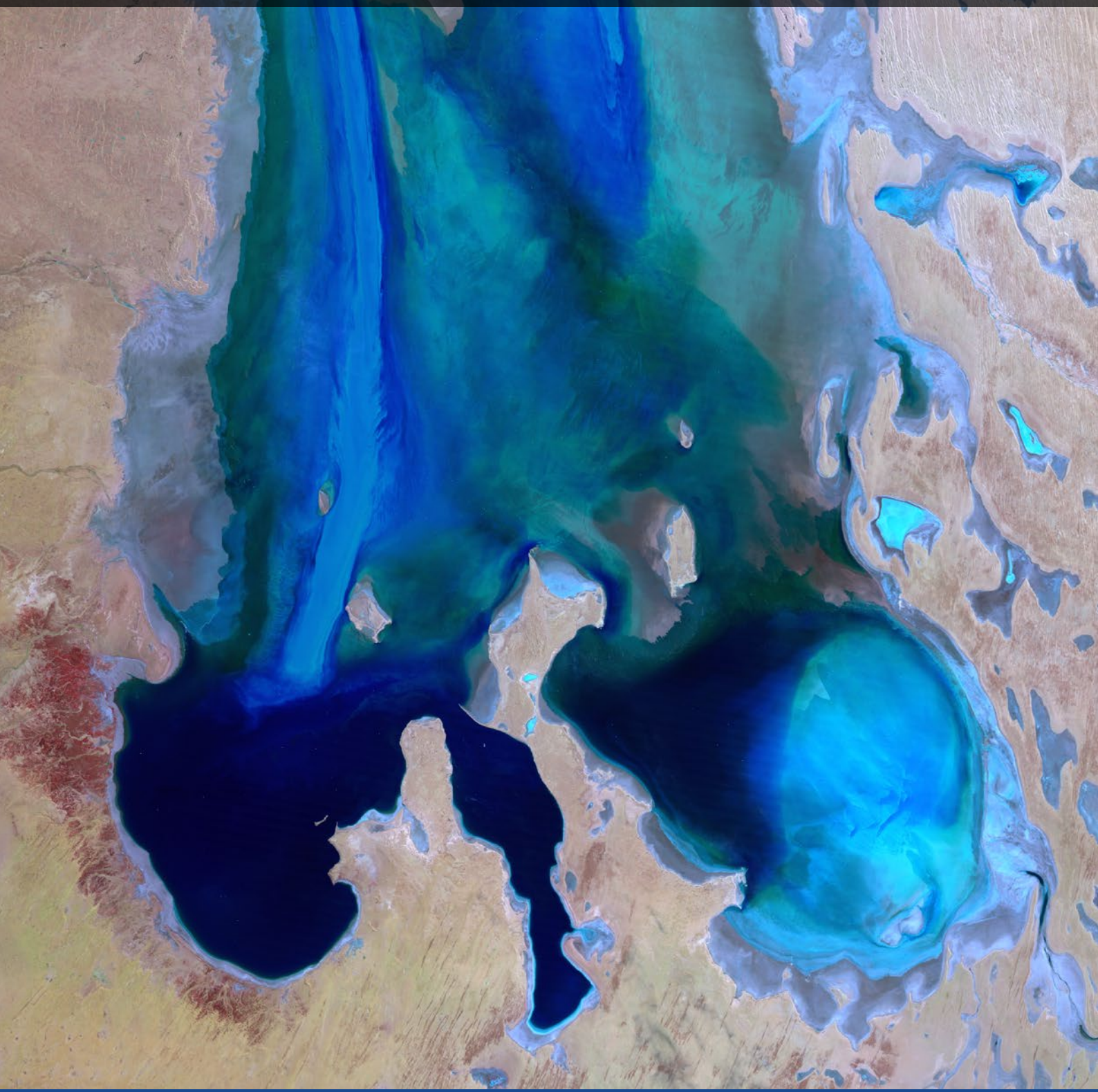
Earth Observation:
Data, Processing and Applications

Volume 2A: Processing—Basic Image Operations



The report is available in PDF format at *http://www.crcsi.com.au/earth-observation-series*
We welcome your comments regarding the readability and usefulness of this report. To provide feedback, please contact us at *info@crcsi.com.au*.

Publisher:
Australia and New Zealand CRC for Spatial Information

ISBN [ONLINE]:
978-0-9943019-5-6

Recommended Citation for Volume 2A:
CRCSI (2017) *Earth Observation: Data, Processing and Applications. Volume 2A: Processing—Basic Image Operations.* (Eds. Harrison, B.A., Jupp, D.L.B., Lewis, M.M, Sparks, T., Phinn, S.F., Mueller, N., Byrne, G.) CRCSI, Melbourne.

Background image on previous page: Landsat-5 image of Lake Eyre, South Australia, acquired on 9 May, 2009, while the lake was filling. Three images are juxtaposed: a colour composite of bands 5, 4, 2 as RGB on the left, a principal component transformation (with PC3, PC4 and PC1 displayed as RGB) in the centre and a 10 class ISODATA classification on the right.
Source: Norman Mueller, Geoscience Australia

Acknowledgements

Production of this series of texts would not have been possible without the financial support of CSIRO, CRCSI, GA and BNHCRC, input from members of the editorial panels and direction from members of the various advisory panels.

Volumes 1 and 2 of this series are based on text originally published in Harrison and Jupp (1989, 1990, 1992 and 1993)[1]. Many illustrations and some text from these publications have been reproduced with permission from CSIRO. Much of the text in Section 3 has been derived from Jupp and Strahler (1996)[2].

Other contributors are gratefully acknowledged:

- reviewers: Alfredo Huete and Laurie Chisholm (various sections), David Hudson and Joshua Sixsmith (parts of Sections 1 and 2), and Tim Malthus (Section 3);
- illustrations: Norman Mueller kindly supplied most of the lovely images; other contributors of graphical material include: Tony Sparks, Megan Lewis, Fuqin Li, David Jupp, Stephen Sagar, Medhavy Thankappan, Edward King, Karen Joyce, Guy Byrne, Chris MacLellan, Tim Malthus, Neil Flood, Thomas Cudahy, Michael Caccetta, Ken Clarke and Dale Roberts;
- Jason Mazur (MIrage Digital) for permission to reproduce Figure 5.15 in this document;
- Paul Cockram and Matej Hrcek, who 'translated' the Harrison and Jupp (1989, 1990) files from Quark Express 2.0;
- Joost Kuckartz for formatting of references and equations;
- Daniel Rawson (Accessible Publication & Template Design) for layout and formatting; and
- Carl Davies (CMDphotographics) for selected graphical illustrations.

We thank those owners of copyrighted illustrative material for permission to reproduce their work. Credits for individual illustrations are provided below the relevant graphic.

1 Harrison, B.A., and Jupp, D.L.B. (1989) *Introduction to Remotely Sensed Data: Part ONE of the microBRIAN Resource Manual.* CSIRO, Melbourne. 156pp.
Harrison, B.A., and Jupp, D.L.B. (1990) *Introduction to Image Processing: Part TWO of the microBRIAN Resource Manual.* CSIRO, Melbourne. 256pp.
Harrison, B.A., and Jupp, D.L.B. (1992) *Image Rectification and Registration: Part FOUR of the microBRIAN Resource Manual.* MPA, Melbourne.
Harrison, B.A., and Jupp, D.L.B. (1993) *Image Classification and Analysis: Part THREE of the microBRIAN Resource Manual.* MPA, Melbourne.

2 Jupp, D.L.B., and Strahler, A. (1996) *Image Brightness and BRDF Workshop Issues.* CSIRO Earth Observation Centre, Canberra.

Table of Contents

List of Figures

List of Tables

List of Excurses

Introduction

Image processing allows digital image data to be displayed, enhanced, analysed or integrated with other datasets. A wide variety of image processing operations can be used to enhance the information context of imagery, such as highlighting or detecting features of interest, changing the image geometry, extracting patterns, removing noise, analysing time series trends or reducing data volume. Image processing methods rely upon repetitive mathematical operations to achieve these results.

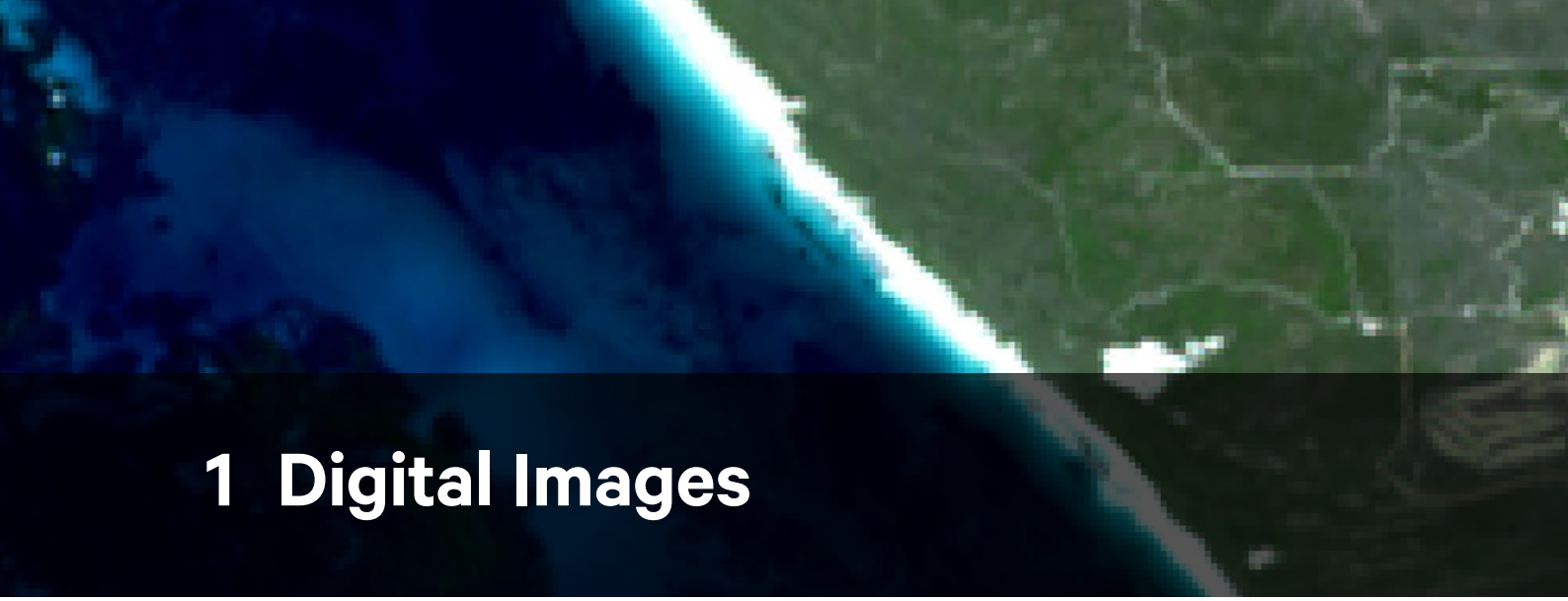
While most image datasets can be processed on individual computers using appropriate software, extremely large datasets may require specialised hardware and software to perform image processing operations with flexibility and efficiency. Increasingly the storage and processing of very large image datasets is undertaken in internet-based (or 'cloud-based') data stores, using open access software libraries.

The following three sections describe:

- digital image structure and formats (see Section 1);
- processing stages and methodologies (see Section 2); and
- calibration approaches (see Section 3).

Contents

Background image on previous page: Landsat-5 image of Wilson's Promontory, Victoria, acquired on 30 January 2011, shown using bands 3, 2, 1 as RGB.
Source: Norman Mueller, Geoscience Australia



1 Digital Images

Any pictorial or spatial data may be represented as a digital, or numeric, image. An image needs to be in digital form to be processed by computer software. Spatial data, such as a digital photograph, is represented as an image using a pattern of non-overlapping cells (or a 'tessellation') with each cell being associated with the observed 'colour' of some portion of the original object or picture. In remotely sensed data, colour usually refers to radiance in discrete wavelength bands or channels. The differences between digital and photographic images, as well as various aspects of image resolution (spectral, spatial, temporal and radiometric), are described in Volume 1B—Section 1 of this series.

1.1 Image Structure

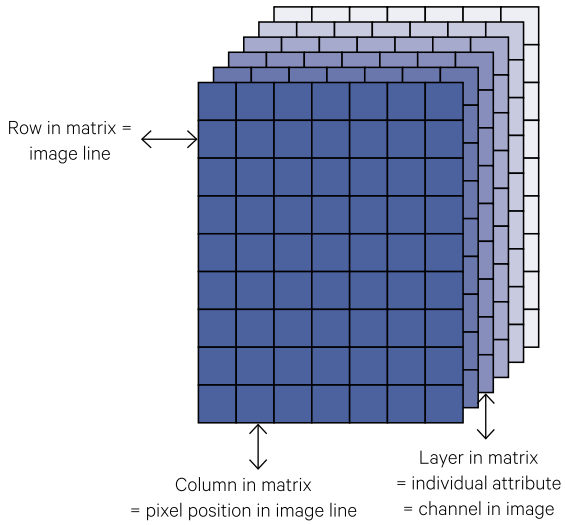
Image processing systems assume that the tessellation used to record the values in an image consists of a regular cell size, and usually represent the tessellation as a grid. In image processing each cell in the image grid is referred to as a 'pixel' (picture element). A single channel in a digital image can be considered as a two-dimensional table, or matrix, with the pixels across a line representing columns, and the lines down the image representing rows. Similarly, in a multi-channel image, each channel can be viewed as an attribute or dimension in a three-dimensional table.

The entries in the table represent an observation of some attribute or variable (see Volume 1A—Section 1). In the case of remotely sensed data, the attribute (or image channel) is a scaled record, in a given wavelength band, of the energy reflected or emitted by the Earth's surface. This concept is illustrated in Figure 1.1. Image pixel locations are generally referenced by their pixel (column) and line (row) positions relative to an origin in the top-left corner of the image (see Section 7 below).

Digital images allow the use of a large range of mathematical and statistical operations to enhance, reduce or identify particular features in spatial data. Such operations analyse the colour, or spectral, characteristics and/or the pattern, or spatial attributes of an image as described in Sections 8 to 10 below and Volumes 2C and 2E. The geometric properties of image data also allow operations to be applied to modify the spatial relationship between features by changing image shape, scale and/or orientation (see Section 7 and Volume 2B). Consideration of image geometry is important when comparing multiple images from the same or different sensors, including time series datasets (see Volume 2D). Ongoing improvements in the capabilities of image processing software enable some geometric operations to be applied to imagery concurrently with other processes.

Figure 1.1 Components of a digital image

A multi-channel image comprises three data dimensions—pixels, lines and channels. Pixels (or 'elements' or 'samples') along the line can be considered as columns in a matrix. Lines in the image are like rows of a matrix, with each channel being an attribute layer in the matrix.



Background image: Landsat-8 image of Deepene Beach, WA, acquired on 17 December 2016. This natural colour composite is based on bands 4, 3 and 2 as RGB.
Source: Norman Mueller, Geoscience Australia

Earth Observation (EO) image pixels contain estimates of surface radiance in particular wavelengths. However, image channels could also contain values that indicate other spatial attributes. Such attributes can be represented as either a scaled measure of some parameter, such as temperature or height at each pixel, or a nominal value that indicates a category, such as soil type, land use or cadastral region. The process of integrating other data sources with remotely sensed imagery is detailed in Volume 2D.

1.2 Image Data Formats

A wide range of file formats have been devised to store image data. This section briefly includes the most commonly encountered generic formats. Various forms of metadata are often added before or after these generic formats in specific proprietary and industry standard file types (see Section 1.3).

As introduced above, image data can be viewed in terms of three dimensions of data:

- pixels across each line;
- lines down the image; and
- channels (or bands) of pixels and lines (see Figure 1.1).

These image data dimensions are generally recorded using one of three standard ordering formats:

- Band Sequential (BSQ);
- Band Interleaved by Pixel (BIP); or
- Band Interleaved by Line (BIL).

The implementation of each format is illustrated in Figure 1.2d, Figure 1.2e and Figure 1.2f respectively. In any of these file formats, all pixels in an image channel have the same fixed range of possible values, such as:

- one bit (range 0–1);
- one byte (usually 8 bits: 0–255);
- one integer (0–32767);
- a signed integer (-32768–32767); or
- floating point (positive or negative numbers including decimal places).

Each file format generally reserves one value in the potential data range to indicate missing values, which is called the 'null' (or no data) value.

Figure 1.2 Image data formats

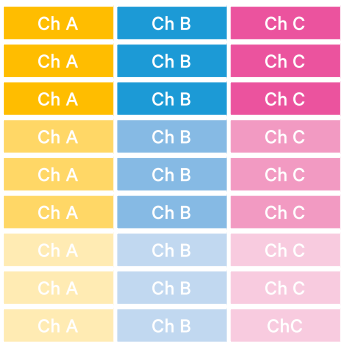
a. Example image data for channel A

b. Example image data for channel B

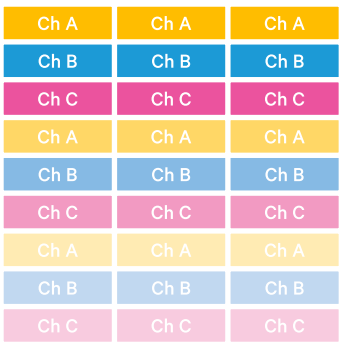
c. Example image data for channel C

d. Band sequential file format (BSQ)

e. Band interleaved by pixel format (BIP)



f. Band interleaved by line format (BIL)



Adapted from: Chuvieco and Huete (2010)

1.2.1 Band sequential (BSQ)

This format records all the data for each band or channel in the image as a continuous block of information (see Figure 1.1d). Within each channel block, the ordering is usually by line, that is:

channel 1:
line 1: pixel 1, pixel 2, ... pixel m
line 2: pixel 1, pixel 2, ... pixel m
...
line n: pixel 1, pixel 2, ... pixel m

channel 2:
line 1: pixel 1, pixel 2, ... pixel m
line 2: pixel 1, pixel 2, ... pixel m
...
line n: pixel 1, pixel 2, ... pixel m
...
...

channel z:
line 1: pixel 1, pixel 2, ... pixel m
line 2: pixel 1, pixel 2, ... pixel m
...
line n: pixel 1, pixel 2, ... pixel m

This format allows data for an individual channel to be retrieved quickly but requires several read operations to get all the data for all channels relating to a particular pixel or line.

1.2.2 Band interleaved by pixel (BIP)

This format records all channels for each image pixel column as sequential file data (see Figure 1.1e):

line 1:
pixel 1: channel 1, channel 2, ... channel n
pixel 2: channel 1, channel 2, ... channel n
...
pixel z: channel 1, channel 2, ... channel n

line 2:
pixel 1: channel 1, channel 2, ... channel n
pixel 2: channel 1, channel 2, ... channel n
...
pixel z: channel 1, channel 2, ... channel n
...
...

line m:
pixel 1: channel 1, channel 2, ... channel n
pixel 2: channel 1, channel 2, ... channel n
...
pixel z: channel 1, channel 2, ... channel n

I'm not dumb. I just have a command of thoroughly useless information (Calvin, Calvin and Hobbes)

This format allows rapid access to columns of image data for all channels so is useful for data requiring pixel-based processing that uses multiple channels, such as many hyperspectral analyses. Access to complete channel information, however, is slower than with BSQ format data.

1.2.3 Band interleaved by line (BIL)

This format records all channels for each image line as consecutive data in the file (see Figure 1.1f):

line 1:
channel 1: pixel 1, pixel 2, ... pixel m
channel 2: pixel 1, pixel 2, ... pixel m
...
channel z: pixel 1, pixel 2, ... pixel m

line 2:
channel 1: pixel 1, pixel 2, ... pixel m
channel 2: pixel 1, pixel 2, ... pixel m
...
channel z: pixel 1, pixel 2, ... pixel m
...
...

line n:
channel 1: pixel 1, pixel 2, ... pixel m
channel 2: pixel 1, pixel 2, ... pixel m
...
channel z: pixel 1, pixel 2, ... pixel m

The BIL format allows all data for a particular image line to be accessed quickly but separates the data for each individual channel so that complete channel information is retrieved more slowly than for BSQ. This format is most appropriate for EO data since many processing operations require the analysis of both multiple channel values and groups of pixels, that is, it is a good compromise between BSQ and BIP. Since various correction techniques process the data for individual image lines as discrete entities, this format enables efficient data access.

1.3 Data Volume

The size of an uncompressed image file, in bytes, can be computed as:

number of pixels x *number of lines* x *number of channels* x *number of bytes per pixel* + *metadata*

To reduce the amount of disk space required to store an image data file, data compression methods can be used in conjunction with various data formats. These methods can be divided into two categories, which describe the impact of compression on the original data:

- loseless—maintains the original data, which can be recovered by decompression; and
- lossy—reduces the original data by permanently removing information that the compression algorithm considers to be redundant, such that the removed information cannot be recovered by decompression. Lossy compression algorithms are suited to visual representation of imagery but should not be used when further processing of image data is anticipated, or for any form of data where specific values represent particular categories or null pixels.

One of the most commonly used compression methods for thematic data is 'run-length encoding' which records the number of sequential pixels along a line which have the same value, then the value itself, rather than one value for each pixel. This lossless method reduces the number of values to be stored in the image file when adjacent pixels in the line frequently have the same value but increases the time required to read or write the file. Such methods are most appropriate for categorical data sets, however their effectiveness on unclassified EO imagery depends on the sensor and ground features being imaged. In general terms, in imagery where adjacent pixels rarely have the same values, there is less reduction in data volume when using this method to encode data values (see Table 1.1).

Table 1.1 Example of data compression rates from run length encoding

Different images were compressed using run length encoding to demonstrate the resulting reduction in data volume. All images were stored in 16 bit integer format, with scaled reflectance values.

Image	Description	Data volume reduced by
Landsat-8 of rural NSW	Full scene, including null pixels on outside edges	25%
	Subset without null pixels	27%
Landsat-5 of NSW south coast	Full scene, including null pixels on outside edges	48%
	Subset with 1/3 water pixels	47%
	Subset of agricultural land	58%
	Subset of classified image containing mixed rural/woodland with minor urban component	78%

Source: Tony Sparks, Icon Water

1.4 Metadata

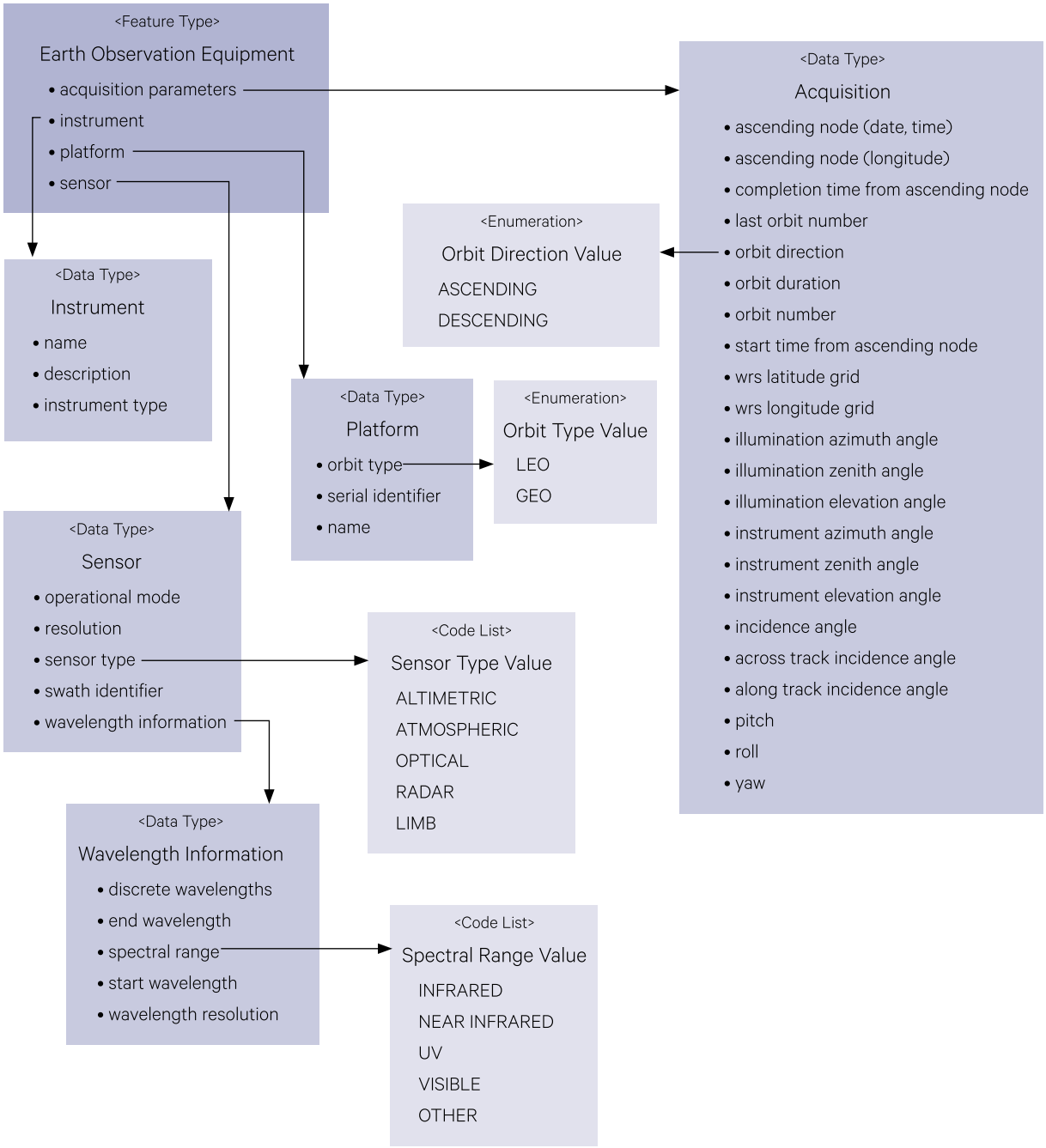
EO image files generally include additional information relating to the source and size of the image and its processing history. This information is called metadata (data about data) and is typically written in an image file as one or more header records before the image data. The goal of metadata for geophysical datasets is that the data becomes more portable, more accessible and 'self-describing' (Retscher *et al.*, 2011). For example, metadata elements for EO datasets include information relating to the acquisition platform, acquisition sensor, data processing history and spatial representation.

Standards for metadata of EO datasets have been developed to maximise consistency in data transfer and distinguish between:

- structural metadata—describes how data components are recorded as data structures; and
- descriptive metadata—describes the data components.

For example, the Open Geospatial Consortium (OGC) metadata describing the equipment used during EO data acquisition is shown in Figure 1.3 and the description of the EO result is given in Figure 1.4 (OGC, 2012). The Generic Earth Observation Metadata Standard (GEOMS) is commonly used to exchange validation data used in conjunction with EO datasets (Retscher *et al.*, 2011).

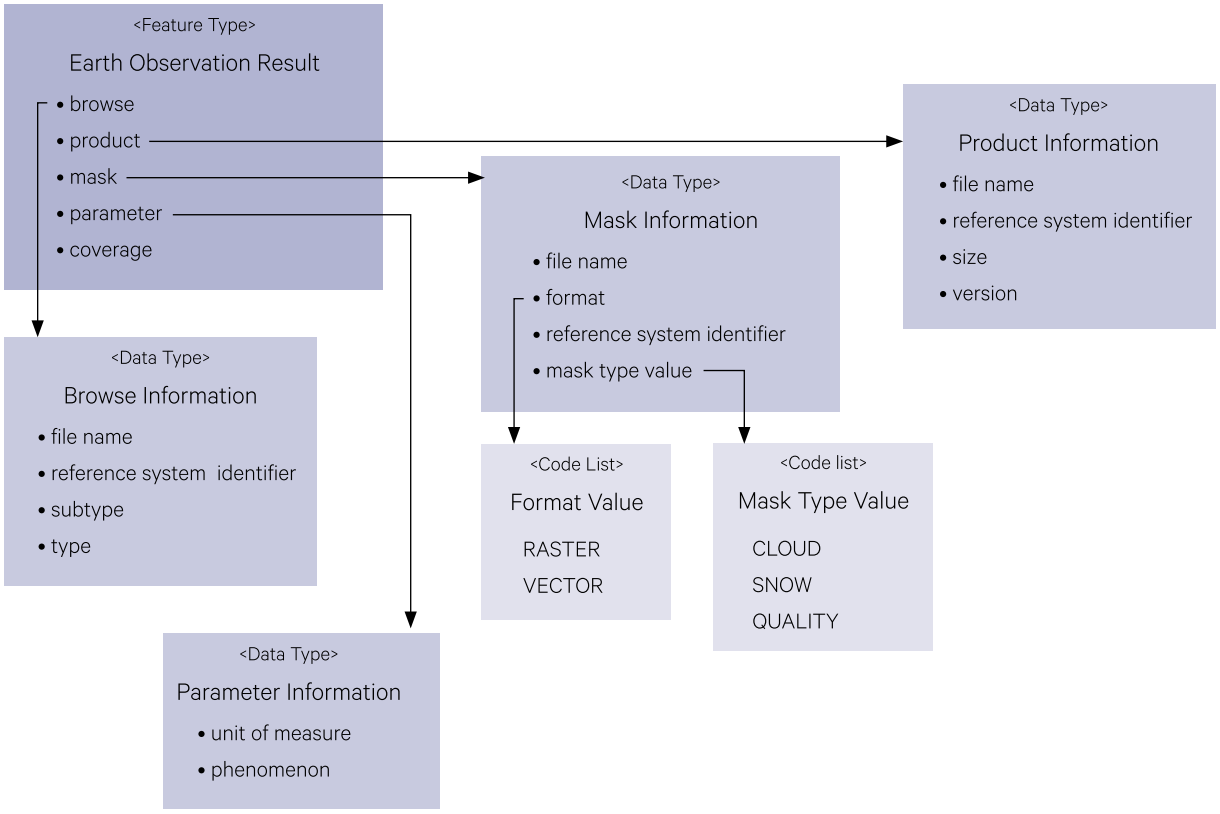
Figure 1.3 Earth Observation equipment metadata



Adapted from: OGC (2012) 8.1.2

The Sensor Web, envisaged as encircling the globe with electronic sensing devices during the next century, will add a new level of sophistication for observing and measuring the Earth, and calibrating and validating EO imagery (van Zyl *et al.*, 2009; Bröring *et al.*, 2011). Accessing and interrogating a system of this scale and complexity will take metadata to a new level of sophistication and utility (Hu *et al.*, 2014).

Figure 1.4 Earth Observation result metadata



Adapted from: OGC (2012) 8.1.4

Excursus 1.1—Metadata Standards

Source: David Hudson, Geoscience Australia

The term 'metadata' is used to describe data about data. There are many different types of metadata and many different metadata standards.

The most common standards used for EO data are:

- data labelling standards:
  - all spatial data—ISO 19115-1 (*https://www.iso.org/standard/53798.html*)
  - raster data—ISO 19115-2 (*https://www.iso.org/standard/39229.html*)
- data transport standards:
  - picture data—Web Map Service (*http://www.opengeospatial.org/standards/wms*)
  - vector data—Web Feature Service (*http://www.opengeospatial.org/standards/wfs*)
  - raster data—Web Coverage Service (*http://www.opengeospatial.org/standards/wcs*)
- EO product standard:
  - CEOS Analysis Ready Data for Land (CARD4L)—(CEOS, 2017)

In the broader context of spatial data, ANZLIC (Australia and New Zealand Land Information Council) is sponsoring the Foundation Spatial Data Framework (FSDF). This ongoing, collaborative initiative provides a common reference to assemble and maintain foundation level spatial data in Australian and New Zealand. The FSDF aims to deliver national coverage of the most current and authoritative foundation spatial data to serve the widest possible variety of users in both government and commercial activities. Ten themes for this standardised and quality-controlled data have been defined:

- administrative boundaries;
- elevation and depth;
- geocoded addressing;
- imagery;
- land cover and land use;
- land parcel and property;
- place names;
- positioning;
- transport; and
- water (ANZLIC, 2017).

1.5 Pixel Quality Flag

Some EO image suppliers include an additional channel with the image data to indicate the data quality of the original image pixel or derived products. For example, many recent Landsat products from Geoscience Australia (GA) include a Pixel Quality Flag value, which is the binary representation of the cumulative sum of 16 test conditions (Sixsmith, 2012). The test conditions used to determine each bit value are summarised in Table 1.2. An example of interpreting this channel is provided in Table 1.3. Such pixel quality flags assist users in checking that the imagery or derived products are reliable and suitable for analysis and interpretation (Sixsmith *et al.*, 2013).

Table 1.2 Pixel quality flag

Example of pixel quality flag values used for Landsat images by Geoscience Australia.

Test Condition	Bit Position (right to left)	Pixel Quality Flag equals 0 if this pixel is	Pixel Quality Flag equals 1 if this pixel is
Band 1 is saturated	0	Saturated in Band 1	Not saturated in Band 1
Band 2 is saturated	1	Saturated in Band 2	Not saturated in Band 2
Band 3 is saturated	2	Saturated in Band 3	Not saturated in Band 3
Band 4 is saturated	3	Saturated in Band 4	Not saturated in Band 4
Band 5 is saturated	4	Saturated in Band 5	Not saturated in Band 5
Band 6-1 is saturated (This is designed to match Landsat-7 ETM+. The thermal band for Landsat-5 TM corresponds to Band 6-1 and this value is duplicated in Band 6-2.)	5	Saturated in Band 6-1	Not saturated in Band 6-1
Band 6-2 is saturated (This is designed to match Landsat-7 ETM+. The thermal band for Landsat-5 TM corresponds to Band 6-1 and this value is duplicated in Band 6-2.)	6	Saturated in Band 6-2	Not saturated in Band 6-2
Band 7 is saturated	7	Saturated in Band 7	Not saturated in Band 7
Contiguity	8	Null value in at least one band	All bands for this pixel contain non-null values
Land or Sea (based on vector data from TM World Borders (*http://thematicmapping.org/downloads/world_borders.php*) spliced with Australian Coastal Boundary using a 100 m buffer into the sea.)	9	Sea	Land
Cloud (based on Automatic Cloud Cover Assessment (ACCA) algorithm; Irish, 2000; Irish ***et al.***, 2006)	10	Cloud	No Cloud
Cloud (Fmask; based on TOAR rather than NBAR as input reflectance; (Zhu and Woodcock, 2012)	11	Cloud	No Cloud
Cloud Shadow (only computed for cloud detected by ACCA algorithm using in-house algorithm)	12	Cloud Shadow	No Cloud Shadow
Cloud Shadow (Fmask; based on TOAR rather than NBAR as input reflectance; (Zhu and Woodcock, 2012))	13	Cloud Shadow	No Cloud Shadow
Topographic Shadow (This feature has not been implemented yet so bit 14 is set to value 0)	*14*	*Topographic Shadow*	*No Topographic Shadow*
To be determined (Default untested value currently set to 0)	*15*	*N/A*	*N/A*

Table 1.3 Examples of pixel quality flag values

Four examples of pixel quality flag values defined in Table 1.2.

Example	Binary Code (for bits 0 to 15)																Equivalent Decimal value	Interpretation of Binary Code
	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0		
1	0	0	1	1	0	0	1	1	1	1	1	0	1	1	1	0	13294	Bands 1 and 5 are saturated (bits 0 and 4 = 0) No null values (bit 8 = 1) Land pixel (bit 9 = 1) Clouds detected by ACCA and Fmask (bits 10 and 11 = 0) No cloud shadow (bits 12 and 13 = 1)
2	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	4095	No saturated bands (bits 0–7=1) No null values (bit 8 = 1) Land pixel (bit 9 = 1) No cloud detected (bits 10 and 11 = 1) Cloud shadow detected by ACCA and Fmask (bits 12 and 13 = 0)
3	0	0	1	1	0	0	1	1	0	1	1	0	1	1	1	0	13166	Bands 1, 5 and 7 are saturated (bits 0, 4 and 7 = 0) No null values (bit 8 = 1) Land pixel (bit 9 = 1) Clouds detected by ACCA and Fmask (bits 10 and 11 = 0) No cloud shadow (bits 12 and 13 = 1)
4	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16383	Clear Land Pixel—no saturation, null values, cloud, or cloud shadow

1.6 Hierarchical Storage Structures

A range of digital storage structures has been developed to access large images more efficiently. Several hierarchical data structures have been developed to optimise storage and retrieval of large image-based datasets. Three implementations of such structures include:

- pyramids of downscaled image tiers which effectively store image at a range of scales (see Volume 1B—Section 1.2). These tiers enable imagery to be rapidly displayed at a range of scales;
- mosaics (also called overviews or tilesets) of smaller images which together cover a larger area, possibly with overlapping extent and different geometric properties (see Section 1.7); and
- datacubes of calibrated, conformal image layers through time (see Section 3.4).

Each of these structures offer different advantages for displaying, interrogating and processing large image-based datasets. As the volume of image data increases and its spatial resolution decreases, efficient access to data is becoming critical for many applications. Some interchange formats used with EO imagery include:

- GeoTIFF (Georeferenced Tagged Image File format)—an open file format based on TIFF for georeferenced raster imagery and embraced by NASA Earth Science data systems;
- HDF (Hierarchical Data Format)—multi-object file format used for all NASA EO data products;
- NetCDF (Network Common Data Form)—set of self-describing, generic data formats (and software libraries) for multidimensional, array-oriented scientific data; and
- JPEG—(Joint Photographic Experts Group)—most common standard for digital photography. Many JPEG implementations use lossy compression, which render smooth variations in tone and colour but reduces image fidelity (see Section 1.7).

1.7 Further Information

Metadata

Open Geospatial Consortium (OGC): (OGC, 2012)

The Generic Earth Observation Metadata Standard (GEOMS): Retscher *et al.* (2011)

Foundation Spatial Data Framework (FSDF): *http://www.anzlic.gov.au*

Sensor Web

Hu *et al.* (2014)

Virtual Constellations: *http://ceos.org/ourwork/virtual-constellations/*

Data structures

GEOTIFF: *http://trac.osgeo.org/geotiff/* (*https://earthdata.nasa.gov/standards/geotiff*)

HDF: *http://nsidc.org/data/hdfeos/*

NetCDF: *https://www.unidata.ucar.edu/software/netcdf/*

JPEG: *https://jpeg.org/jpeg2000/index.html*

Tileset Concepts and Terminology: *http://www.microimages.com/documentation/FeatureSummaries/2010/Tileset%20Terminology.pdf*

1.8 References

ANZLIC (2017). *Australian and New Zealand Land Information Council (ANZLIC) website*. Retrieved from *http://www.anzlic.gov.au*.

Bröring, A., Echterhoff, J., Jirka, S., Simonis, I., Everding, T., Stasch, C., Liang, S., and Lemmens, R. (2011). New Generation Sensor Web Enablement. Sensors, 11(3), pp. 2652. doi:*http://dx.doi.org/10.3390/s110302652*.

CEOS (2017). *CEOS Analysis Ready Data for Land (CARD4L) Description Document*. Retrieved from: *https://www.google.com.au/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiR-_OCosLXAhWLnJQKHRpUCOsQFggoMAA&url=http%3A%2F%2Fceos.org%2Fdocument_management%2FMeetings%2FPlenary%2F30%2FDocuments%2F5.5_CEOS-CARD4L-Description_v.22.docx&usg=AOvVaw2r8F89fp--1nxlALHTQsAi*.

Chuvieco, E., and Huete, A. (2010). *Fundamentals of Satellite Remote Sensing*. CRC Press, Boca Raton.

Hu, C., Guan, Q., Chen, N., Li, J., Zhong, X., and Han, Y. (2014). An Observation Capability Metadata Model for EO Sensor Discovery in Sensor Web Enablement Environments. *Remote Sensing*, 6(11), pp. 10546.

Irish, R. R. (2000). *Landsat 7 automatic cloud cover assessment*. Paper presented at the SPIE 4049, Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI, Orlando, United States.

Irish, R. R., Barker, J. L., Goward, S. N., and Arvidson, T. (2006). Characterization of the Landsat-7 ETM+ Automated Cloud-Cover Assessment (ACCA) Algorithm. *Photogrammetric Engineering & Remote Sensing*, 72(10), pp. 1179-1188. doi:*http://dx.doi.org/10.14358/PERS.72.10.1179*.

OGC (2012). *Earth Observation Metadata Profile of Observations and Measurements* (OGC 10-157r3) (J. Gasperi, F. Houbie, A. Woolf, and S. Smolders Eds.). Open Geospatial Consortium.

Retscher, C., de Maziere, M., Meijer, Y., Vik, A. F., Boyd, I., Niemeijer, S., Koopman, R. M., and Bojkov, B. (2011). *The Generic Earth Observation Metadata Standard (GEOMS), Version 1.0*. Retrieved from *http://avdc.gsfc.nasa.gov/PDF/GEOMS/geoms-1.0.pdf*.

Sixsmith, J. (2012). *Unlocking the Landsat Archive Work Package 4.1. Pixel Quality Documentation* (GA TRIM Number D2012-212824). Geoscience Australia.

Sixsmith, J., Oliver, S., and Lymburner, L. (2013). *A hybrid approach to automated landsat pixel quality*. Paper presented at the IEEE Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, Australia.

van Zyl, T. L., Simonis, I., and McFerren, G. (2009). The Sensor Web: systems of sensor systems. *International Journal of Digital Earth*, 2(1), pp. 16-30. doi:*http://dx.doi.org/10.1080/17538940802439549*.

Zhu, Z., and Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118, pp. 83-94. doi:*http://dx.doi.org/10.1016/j.rse.2011.10.028*.

2 Processing Stages

Image processing techniques can be applied to a wide variety of data in a diverse range of disciplines. These techniques rely on 'spectral' and/or 'spatial' analysis. In spectral analysis, the values of individual pixels in an image are used to identify particular features. In spectroscopy, for example, identification of a sample is based on its spectral response in a large number of channels. Alternatively, in spatial analysis, the patterns formed by groups of pixels in an image may be used to identify or enhance features on the basis of their shape, size or proximity. For example, image spatial patterns are particularly important in visual interpretation of images. Image processing of EO data combines both spectral and spatial analyses to interpret, analyse and model image data.

2.1 Data Provenance

At all stages of processing EO imagery, analysts need to be mindful of the image data provenance, that is:

- its intrinsic resolutions and extents in terms of spectral, spatial, radiometric and temporal properties (see Volume 1); and
- any pre-processing that has been applied to the data (see Section 3 below).

Selection of appropriate processing sequences for calibration, analysis, modelling and integration of EO datasets requires both an understanding of data provenance and insight into the assumptions and constraints of the component algorithms. All EO products should be produced using mature, operational processing pathways, which are documented in terms of:

- data sources (primary EO inputs, metadata parameters and any ancillary data used);
- major algorithms, described in a relevant Algorithm Theoretical Basis Document (ATBD);
- processing sequence, as a series of steps from input to output;
- validation of product or its underlying algorithm; and
- statement of accuracy and limitations relating to product and/or input data.

It should be noted that image data can represent two fundamental types of information. The value of each pixel in a given image channel is derived from either:

- a continuous range of values for some characteristic, such as the acquired radiance in optical or thermal imagery; or
- a set of categories, such that different categories (or classes) are represented by different channel values. The results from classifying EO imagery into surface cover classes, for example, are generally encoded into a single classification channel (see Section 9 and Volume 2E).

The types of image processing tools and methods that are appropriate for continuous data may not be appropriate for categorical data and, conversely, processes that can be applied to categorical data may not be suitable for continuous data. In this context, knowledge of data provenance is essential for meaningful and reliable results.

Image products derived from EO data sources have traditionally been grouped in terms of their processing history. Here, image products will be considered in terms of two broad types:

- data products—corrected image observations from EO sensors (see Section 2.1.1); and
- information products—derived from EO data products for specific applications (see Section 2.1.2).

Background image: Astronaut photo of Roebuck Bay, just south of Broome, WA, taken on 11 June 2015 from the International Space Station (ISS) at an altitude of 402 km. **Source:** NASA. Retrieved from: *https://eol.jsc.nasa.gov/SearchPhotos/photo.pl?mission=ISS044&roll=E&frame=885*

2.1.1 Data products

In the context of EO imagery, data products comprise corrected image observation values acquired from EO sensors. The observation values are presented in grid cells and represented as radiance and/or digital numbers. The corrections applied to these values include geometric, radiometric, and atmospheric adjustments to compensate for known distortions in the original EO image data.

Preparation of data products does not involve interpretation models or classification procedures. These products provide EO data for users to process for their own purposes and correspond to the traditional EO image scenes, such as terrain-corrected (also known as ortho-corrected) and map-oriented (with no terrain correction) imagery. Examples of GA EO data products include the Australian Reflectance Grid (ARG) suite of products. The ARG25 product, for example, delivers an estimate of surface reflectance from Landsat TM/ETM+ data using physical rather than empirical models. This product aims to deliver imagery in which differences in the reflective value(s) between images that were acquired at different times by different sensors[2] will be primarily due to on-ground changes in biophysical parameters rather than artefacts of the imaging environment. These products are processed to deliver standardised surface reflectance measures and, in future, are intended to be conformal to the National Nested Grid (NNG; see Section 7).

2.1.2 Information products

Information products derived from EO data are also available for specific applications. These products have been developed to highlight specific features of interest for particular users and comprise interpreted values derived from classification and/or modelling techniques applied to the corrected image values. These are represented as grid cell values and can depict a broad range of information, such as:

- vegetation indices (NDVI, EVI—see Volume 3A);
- temperature (K);
- position (x, y coordinates);
- fractional cover; or
- land cover (e.g. water, bare soil, vegetation type).

Data channels in these products can be either continuous (such as range of NDVI values) or categorical (such as land cover categories). Examples of information products produced by GA include the Australian Geographic Reference Grid (AGRI), Sentinel Hotspots, the Dynamic Land Cover Dataset (DLCD) and the Intertidal Extent Model (see Excursus 2.1).

You can have data without information,
but you cannot have information without data.
(Daniel Keys Moran)

2 Notwithstanding differences due to variations in spectral bandwidth between sensors.

Excursus 2.1—Intertidal Extents Model

Source: Stephen Sagar, Geoscience Australia
Further Information: (Sagar *et al.*, 2017)

An example of continental scale modelling utilising the Landsat archive

The InterTidal Extents Model (ITEM) (Sagar *et al.*, 2017) provides valuable information regarding the exposed intertidal zone—the area between the land and sea that can be observed between the highest and lowest tide. Traditionally, surveying this dynamic landscape has been difficult, especially when mapping large expanses of shoreline—some 50,000 km for the Australian coast.

The process used to create the ITEM product leverages the full 28 year time series archive of Landsat data, managed in a countrywide, analysis ready data platform, known as the Australian Geoscience Data Cube (AGDC; see Section 3.4). To fully exploit the time series of data, ITEM moves away from a single scene analysis method to a full ensemble modelling framework by implementing two key concepts:

- through the use of an ancillary tidal model, the ensemble of Landsat observations is reframed and reordered from the traditional time series domain to one based on tidal height; and
- the full reordered ensemble of data is then examined using a median estimator to account effectively for data noise and incompleteness.

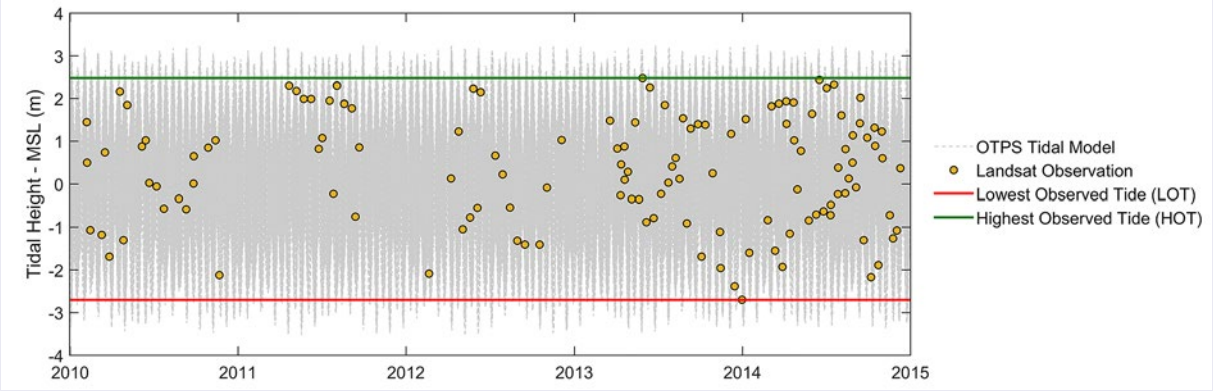
This concept illustrates how EO data analysis can be driven by other environmental attributes, enabling a new range of applications to grow from the traditional time series approach.

Processing steps

1. The modelling process to derive the product utilises a continental scale tidal model developed by Oregon State University (Egbert and Erofeeva, 2002, 2010) applied to the 221 AGDC cells that cover the Australian coastline.
2. Each Landsat tile in these coastal cells is attributed with a tidal height, relative to Mean Sea Level (MSL), corresponding to the time of observation acquisition. Tidal heights are modelled at a single constant location for each 1° by 1° cell, determined manually and based on the complexity of the coastline contained within the cell.
3. The time series of observations for each AGDC cell is then sorted based on the modelled tidal height, and split into subsets representing 10% percentile intervals of the Observed Tidal Range (OTR) for that cell. Due to the Sun synchronous nature of the various Landsat sensor observations, it is unlikely that the full physical extent of the tidal range in any cell will be observed. Hence, terminology has been adopted for the product to reflect the highest modelled tide observed in a given cell (HOT) and the lowest modelled tide observed (LOT; see Figure 2.1). These measures are relative to MSL and have no consistent relationship to Lowest Astronomical Tide (LAT) and Highest Astronomical Tide (HAT).
4. Each tile observation is masked for pixel quality based on the AGDC PQA layer to exclude pixels flagged for cloud, band saturation and contiguity (see Section 1.5). The Normalised Difference Water Index (NDWI) (McFeeters, 1996) is calculated for each tile, and a pixel-based median NDWI composite is derived for each 10% percentile interval of the cell's OTR.
5. NDWI composites derived for each cell are used to estimate the extent of exposed substratum at each 10% interval of the OTR through a classification of water/non-water pixels, and combined to create the ITEM Relative Extents Model (see Figure 2.2).
6. Variability of the NDWI values contributing to the tidal interval median composites is analysed for each pixel. This produces a confidence layer output that indicates regions in which the observed water/land interactions are not driven by tidal factors. This layer helps to identify modelling errors and issues such as coastal instability over time.

Figure 2.1 The concept of tidal attribution and Lowest (LOT) and Highest (HOT) Observed Tide.

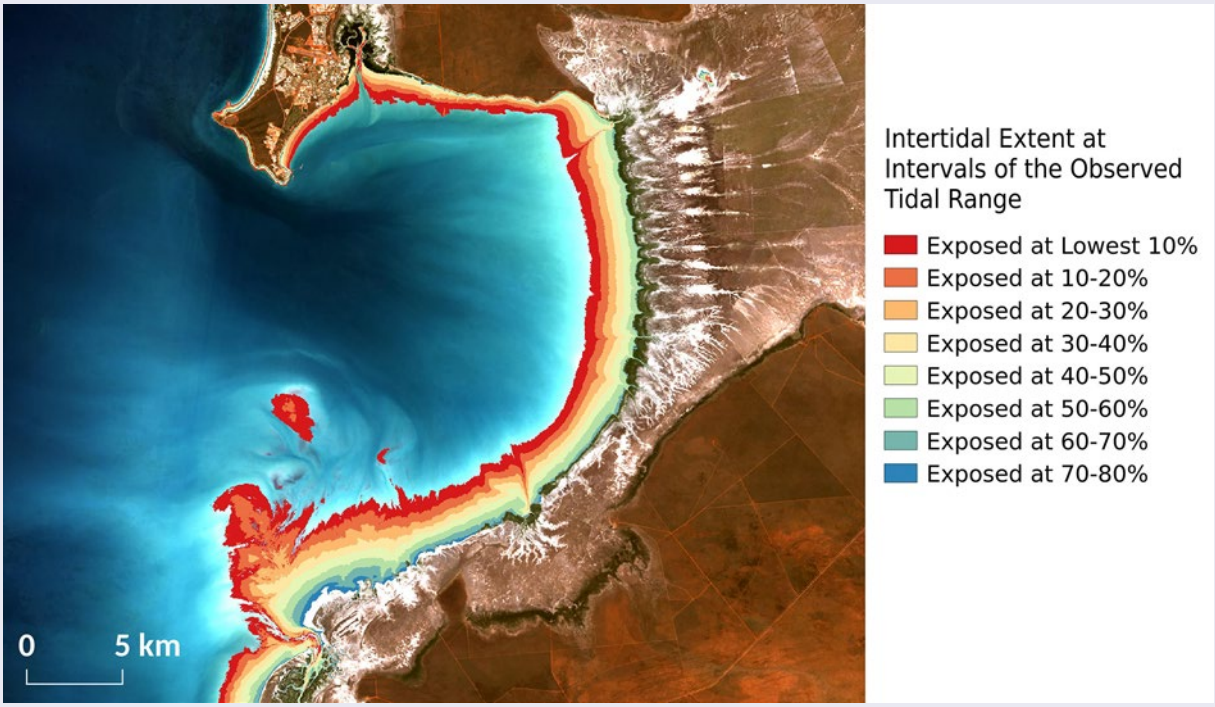
Modelled tidal heights for Landsat observations shown in relation to the full tidal model. Note how the observed tidal range (HOT-LOT) differs from the full tidal range.



Source: Stephen Sagar, Geoscience Australia.

Figure 2.2 Roebuck Bay Example

Example of the ITEM Relative Extents Model showing the intertidal flat extents and topography at Roebuck Bay, WA, a site of ecological significance for migratory shorebirds.



Source: Stephen Sagar, Geoscience Australia.

The modelling approach enables the creation of a continuous intertidal extent map for the full Australian coastline. Since tidal extent is a function of the tidal height and the shoreline elevation, the data set also provides relative topography, and thus a framework to derive coastal elevation models. This is an example of how EO products can act as a complement to existing data, such as ship-based bathymetry, enabling a more realistic representation and a deeper understanding of Australia's vast coastline.

The knowledge provided by ITEM can contribute to a broad range of applications, including environmental monitoring applications for migratory bird species, habitat mapping in coastal regions, hydrodynamic modelling, and examining the geomorphology of the intertidal zone.

2.2 Processing Methodologies

Although some image processing exercises will not include all stages, processing of EO imagery can be considered within the context of three broad functional stages:

- descriptive interpretation—this is generally an initial process of visually delineating and identifying features or patterns in an image. Digital image processing can be useful during this stage to improve feature contrast and colour (see Sections 4 and 5) or modify image geometry (see Section 7 and Volume 2B);
- exploratory data analysis—involves using statistical tools to identify features and patterns in the image. This stage produces results such as land cover maps from EO imagery; and
- modelling—relates image values to data from other sources and accesses a wide variety of transformations to determine the relationships between different data types.

Table 2.1 uses some example applications to illustrate the relationship between different processing stages and spectral/spatial data attributes. Image processing is primarily used for exploratory data analysis but draws on both colour and shape (that is spectral and spatial, data and tools) and can be used to prepare data for both descriptive activities and modelling.

This Volume introduces image processing methods in terms of these three stages. Sections 4 to 6 cover the basic techniques that are relevant to visual image interpretation, including image enhancement, geometry, display and hardcopy. The image analysis topics of statistics, classification, and segmentation are described in Sections 8 to 10. Various image transformations, which may be used for interpretation, analysis or modelling, are detailed in Volume 2C. The detailed processing involved with more advanced functions of image rectification and classification are covered in Volumes 2B and 2E respectively. Processing of multiple images and their integration with other datasets is discussed in Volume 2D.

Table 2.1 Image processing stages

Stages of processing spectral and spatial data can be considered as description, analysis and modelling. Image processing is primarily used for exploratory data analysis and involves consideration of both the spectral and spatial attributes of image data.

Data Attribute		Stages of Processing		
Attribute	Example	Description	Analysis	Modelling
Colour	Spectra	Compiling spectral libraries, eg. AU-SPECCHIO	Compiling image time series, eg. GA data cube (see Section 7.4.2)	Quantifying identified image features, eg. oil spill analysis (see Volume 3B)
Imagery	Spatial Patches	Identifying major spatial features, eg. labelling land cover categories	Exploratory data analysis methods, eg. analysing area and distribution of land covers over time	Developing mathematical model to predict image values, eg. bio-optical modelling of water constituents (see Volume 3B)
Shape	Spatial Patterns	Interpreting spatial patterns, eg. aerial photography interpretation	Spatial analysis methods, eg. mathematical morphology	Inferring three-dimensional structure, eg. stereology/tomography or lidar modelling of forests (see Volume 3A)

2.3 Modelling

In many application areas, physics-based models are being developed and standardised to relate image values to measurable surface properties. To ensure that the value of EO imagery is optimised, such methods need to be fostered and endorsed by the wider scientific community. Processing methodologies for specific application areas are detailed in Volume 3.

This section overviews aspects of modelling that are relevant EO analyses. Subsequent volumes contain examples of specific models derived using EO datasets in various applications. The following text and sub-sections provides relevant background information on modelling terminology, prerequisites, and constraints.

2.3.1 Modelling Overview

Models describe various aspects of the physical world, including interactions and dynamics, using mathematics. They are typically applied to current datasets, which quantify current conditons, to predict future datasets, which should infer future conditions. Mathematical models can be developed which relate known real world properties to EO measurements. The modelled relationship can then to used to infer one entity from the other, either as a:

- forward model—to compute likely EO measurements from known real world properties; or as an
- inverse model—to compute likely real world properties from EO measurements (see Section 2.3.2).

A range of mathematical and algorithmic tools are used to develop appropriate models. Typically simpler, more robust models are less flexible and transferable than complex and computationally intensive models. Mathematical models are generally grouped into three categories:

- empirical or analytical models—use simple stochastic relationships, such as regression curves, between input and output data (see Section 2.3.3);
- semi-empirical or semi-analytical models—use *a priori* knowledge to determine a relationship between input and output data (see Section 2.3.4); and
- physics-based models—use knowledge about physical relationships between input and output data to define a relationship (see Section 2.3.5).

The unique perspective of EO imagery offers an attractive opportunity to model different aspects of planet Earth using appropriately-scaled data. A wide variety of models have been developed to interact with EO datasets for specific calibration, such as corrections for atmospheric and BRDF effects (see Section 3.3.2 and Appendices 1 and 2) and myriad applications, including phenology, water optics, carbon cycling, and climate studies. Some of these models are discussed in the context of specific application areas in Volume 3.

2.3.2 Forward and inverse models

Forwards models use data from the 'real world' to estimate a set of 'modelled' data, which may relate to a future time or state, or an alternate measurement mode. For example, in the context of hydrological EO, forward models start with known water properties and determine the upwelling radiation that would be measured by EO sensors.

Inverse models essential invert the forward modelling process, that is, they estimate 'real world' data from a set of 'modelled' data. In the context of EO imagery, inversion algorithms operate on a per pixel basis to determine the best combination of input parameters that would explain the spectral measurements in an EO image (Dekker *et al.*, 2001).

2.3.3 Empirical or analytical models

By definition, empirical models are based on data, rather than theory. Relevant measurements (or observations) of the 'input' variable and 'output' variable are used to develop, calibrate and test the relationship between them. The resulting model correlation enables an input value to be converted to an output value. Such models need to accurately relate the input and output variables and define the reliable constraint range(s) for input values. Empirical models for the remote sensing of surface waters, for example, relate the optical properties of water to their observed optical measurements in EO imagery (see Volume 3B).

2.3.4 Semi-empirical or semi-analytical models

Like empirical models, semi-empirical or semi-analytical models are based on data, but involve assumptions, approximations or generalisation to simplify the modelling process. *A priori* knowledge, for example, may be used to define the relationship between input and output data. In the context of EO, this may involve selecting the most appropriate spectral bands to develop a model. However, while such tailoring of data may improve model accuracy, the resulting modelling relationships may be less transferable to other locations and other sensors.

2.3.5 Physics-based models

Physics-based models use known physical laws to define the relationship between input and output variables. Physics-based models of the water column, for example, attempt to account for the energy transfer process through both the atmosphere and the waterbody (see Volume 3B). These models are more sophisticated, flexible and transferrable than empirical or semi-empirical models, but are signficantly more computationally intensive and require more detailed parameterisation data.

Essentially, all models are wrong, but some are useful
(George E. P. Box)

2.4 Further Information

Image Processing

The Pixel Page: *http://home.earthlink.net/~castleman/*

Image Processing Systems

ERDAS Imagine: *http://www.hexagongeospatial.com*

ERMapper: *http://www.hexagongeospatial.com*

ENVI: *http://www.harrisgeospatial.com*

TNTmips: *http://www.microimages.com*

2.5 References

Dekker, A. G., Brando, V., Anstee, J. M., Pinnel, N., Kutser, T., Hodgenboom, E., Peters, S., Pasterkamp, R., Vos, R., Olber, C., and Malthus, T. J. (2001). Imaging Spectrometry of Waters. Chapter 11 in '*Imaging Spectrometry: Basic Principles and Prospective Applications*' (Eds: F. van der Meer, and S. M. de Jong). Springer, Netherlands.

Egbert, G.D., Erofeeva, S.Y., 2010. *The OSU TOPEX/Poseiden Global Inverse Solution TPXO* [WWW Document]. TPXO8-Atlas Version 10. URL *http://volkov.oce.orst.edu/tides/global.html*.

Egbert, G.D., Erofeeva, S.Y., 2002. Efficient Inverse Modeling of Barotropic Ocean Tides. *J. Atmospheric Ocean. Technol.* 19, 183–204. doi: *https://doi.org/10.1175/1520-0426(2002)019%3C0183:EIMOBO%3E2.0.CO;2*

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* 17, 1425–1432. doi:*10.1080/01431169608948714*

Sagar, S., Roberts, D., Bala, B., Lymburner, L., 2017. Extracting the intertidal extent and topography of the Australian coastline from a 28 year time series of Landsat observations. *Remote Sens. Environ.* 195, 153–169. doi:*10.1016/j.rse.2017.04.009*

3 Calibration

Various characteristics of EO imagery are introduced in Volume 1. While an observation recognises both the existence of the observed object and the action of recording something about it, a measurement provides a scaled record of such interactions and one that can be directly compared with similarly scaled measurements (see Volume 1A—Section 1). In order to relate EO observations to physical measurements of a target, images need to be calibrated to a known standard. Alternatively, to enable inter-comparisons, multiple images can be calibrated to a relative standard.

Section 3.1 provides an overview of calibration in the context of EO imagery while Section 3.2 reviews the need for calibration of remotely sensed imagery in terms of spectral, spatial, radiometric and temporal factors. Specific processing procedures that can be used to calibrate EO imagery are introduced in Section 3.3.

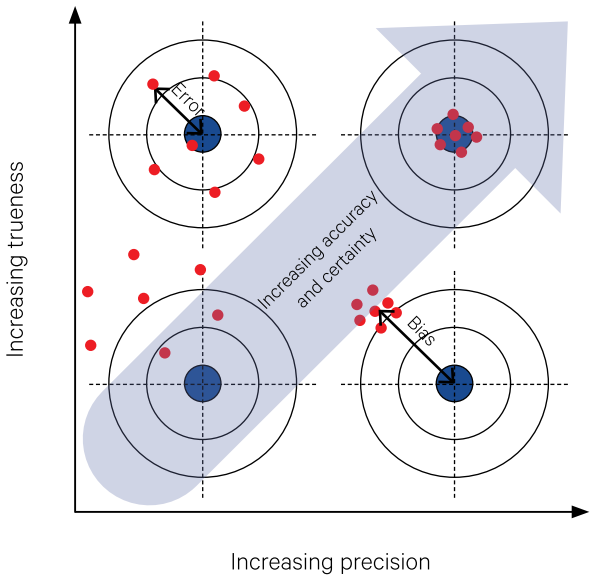
3.1 About Calibration

At this point, it is worth pondering the fact that all measurements of physical quantities are subject to uncertainties. Accordingly, it is inevitable that repeated measurements of the same quantity would yield variable results. Thus, every measurement implicitly comprises both a numeric value to estimate the quantity and a degree of uncertainty to indicate the inherent variability of the estimate. An accurate estimate of 'truth' generally involves repeated measurements of some traceable reference standard (see Figure 3.1; also ISO 57251).

When you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.
(Lord Kelvin, 1883)

Figure 3.1 Precision versus trueness

Trueness—closeness of agreement between the average value from a large series of test results and an accepted true value. Precision—closeness of agreement between independent measurements of a quantity under the same conditions.



Background image: Landsat-5 TM image of Lake Eyre acquired on 28 March 2011, displayed using bands 5,4,2 as RGB. A slanting striping pattern, resulting from imperfect calibration of the 16 TM detectors, is visible in the central water features. **Source:** Norman Mueller, Geoscience Australia

As illustrated in Figure 3.1, to appreciate the uncertainty inherent in all measurements, it is necessary to distinguish between:

- Trueness—closeness of agreement between the average value from a large series of test results and an accepted true value; and
- Precision—closeness of agreement between independent measurements of a quantity under the same conditions.

For EO data, the Committee of Earth Observation Satellites (CEOS) and Malthus and Li (2015) define relevant terminology as:

Calibration—process of quantitatively defining the responses of a system to known, controlled signal inputs;

Cross-Calibration—process of transferring the well-known radiometric calibration of one satellite sensor to another poorly-calibrated sensor via near-simultaneous imaging of a common ground target;

Traceability—property of a measurement result, which relates that result to a stated metrological reference through an unbroken chain of calibrations of a measuring system or comparisons, each contributing to the stated measurement uncertainty;

Uncertainty—parameter that characterizes the dispersion of the quantity values attributed to a measured mean, based on the information used;

Vicarious Calibration—techniques that make use of natural or artificial sites on the surface of the Earth for post (launch or flight) calibration of airborne or space-borne sensors; and

Validation—process of assessing, by independent means, the quality of the data products derived from the system outputs.

Being acquired by remote sensors, EO data necessarily involve one or more levels of inference or indirection (see Volume 1A—Section 1). As detailed in Volume 1, mechanisms for image acquisition and sampling inevitably produce distortions within the resulting data. These distortions can occur in terms of four intrinsic attributes:

- spectral—specific wavelengths recorded;
- spatial—location, area and scale of record;
- radiometric—energy levels recorded; and
- temporal—time of record.

Examples of image distortions include wide scan angles (resulting in larger ground pixels at the edges of a scan than in the centre) and varying Sun and observer positions (resulting in brightness variations which are unrelated to target brightness). Such distortions can impact interpretation, analysis and modelling of individual images and sets of imagery.

Typical processing scenarios include:

- single images, where classification and other image processing is predicated on consistent values from similar target surface covers;
- mosaics of images, which are intended to produce a consistent composite image for spectral image processing; and
- multiple images, such as time series datasets or imagery acquired by different sensors, where changes in the target surface are being detected between images and should not be confused with brightness variations resulting from imaging artefacts.

In each case, the underlying processing algorithms assume consistency between image values, and variation due to imaging artefacts will violate this assumption and undermine the credibility of analysis results. Thus, before EO data can be analysed for particular applications, the recorded observations need to be calibrated to some acceptable standard.

Calibration can be either:

- absolute—relates the image data range to a physically-measured parameter; or
- relative—rescales the radiance measurements of one or more images to match those of a reference image.

The primary objective in EO image calibration is to ensure a spatial and temporal consistency, or 'stationarity', in which a specific target with a particular reflectance will be the same at all locations and in all images within a given dataset. Image calibration is essential when multiple images are being compared digitally (see Volume 2D). Obviously, the degree to which this needs to be accomplished depends on the specific application. Ultimately, the value of spending time and effort on image correction depends on the value of the final product, as judged by the client, since the total cost of any image product needs to reflect its value to the end user.

As discussed in Volume 1B—Section 9, from an economic perspective the analysis of EO imagery involves a range of costs, including project planning and management, image and ancillary data acquisition, data processing and integration, and product presentation. When a project aims to deliver a low-cost product, the benefits of correcting for effects such as image brightness variations or atmospheric correction may not be justified (Song *et al*., 2001). Conversely, where product accuracy is crucial, the effort required to correct for image distortions would be appropriate. Given the extensive range of approaches and methods that can be used to model or correct for image distortions, the 'best' method will vary for each specific application, its intended use and the 'value' of the processing results to the user.

The importance of EO image calibration is increasingly recognised and significant effort has been expended in recent years to standardise procedures and establish reference sites for this purpose. In Australia, the TERN AusCover Remote Sensing Data Facility has compiled *Good Practice Guidelines* for calibration and validation activities of EO data products (Held *et al.*, 2015). Such guidelines are essential for the generation of biophysical and geophysical products from EO datasets, especially when the underlying data for a given product is derived from a variety of sensors.

3.2 Optical Image Distortions

The following sub-sections consider image distortions in terms of:

- spectral (see Section 3.2.1);
- spatial (see Section 3.2.2);
- radiometric (see Section 3.2.3); and
- temporal (see Section 3.2.4) factors.

While calibration processing is traditionally discussed in terms of radiometric and geometric corrections, the following structure has been adopted to emphasis both the sources of image distortions that are relevant to EO and the need to select appropriate calibration algorithms. In practice, many processing systems and data suppliers provide packaged solutions for image calibration; however, an awareness of the inherent assumptions underlying the convenience of such options is essential for maximising the value and accuracy of EO products.

3.2.1 Spectral

Spectral resolution, density and extent of EO datasets are introduced in Volume 1B—Section 1.1. In the context of spectral characteristics, calibration relates to standardising the reflectance observed by a remote sensor to account for radiance variations resulting from changes in atmospheric conditions and/or illumination and viewing interactions. Note that the term spectral calibration is used to ensure that an optical sensor measures reflectance within a defined range of wavelengths, a process that needs to be applied to the sensor before data acquisition.

The basic geometry used to describe illumination and viewing positions in EO is illustrated in Figure 3.2 (see also Volume 1B—Section 3.5). Ideally, the Earth's surface would be scanned with the illumination and viewing positions directly overhead each image pixel, with no interference from the atmosphere or surface topography. The calibration process attempts to convert EO data to this ideal, or at least reduce the pixel-to-pixel variation that results from atmospheric and illumination/viewing geometry factors. Such calibration is particularly important when multiple image acquisitions are being analysed.

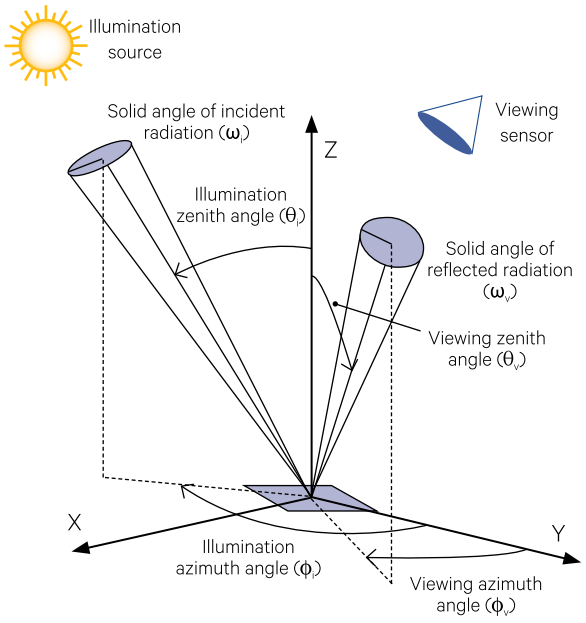
Optical, shortwave and thermal image data from airborne and satellite platforms have varying degrees of angle-dependent brightness variation which change with Sun position, atmospheric conditions, land surface type and sensor characteristics (see Volume 1B). If left uncorrected, these broad level variations can make it difficult to use these data with standard image processing and interpretation methods.

The impact of these brightness variations is most obvious in scanners with wide scan angles (for example, airborne scanners can have scan angles (field of view: FOV) up to ±90°). In imagery acquired from a nadir-viewing position, such brightness variations derive from the:

- atmospheric component phase functions (see Appendix 1); and
- angle-dependent reflectance from the surface of Earth (or the Bi-Directional Reflectance Distribution Function: BRDF; see Appendices 1 and 2 and Volume 1B—Section 3).

Figure 3.2 Geometry of illumination and viewing positions

Basic geometry of illumination source (i) and viewing sensor (v) relative to a target surface



Brightness variations can also occur in imagery acquired by off-nadir viewing angles. In this case, the sensor may view the sides of ground features, such as the woody structures of trees or shadows beneath the canopy. This effect is emphasised for foliage with vertically-oriented leaves (such as *Eucalyptus* trees).

The atmospheric contribution to scan line brightness variations results from two factors:

- varying atmospheric path lengths at different parts of the scan; and
- variations in the atmospheric scattering phase variation (the composite phase function of the atmospheric constituents), which describes the way radiation is scattered out of a volume relative to the direction it enters.

In optical remote sensing, air molecules and aerosols account for much of the scattering, and the effects are greatest at shorter wavelengths. For example, if the aerosol atmospheric turbidity is high, the atmospheric brightening in the blue-green wavelengths will be visible and significant in most remotely sensed data. In thermal image bands, the angular variation is predominantly due to the changing path length; thus the major atmospheric effect is due to water vapour along the path.

These atmospheric and surface effects are present in all satellite data, as well as in aerial photography and scanner data, depending on the variation in the phase angle between the Sun, target and sensor among the pixels in the resulting image. The impact of topography on EO values depends on scale (or pixel size), leading to quite complex image brightness variations (see Appendix 1).

The land surface itself contributes to the broad level brightness variations common to all EO images. The base 'colour' and brightness is a function of the spectral properties of the material types making up the scene (e.g. reflectance and transmittance from leaves, or soil grain mineralogy) plus structural effects which give rise to angular variations (see Volume 1B). The angular effects can be ascribed to three main factors:

- volume effect—as the relative illumination and view angles vary within an image, changes in radiation path lengths and extinction on complex surfaces result in a Volume BRDF that has some similarity to the variations induced by the atmospheric phase function. This effect depends on surface structure and in vegetation it varies with leaf density, leaf angular distributions and total leaf area;
- occlusion (or hotspot) effect—is a more specific effect induced by the fact that the shadows cast by the Sun represent parts of the surface that are occluded (not visible to the Sun). Some of these occluded areas are not visible to the sensor either and these can be considered as those shadows that would be cast by the Sun if it could be located in both the viewing and illumination positions. As the Sun and sensor positions move, different surface areas become occluded to one or both positions, resulting in image brightness variations. For example, when the viewing and illumination positions are aligned, no shadows are visible and the imaged scene appears bright; conversely, when the viewing and illumination angles are 'opposed' many shadows are observed, resulting in lower image brightness; and
- specular (or glint) effect—is most pronounced on water surfaces. This refers to the surface 'mirror' (or Fresnel) reflectance in which the radiation is usually unaltered by the surface material from which it is reflected. On land surfaces it is a composite of reflections diffused from facets of varying angular positions and specularity. In Australia, *Eucalyptus* leaves are especially specularly reflective due to their waxy coating. On water, glint is probably the major component of the BRDF and the hotspot effect does not occur.

Atmospheric and surface effects interact to produce a composite effect that varies with Sun, target and sensor geometry. The effects are more obvious in aerial photography and video images where the central perspective means that at high Sun angles the hotspot and specular points are within the angular radius of the frame. For airborne scanners and satellite borne scanners it is possible to avoid the brightest hotspot and glint points by careful consideration of the scan geometry. However, since the variations are significant for all angles, the effects cannot be ignored in any image.

The scene and atmospheric brightness variations are all minimised if the sensor scans at right angles to the principal plane (the plane formed by the vectors joining the Earth's centre to the sensor and to the Sun). For an airborne scanner, this means flying in a level attitude (neutral pitch), with the Sun directly in front of, or behind, the flight line. If the hotspot or specular points need to be avoided at high Sun times, the scanning sensor may be tilted forward or backward. However, these 'avoiding' strategies are not necessarily easy to implement with acquisition of airborne imagery.

EO imagery is commonly pre-processed to correct for atmospheric effects, and may also be corrected for BRDF (Schaaf *et al.*, 2002) and topographic effects (Richter *et al.*, 2009). The impact of these corrections is illustrated in Section 3.3.2 below. Appropriate processing techniques can also correct for missing and misaligned data in an image (see Volume 2C). When significant portions of an image are reconstructed by image processing, the accuracy of the 'infilled' data should be verified independently. Specific procedures used in image calibration to account for brightness variations are considered in Section 3.3.2 and Appendix 1.

3.2.2 Spatial

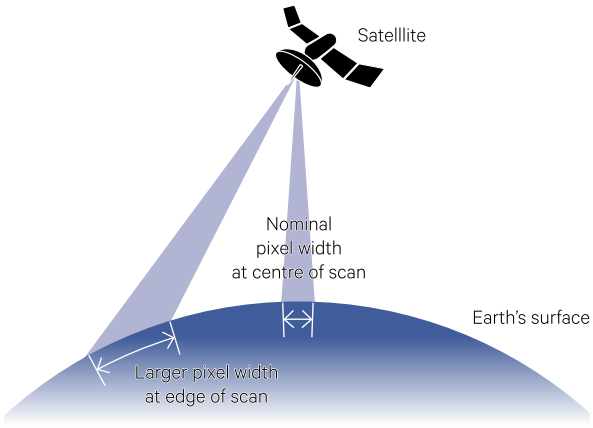
Factors determining the spatial resolution, extent and density of EO imagery are detailed in Volume 1B—Section 1.2. Geometric distortions are inherent to EO imagery due to the image acquisition process (see Volume 2B for details). These distortions broadly result from the composite impact of:

- variations in platform height, direction, speed and stability (most relevant to airborne imagery);
- variations in the viewing position of the sensor when acquiring each pixel in the image;
- scanning mechanism of the sensor (including errors in operation); and
- Earth curvature and rotation.

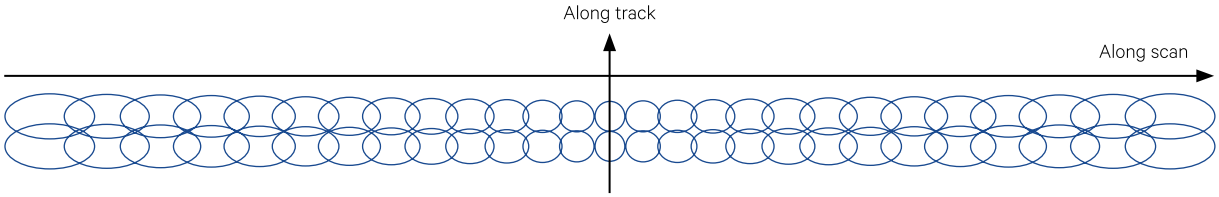
The net effect of these factors is image data with pixels of varying size so that each pixel does not represent the same ground area. This is illustrated in Figure 3.3 for the pixel geometry of AVHRR imagery. In addition, as described in Volume 1B—Section 2, characteristics of the imaging optics can introduce artefacts. Such artefacts mean that not all parts of a ground pixel area are imaged uniformly and a given image pixel may include contributions from neighbouring pixel areas.

Figure 3.3 AVHRR pixel geometry and correction of spatial distortions

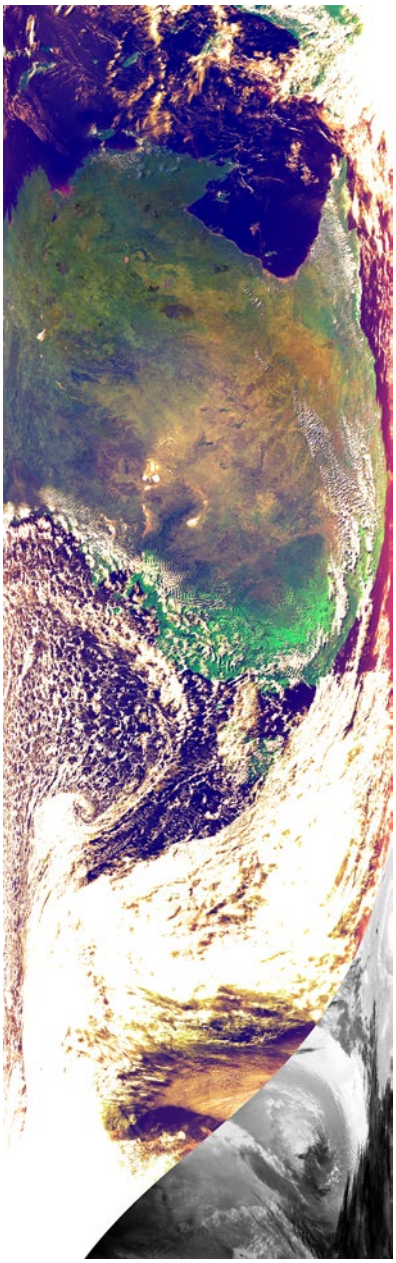
a. For wide scan angles, as occurs with AVHRR, pixel size increases away from the centre of an image swath. This effect is further exaggerated by Earth curvature, especially for sensors imaging wide swaths such as AVHRR.



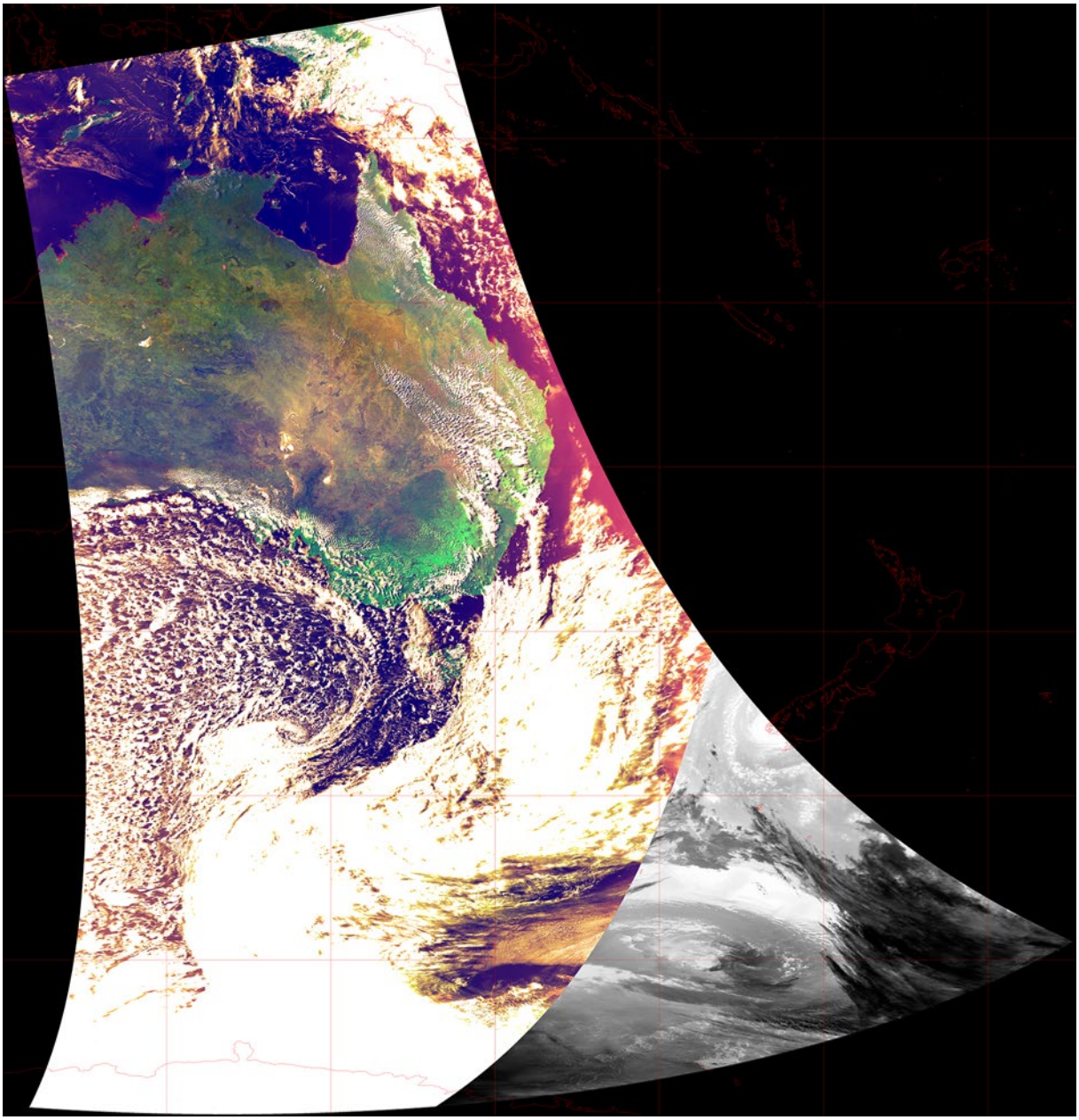
b. When looking vertically down to Earth at the centre of each scan line (or nadir), the sensor detects radiance from a circular ground area. As the sensor looks away from the image centre (off-nadir viewing), a larger ground area is imaged in both the along-scan and along-track directions. A greater increase in imaged area occurs in the along-scan direction, resulting in the optical pixel shape changing from a circle to an ellipse. Near the edges of the image swath, the ground area being recorded as a single pixel is significantly larger than at the centre, resulting in distinct geometric distortions in such imagery (see Volume 2B). Note that there are over 2,000 pixels per line in AVHRR imagery; here the number is significantly reduced for simplicity.



c. This example shows the original image swath acquired by NOAA-19 on 3 September 2017 before and after geometric rectification. For daytime acquisitions, two reflectance channels are used to create a false colour composite. Once the Sun has set, a thermal channel is used, resulting in the greyscale section in the southeast of this image.



d. Rectified Image swath for image shown in Figure 3.3c resampled to 0.01° grid (~1 km pixel size) and aligned with geographic projection.

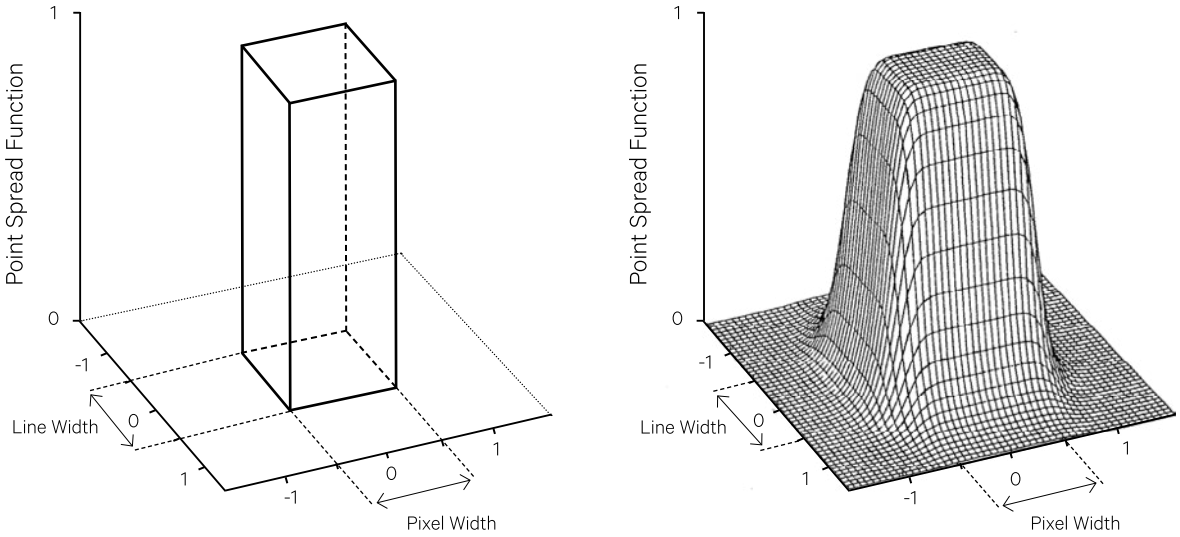


Adapted from: a and b. Breaker (1990), c and d. Sourced from Edward King, CSIRO

Figure 3.4 Example of Point Spread Function

a. Idealised PSF—the radiance measured for a single pixel should form a stepped function (Cracknell, 1998)

b. Modelled PSF for AVHRR pixel—X and Y axes in this figure indicate the central pixel covering the area demarked by elements -0.5 to +0.5 and lines -0.5 to +0.5. Quite clearly, a significant proportion of the radiance recorded for this single AVHRR image pixel originates from the adjacent pixels.



Adapted from: Mannstein and Gesell (1991),

One measure of imaging fidelity is the Point Spread Function (PSF; see Volume 1B—Section 2.1.1.1). The PSF of an 'ideal' pixel is compared with the modelled PSF for an AVHRR pixel in Figure 3.4, showing the extent to which neighbouring pixels can contribute to the recorded radiance values of the central pixel.

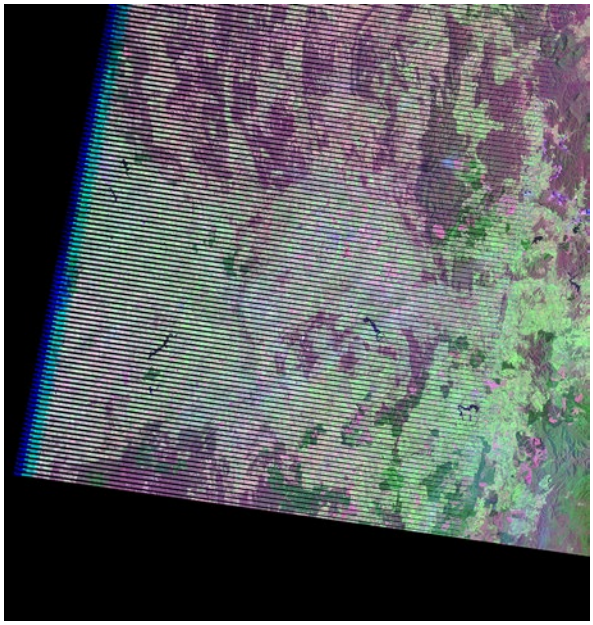
Spatial distortions in images can change their shape, orientation, scale and internal proportions, and do not allow them to be accurately registered with other spatial datasets, including other EO images from the same or different sensors. These distortions are most obvious in images acquired by scanners with wide-scan angles (see Figure 3.3).

Some distortions also result from scanner malfunction, as occurred with the failure of Landsat-7's ETM+ Scan Line Corrector (SLC) in 2003. The SLC corrects the action of the scanner for the motion of the satellite. With the SLC inoperable (termed SLC-Off) the scan lines are not parallel, but form a zig-zag pattern, leaving regular null-data gaps that widen from the centre of the image to the eastern and western edges of the swath (Barsi *et al.*, 2007; USGS, 2017). The resulting missing data is estimated to comprise 22% of the scene area. Two examples of images impacted by this anomaly are shown in Figure 3.5.

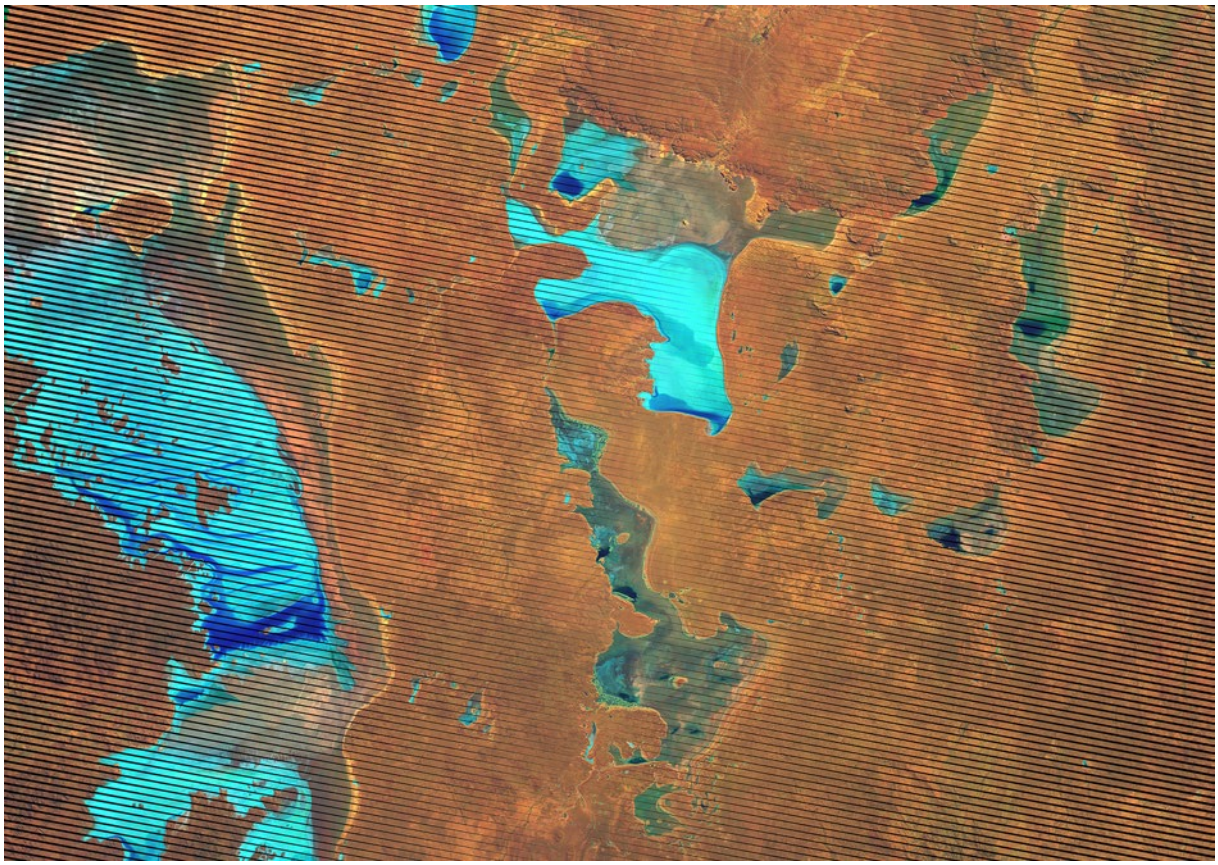
Figure 3.5 Effect of Landsat-7 ETM+ SLC malfunction

These images shows the pattern of missing data that characterised all Landsat-7 ETM+ imagery following the failure of the scan line corrector (SLC) on 31 May 2003.

a. Western portion of Landsat-7 image over Bathurst, NSW, (path 90 row 83) acquired on 2 January 2012.. Bands 5, 4, 2 and RGB.



b. Landsat-7 image over Lake Gairdner and Woomera, SA, acquired on 10 July 2017. (5,4,2 as RGB).



Source: a. Tony Sparks, Icon Water, b. Norman Mueller, Geoscience Australia

The geometry of an EO image can be modified to remove known distortions and register the image to a different (map or image) projection (see Figure 3.3). This is particularly important when the imagery is to be used with GPS or GIS data, or when two or more images are to be compared digitally. Image distortions in airborne data that result from aircraft instability can be avoided using stabilised platforms and inertial navigation systems. Geometric distortions in satellite borne sensors are well-defined and routinely corrected by data suppliers before distribution of data products.

The accuracy of any geometric rectification exercise is critical to the locational accuracy of image features. Whenever multiple images are to be processed as a single dataset, particularly for image time series, image registration needs to be accurate to ensure that detected differences are due to ground changes rather than misregistration (see Volume 2D). Ideally, a spatially calibrated image would represent a regular grid placed on the Earth's surface. While in reality this ideal is impossible to achieve, geometric correction should produce an image in which specific locations can be confidently situated, and pixels in different image acquisitions can be reliably compared. Specific processing procedures for geometric rectification of EO imagery are discussed in Sections 3.3.1 and 7, and Volume 2B.

3.2.3 Radiometric

The concepts of radiometric resolution, density and extent are introduced in Volume 1B—Section 1.3. Radiometric calibration accounts for variations in the measured reflectance due to imperfect operation of sensor components (see Volume 1A—Section 13). These variations may be:

- random—data is not acquired correctly, or at all, for some pixels, creating missing or spurious values for individual pixels, sequences of pixels, or whole lines in the image. Missing values can be interpolated from neighbouring pixels for image continuity, but the interpolated values are merely cosmetic. Spurious values may be more difficult to detect, but various algorithms have been proposed to correct specific datasets (see Sections 3.3.2 and 8.3 below and Volume 2C).
- systematic—miscalibration between multiple detectors in a sensor can result in visible striping artefacts in the acquired image. Examples of this radiometric distortion include Landsat TM 16-line horizontal striping and ALI vertical striping (see Figure 3.6).
- gradual—aging may result in deterioration of individual detectors over time, resulting in radiometric variations in multi-date imagery from the same sensor.

Pixel digital numbers, which are assumed to be proportional to millivolts of radiation, are related to measured radiance by an appropriate gain factor for each imaged channel:

$$L_j = G_j dn_j$$

where

L_j is the radiance data in band j ($W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$);
G_j is the gain factor for band j; and
dn_j is the digital number.

Thus, to both maintain precision and avoid overflow of saturated pixels, image data is supplied in (SI) units of $W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$ multiplied by the gain factor (see Volume 1A—Section 13.3).

EO sensors are routinely calibrated prior to launch (for satellites), flights (for airborne) or field use (for ground-based sensors). Sensing instruments can, however, change during operation, due to a range of factors, including aging or failure of components. Since the operation of satellite sensors can vary after launch and/or degrade in orbit, their radiometric performance is routinely checked post-launch, via well-established on-board calibration processes, with the aid of both on-board calibrators and ground reference points (Malthus and Li, 2015).

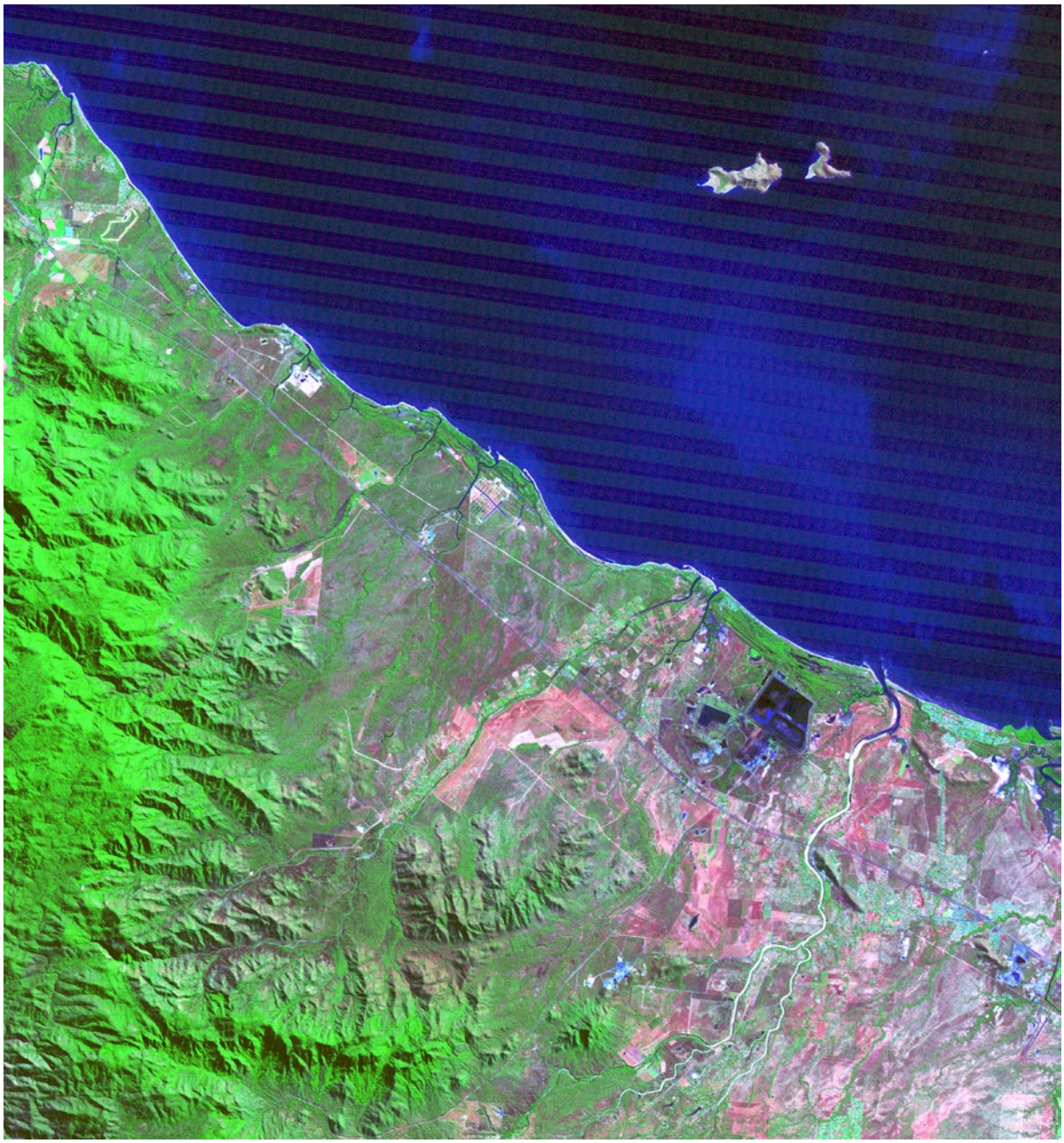
For example, the on-board calibration system for Landsat-7 ETM+ relies on three components to calibrate optical bands to within 5% absolute uncertainty (Markham *et al.*, 2004):

- Internal Calibrator (IC)—two lamps used to check detector efficiency against a dark, reference background after each scan line;
- Partial Aperture Solar Calibrator (PASC)—auxiliary optics that became saturated in most bands, so is no longer used for calibration; and
- Full Aperture Solar Calibrator (FASC)—a deployable diffuser panel that provides absolute calibration against the rising Sun.

This calibration system yields reliable estimates of the uncertainty of the absolute radiometric calibration associated with each image channel (see Figure 3.7). Once in orbit, Landsat-7 ETM+ was briefly cross-calibrated against Landsat-5 TM, since initially both satellites were in 'tandem' orbits and thus acquired imagery over the same targets only 10–30 minutes apart (Teillet *et al.*, 2001). Various research teams also regularly perform vicarious calibration of ETM+ and TM reflective bands (Thome *et al.*, 2004). A detailed description of the more recent Landsat-8 calibration is referenced in Section 3.5.

Figure 3.6 Systematic radiometric miscalibration

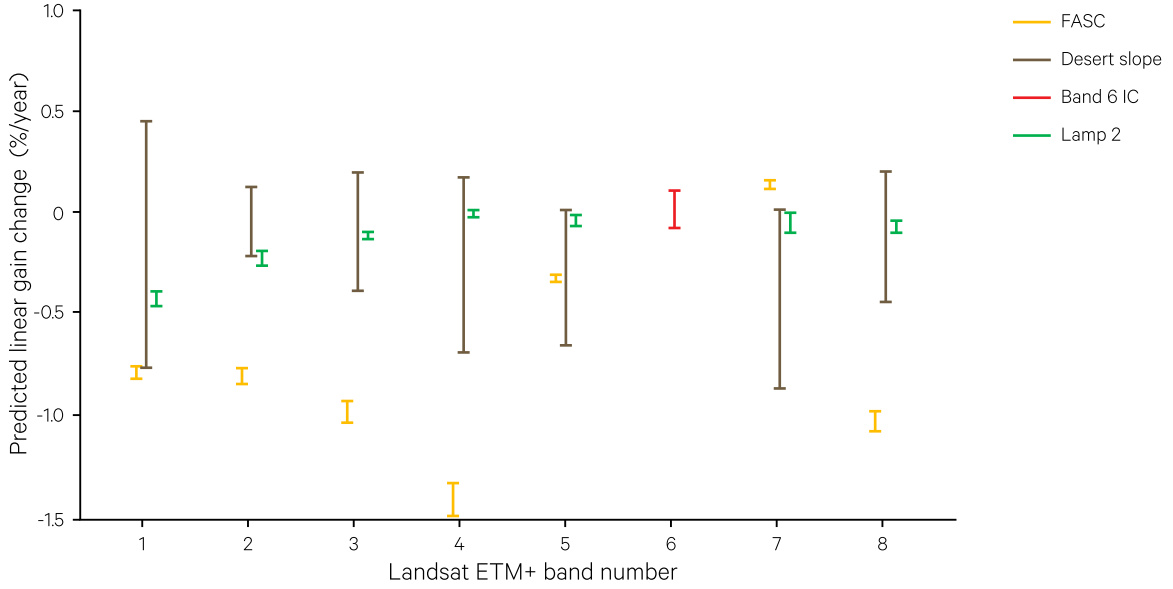
Landsat-5 TM image acquired on 31 July 2003 over northern fringe of Townsville and Paluma Range National Park, Queensland, displayed using bands 5, 4, 2 as RGB. This image shows a striping pattern that was typical of the 16 detector swath in this sensor. In this case, the striping is most visible in the ocean.



Source: Tony Sparks, Icon Water

Figure 3.7 Estimated annual % gain change per year for Landsat-7 ETM+ calibrators

Error bars are the 95% confidence limits in the slope estimate. In the reflective bands, the Full Aperture Solar Calibrator (FASC) gain change is clearly different than the lamp 2 and the desert sites. The lamp 2 change is significant, but could be due to changes in the Internal Calibrator system. The desert sites slopes (straight average) are not significantly different than zero, so indicate there is no change in gain. However, they do not preclude the small changes indicated by lamp 2.



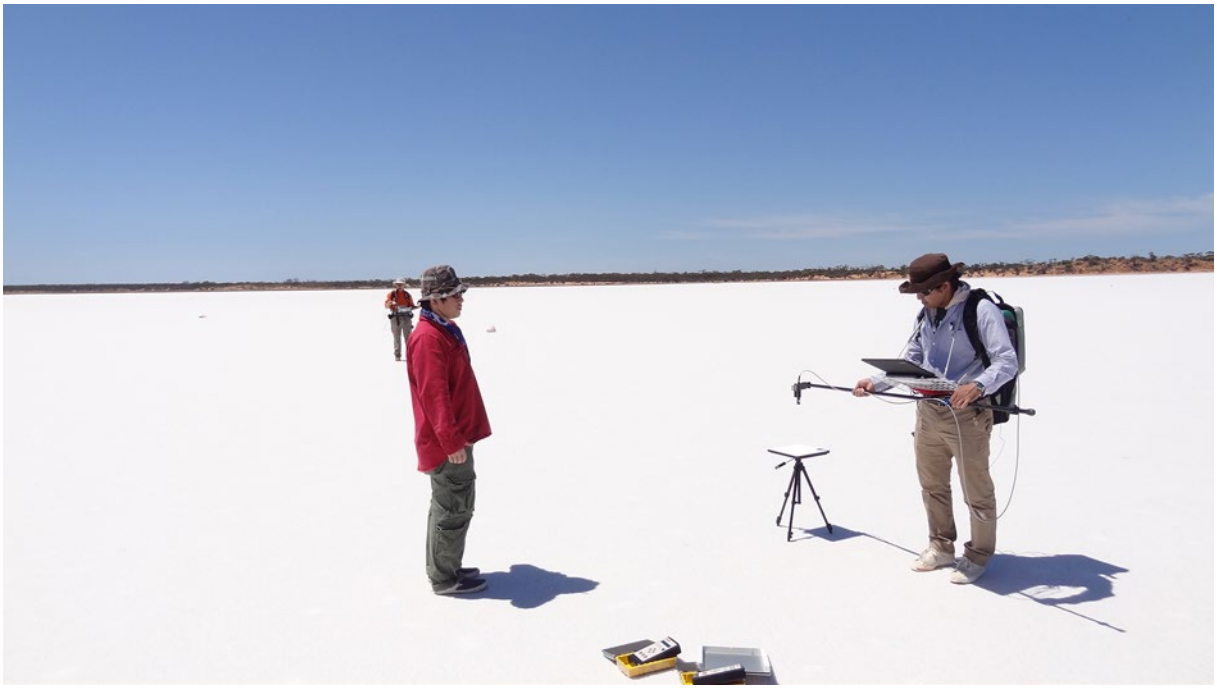
Source: Barsi *et al.* (2007) Figure 11

Most satellite sensor calibration systems have proven to be consistent within individual images and can maintain consistency for several months, but degradation can occur over longer time periods (Thome *et al.*, 1997). Additionally ground-based references (Slater *et al.*, 1987; Slater *et al.*, 1996) and celestial bodies (Kieffer and Wildey, 1985; Kieffer *et al.*, 2003) are used to calibrate satellite sensors over time. Ideal reference points are spatially uniform, spectrally invariant, bright, flat, diffuse surfaces, preferably situated in elevated, inland, arid locations (Teillet *et al.*, 2007). As illustrated in Figure 3.8, dry salt lakes offer suitable reference sites in Australia (Malthus and Li, 2015), and deserts and glaciers provide similar opportunities elsewhere. Some of the approaches used for correcting radiometric distortions are outlined in Section 3.3.2 and Section 8.3.

Figure 3.8 Validation sites for vicarious sensor calibration

a. Lake Frome, SA

b. Lake LeFroy, WA



Source: a. Sue Campbell and Guy Byrne, CSIRO, b. Chris MacLellan, University of Edinburgh

3.2.4 Temporal

To reliably analyse the trends in EO time series imagery, all pixel values must be directly comparable. In addition to correcting the spectral and radiometric distortions mentioned in Sections 3.2.1 and 3.2.3 above, all multi-temporal imagery requires pre-processing to correct for:

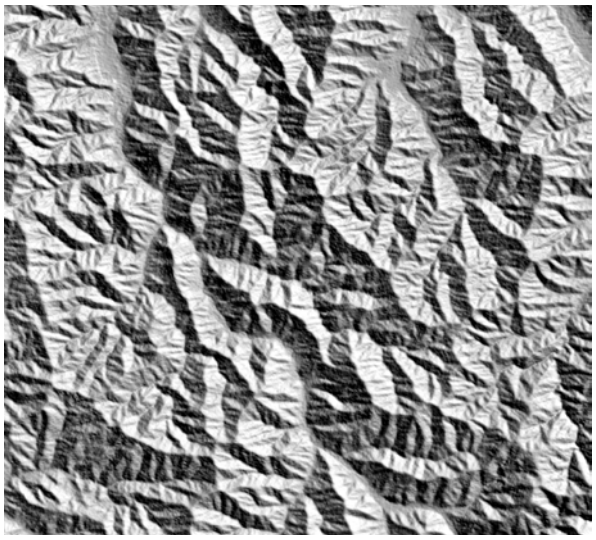
- diurnal and seasonal variations in solar zenith angles, which impact both the overall image brightness and topographic shading (see Figure 3.9); and the
- variations in atmospheric conditions, such as water vapour (as cloud, cloud shadow or snow) or aerosols (see Appendix 1).

Figure 3.9 Seasonal solar illumination variations

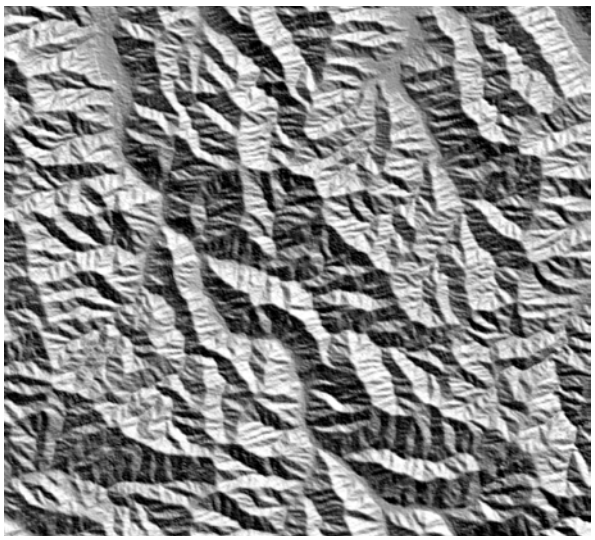
An elevation image was used to simulate the surface shading effects for different seasonal sun positions.

a. Summer (22 December)

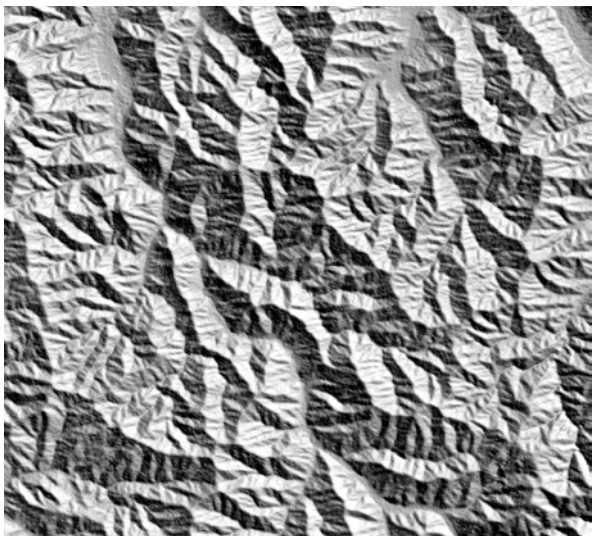
b. Autumn (22 March)



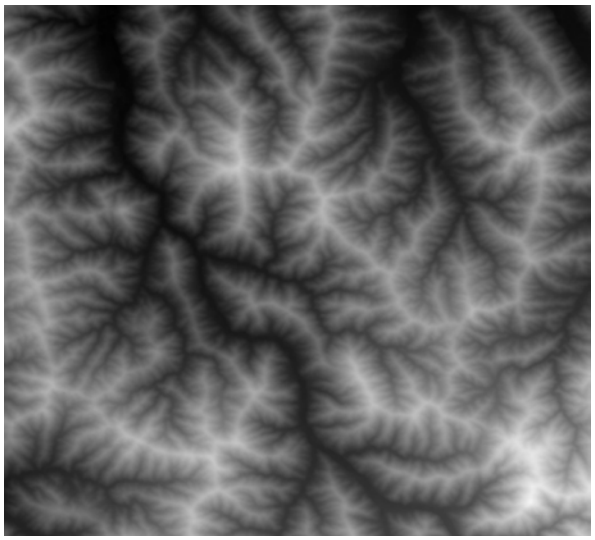
c. Winter (22 June)



d. Spring (22 September)



e. Elevation image displayed with light shades corresponding to high elevation and dark shadows indicating low elevation



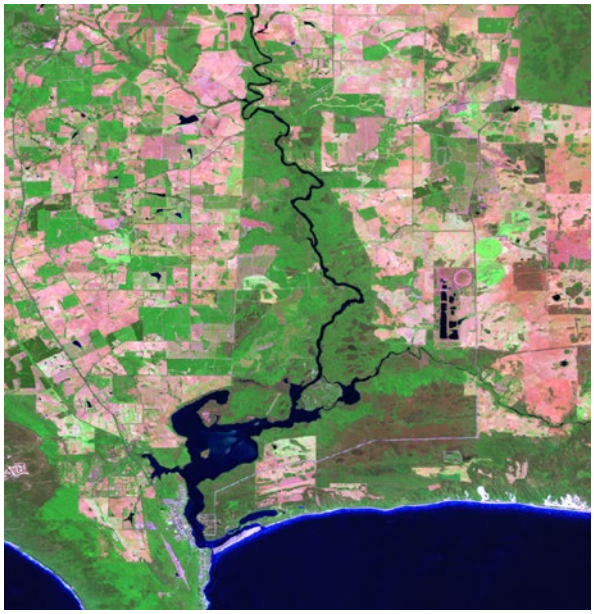
Source: Fuqin Li, Geoscience Australia

Obviously, multi-temporal imagery is likely to show seasonal changes in vegetation cover and inundation, which may or may not be the focus of a particular change detection analysis. Imagery from different sensors is often acquired at different times of the day, which may also reflect diurnal cycles in vegetation, soil moisture and/or water cover. An example of the variations that can occur in imagery acquired in different seasons is shown in Figure 3.10. Specific recommendations for selection of multi-temporal imagery to detect surface cover changes are addressed in Volume 2D.

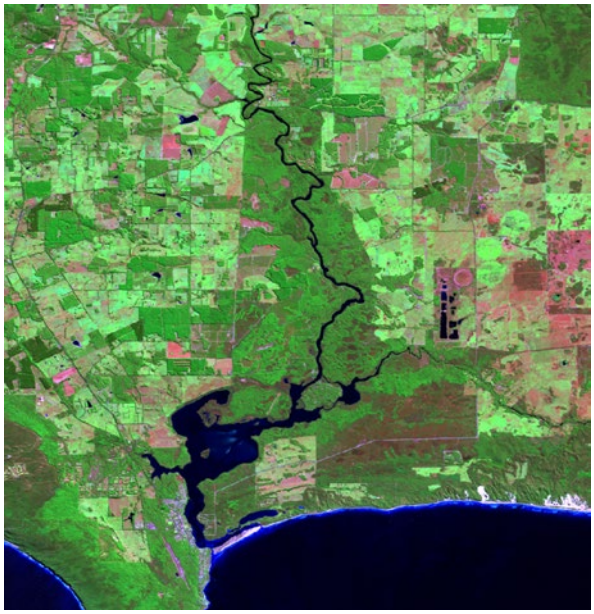
Figure 3.10 Temporal variations in EO imagery

Landsat-8 imagery acquired over Augusta, WA, in four different seasons. Images are displayed as false colour composites, with band 6 as red, band 5 as green, and band 3 as blue.

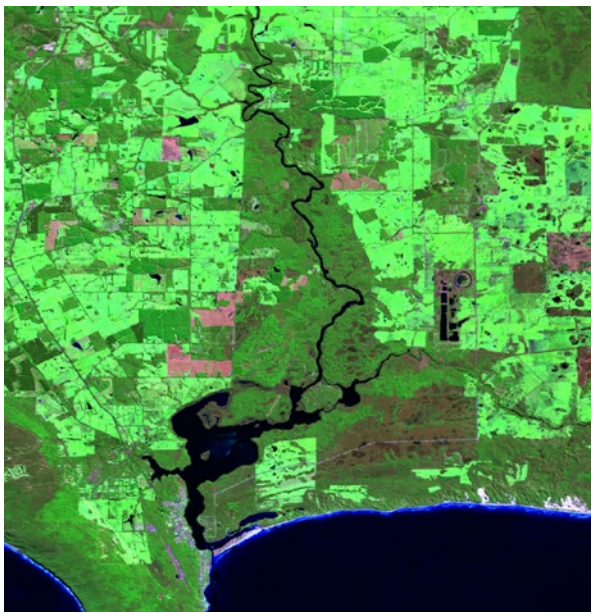
a. Summer—13 January 2015



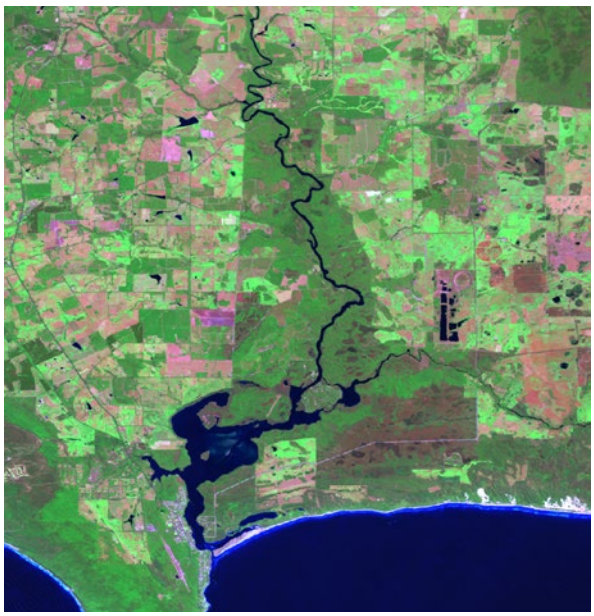
b. Autumn—19 April 2015



c. Winter—25 August 2015



d. Spring—13 November 2015



Source: Norman Mueller, Geoscience Australia

3.3 Optical Image Corrections

Prior to interpretation and analysis, EO imagery is routinely corrected for the various distortions discussed in Section 3.2. Correction methods are grouped into two broad categories:

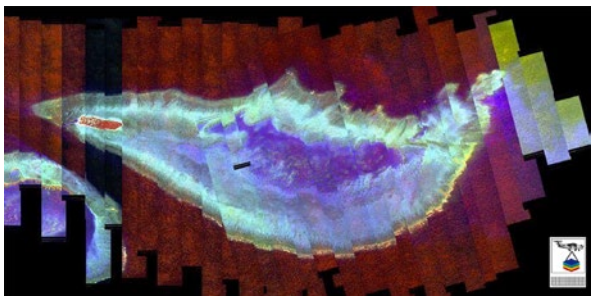
- geometric correction—removing or minimising spatial distortions in the image (see Section 3.3.1); and
- radiometric correction—removing or minimising spectral, radiometric and temporal distortions (see Section 3.3.2).

An example of CASI imagery, acquired from multiple flight lines, that has been corrected for both geometric and radiometric distortions is shown in Figure 3.11. Such corrections are essential in order to digitally compare multi-date images (see Volume 2D).

Figure 3.11 Geometric and radiometric corrections

CASI imagery acquired on 1 June 2002 (western portion) and 3 June 2002 (eastern portion) over Heron Island, Queensland, displayed as natural colour composite (bands 643.7 nm, 522.0 nm, 459.6 nm as RGB).

a. Raw image data—multiple flight lines visible with inconsistent spatial and radiometric features



b. Corrected image—all original images calibrated to surface reflectance , with cross track and deglint corrections applied. (Note: deglinting results in exposed features, such as the island, being obscured.) Image geometry has been rectified to a consistent grid comprising 1 m ground resolution elements.

Source: Karen Joyce, James Cook University

3.3.1 Geometric

Geometric correction accounts for spatial distortions in EO imagery that result from platform and sensor movement and operation, and/or Earth curvature and rotation (see Section 3.2.2 above and Volumes 1A and 2B). Such corrections generally involve three logical processing stages:

- rectification—the process of modelling geometric characteristics of EO imagery so that the image geometry accurately represents the geometric features of the Earth's surface;
- registration—the process of geometrically matching different spatial datasets, such as imagery and/or maps, so that positions in one dataset may be accurately located in others; and
- resampling—modifying the geometry of an image (which may be from either an EO or map data source). This process usually involves rectification and/or registration.

The rectification process involves using well-established models for the sensor and platform, often in conjunction with ground control points (GCP—any features in the study area that can be detected and precisely located on both a map and image), to regularise the image grid. GCP can also be used to obtain external reference information when platform tracking and attitude data are not known precisely.

The process of rectification maps the relationship between locations in the acquired image and a regular, defined (but potentially arbitrary) grid, whereas the process of registration relocates the image data into another defined grid, such as a map projection or another image, possibly with reference to cartographical models. Parameters from both rectification and registration models are used to resample imagery, enabling it to overlay a convenient pre-defined grid.

Image providers generally supply imagery that has been corrected for known geometric distortions and resampled to a defined grid. Such ortho-corrected datasets should state the calibration accuracy and uncertainty, and need to be used with awareness of their pre-processing assumptions and constraints. Multi-temporal studies require all imagery to be resampled to a common grid (see Volume 2D).

An example of the impact that the rectification and resampling processes can have on image geometry is shown in Figure 3.12. The original CASI airborne scanner image over Heron Island Reef displays significant geometric distortion due to platform instability during image acquisition (see Figure

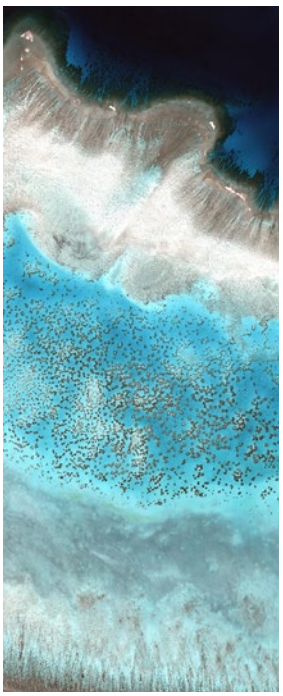
3.12a). Before the image can be usefully compared with other spatial data, these geometric distortions need to be corrected (see Figure 3.12b). Processing methods involved in image rectification, registration and resampling are introduced in Section 7 below and detailed in Volume 2B.

Figure 3.12 Geometric correction of CASI imagery

Imagery acquired by the airborne scanners can exhibit significant geometric distortion due to aircraft movement during image acquisition (see Figure 3.11 for image details). In this case the image also needed to be flipped top to bottom (see Section 7.2.2).

a. Original image

b. Corrected image



Source: Karen Joyce, James Cook University

3.3.2 Radiometric

As detailed in Section 3.2above and Volume 1, the radiance measured by EO optical sensors includes the target reflectance plus a range of effects related to image acquisition, including atmospheric scattering and absorption, surface BRDF, terrain illumination and scanner operation. These acquisition-related effects may be corrected either by simple transformations to reduce their impact or by applying physics-based models with appropriate data (see Section 8.3). For example, the impact of correcting Landsat TM/ETM+ imagery for atmospheric and BRDF effects is illustrated in Figure 3.13.

The logical components of radiometric correction include:

- atmospheric correction—uses vicarious calibration data (see Section 3.1) to correct for atmospheric effects, either by:
  - 'top-down' modelling (atmospheric correction models convert 'top-of-atmosphere' sensor data to reflectance from the ground); or
  - 'bottom up' modelling (radiative transfer models and atmospheric properties convert target reflectances measured on the ground to top-of-atmosphere radiance).
- surface BRDF minimisation—reduces differences due to changes in viewing and illumination positions within and between images, using BRDF parameters or co-incident image acquisitions (Collings *et al.*, 2010; Wu *et al.*, 2010; Collings *et al.*, 2011; Flood *et al.*, 2013). Such methods are detailed in Appendices 1 and 2.
- terrain illumination correction—modifies terrain shading effects using geometrical parameters derived from accurately registered Digital Surface Models (DSM) (as in Li *et al.*, 2012). Topographic correction procedures are further discussed in Appendix 1.
- instrument calibration—uses on-board and/or vicarious reference data to correct for detector differences. Correction of instrument errors may involve destriping algorithms to remove systematic patterns (such as histogram equalisation—see Volume 2C), or specialised algorithms for random distortions (see Section 8.3).

Absolute radiometric correction methods convert pixel digital values to radiance at the sensor, before using atmospheric/BRDF models to estimate surface reflectance(s) (Paolini *et al.*, 2006). An example of multi-scene cross-calibration of 140 ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) images is shown in Figure 3.14. This processing was undertaken as part of a project developing a suite of mineral maps for Australia based on hyperspectral and multispectral EO imagery (Cudahy *et al.*, 2008; see also Excursus 8.3). The seamless mosaic was generated using:

- a physical model that normalises (most of) the intrinsic EO image complications, such as topographic shading, surface directional scattering, solar illumination and residual atmospheric and instrument effects;
- transformation of the ASTER mosaic to (apparent) surface reflectance using least-squares regression between the ASTER data and atmospherically-corrected HyMap airborne imagery; and
- validation of the accuracy of the ASTER cross-calibration based on 60 coincident regions of interest.

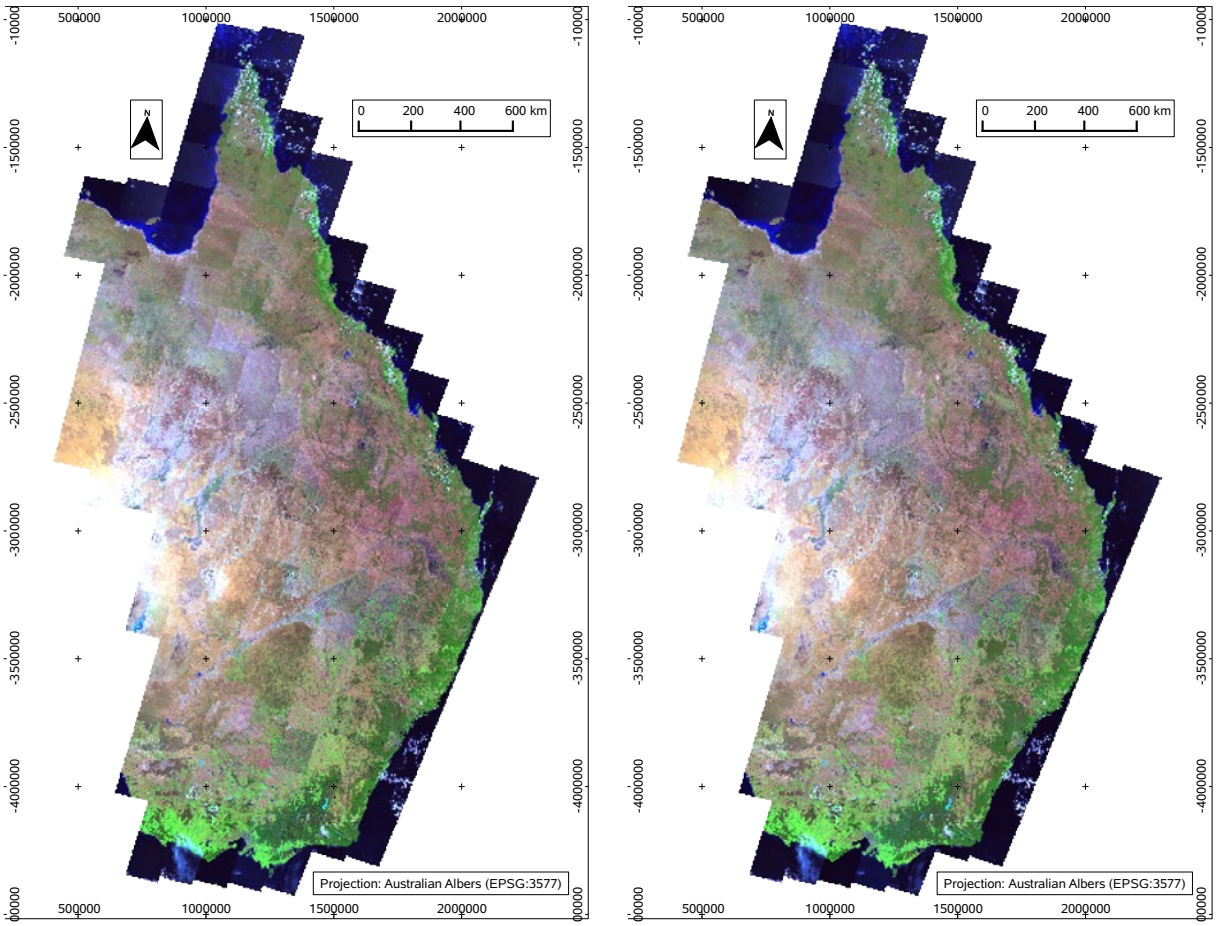
Figure 3.13 Surface reflectance correction

This example of surface reflectance correction applied to Landsat TM imagery for eastern Australia uses a method described by Flood *et al.* (2013). This method uses the 6S Radiative Transfer Model to correct for atmospheric effects and bi-directional reflectance modelling to remove the effects of topography and angular variation in reflectance. Note that some atmospheric effects remain in the corrected imagery due to limitations in input parameters. For example, aerosol modelling assumes a fixed Aerosol Optical Depth (AOD) of 0.05 which would not be valid in smoky or hazy conditions.

This Landsat TM mosaic of eastern Australia was compiled from imagery acquired between August and October 2006. These months experienced prolonged dry weather so vegetation changes were minimal.

a. Standard top-of-atmosphere reflectance

b. Corrected surface reflectance image



Source: Flood *et al.* (2013) using imagery from USGS

Integrated modelling approaches are now being used to generate a range of standard, calibrated image products, such as the ARG25 version of Landsat TM imagery supplied by GA, which is corrected for multiple radiometric distortions (see Section 3.4). Analysis of statistics related to radiometric distortions are further discussed in Section 8.3.

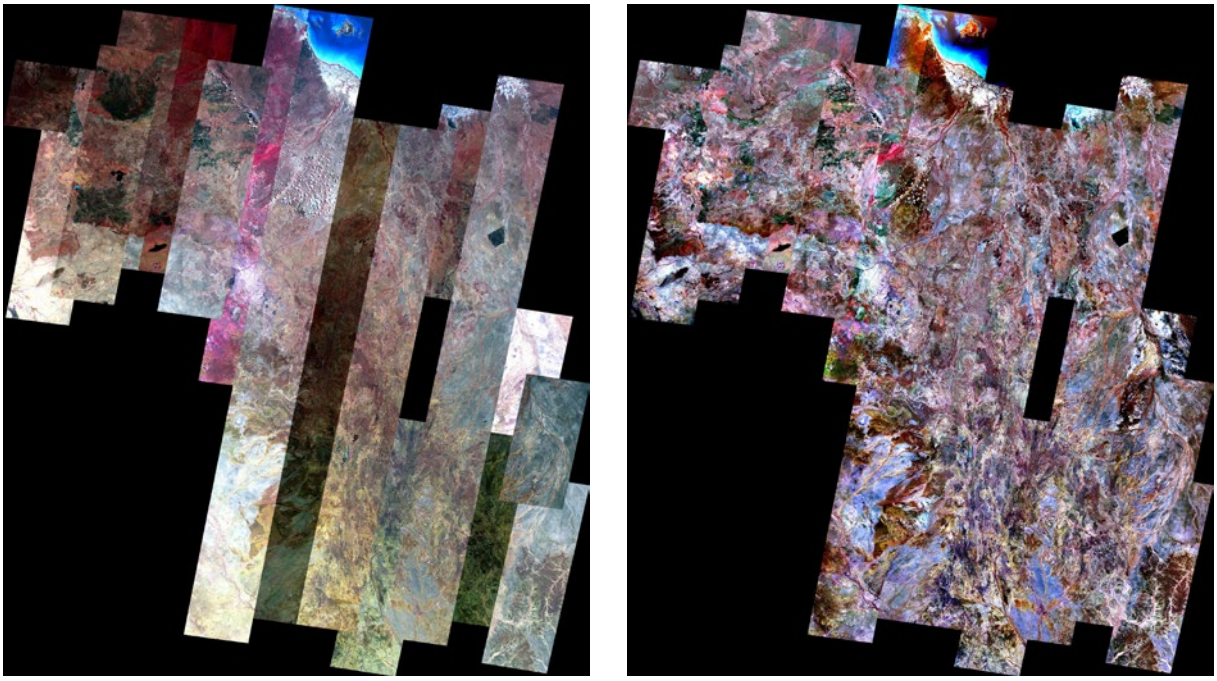
One network of corner reflectors that has been established in Australia to calibrate SAR instruments is described in Excursus 3.1.

Figure 3.14 ASTER image mosaic

This mosaic over the Mount Isa region, Queensland comprises 140 ASTER image scenes acquired over a five year period.

a. Level 1B (radiance-at-sensor) mosaic with SWIR cross-talk correction

b. CSIRO cross-calibrated mosaic after reduction to (apparent) surface reflectance.



Source: Thomas Cudahy, C3DMM

Excursus 3.1—Australian Geophysical Observing Network

Source: Medhavy Thankappan, Geoscience Australia

Calibration of Synthetic Aperture Radar instruments using corner reflectors

The performance of Synthetic Aperture Radar (SAR) instruments is verified by internal and/or external calibration. In this context, calibration is the process of characterising the end-to-end performance of the SAR system to measure the backscattered signal. Internal calibration involves characterisation of the radar system performance using signals from devices built into the sensor system, while external calibration involves the use of ground-based points or distributed targets.

In external calibration, the performance of the SAR instrument is related to a known measurement standard. Point targets, such as radar reflectors or active transponders, can be used for this purpose. A radar reflector is a passive device that reflects incoming radar energy back to the source and can take different forms. The trihedral radar reflector, commonly referred to as a 'corner reflector', which strongly reflects the incident radar energy by virtue of its shape, is structurally robust, more tolerant of alignment errors during deployment compared to other forms, and a popular choice for ground-based calibration of SAR satellites.

The Radar Cross Section (RCS) of an imaged target is a measure of the size of that target as seen by the imaging radar. Distributed targets of known RCS such as agricultural fields, tropical rain forests, or boreal forests, could also be used for calibration. Corner reflectors are routinely used, as they have low maintenance and are available at low cost compared to active devices such as transponders, which also need power for operation. Corner reflectors also exhibit a large RCS relative to their small size, which is maintained over a wide range of incidence angles, ensuring their proper identification in the SAR image.

The Australian Government invested AUD$23 million in 2010, towards building the Australian Geophysical Observing System (AGOS), to enable highly accurate spatial and temporal estimation of large-scale

ground deformation. GA implemented the geospatial component of AGOS, which includes Global Navigation Satellite System (GNSS) instrumentation, a network of high precision GNSS monuments, a robotic GNSS antenna calibration facility, a Synthetic Aperture Radar (SAR) data repository and a network of passive radar corner reflectors in areas specific to AGOS research interests, and designed such that they can be used to monitor crustal deformation and perform routine calibration of SAR instruments on spaceborne platforms.

In order to identify the optimal design of corner reflectors to be deployed as part of the AGOS infrastructure and ensure suitability for calibration applications, GA manufactured 18 corner reflector prototypes with different sizes and material finishes for evaluation. A triangular trihedral design was chosen for the corner reflectors because of the simplicity of manufacture, long term structural rigidity and relative stability for large RCS. Garthwaite *et al.* (2015a) provide a comprehensive report on the design and size choices for the corner reflectors.

The corner reflector prototypes were characterised at the Defence Science and Technology Organisation (DSTO) ground reflection range in Adelaide, by comparing actual RCS measurements with the expected theoretical values and quantifying the change in RCS at different azimuth and elevation angles (Drake and Hatty, 2013). The setup of the corner reflectors for characterisation at the ground reflection range is illustrated in Figure 3.15 and Figure 3.16. Results from the characterisation of the corner reflectors showed that the RCS performance of the prototypes was comparable to theoretical values (Thankappan *et al.*, 2013).

In order to evaluate actual field performance, the corner reflector prototypes were then temporarily deployed at a test site north of Canberra for a five month period. Performance testing involved data acquisitions using SAR satellites at X- and C-band to verify that the observed RCS of the corner reflectors are comparable to theoretical values for calibration of SAR instruments. SAR image acquisitions over the test site showing the response from the corner reflectors can be seen in Figure 3.17.

Figure 3.15 Configuration of trial corner reflectors

Schematic of the experimental setup for corner reflector RCS measurements at the DSTO ground reflection range. The vertical axis that bisects the reflector corner apex also bisects the centre of the turntable used to rotate the reflector in azimuth.

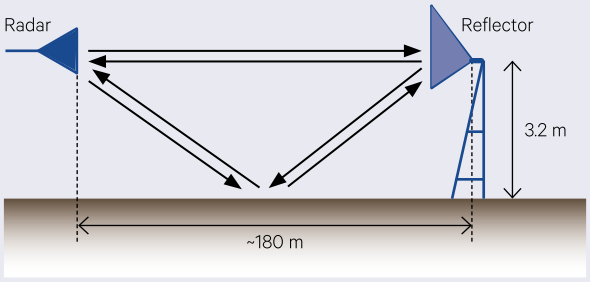


Figure 3.16 Corner reflectors

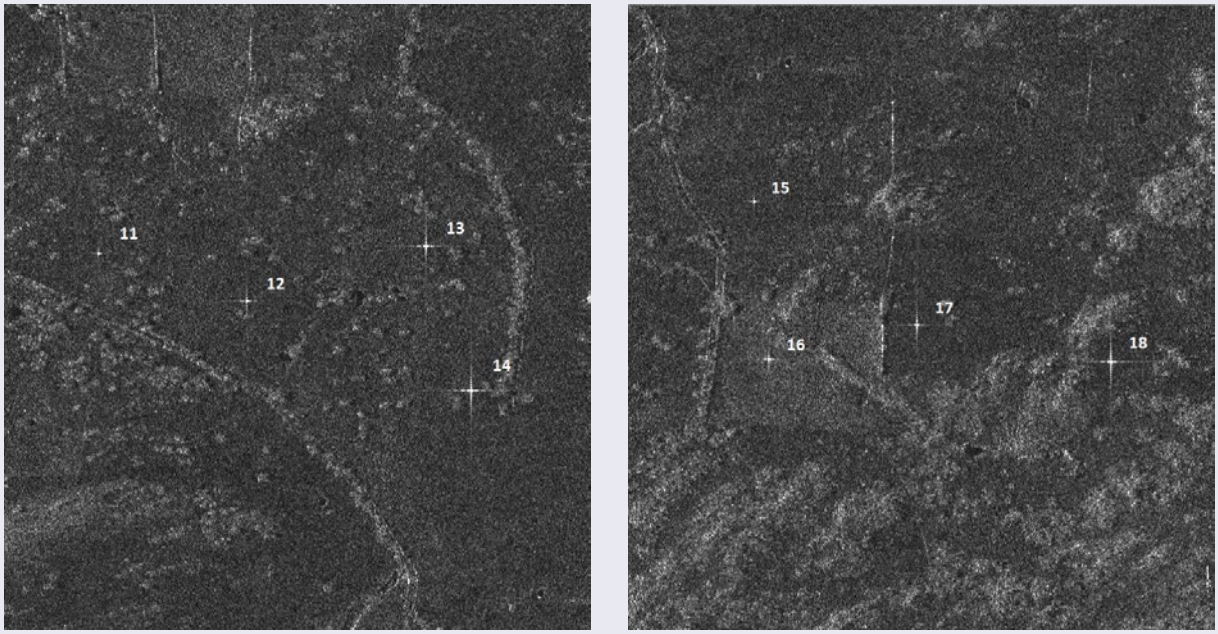
Aperture views of a 1.5 m corner reflector at the DSTO ground reflection range with different alignments:

a. Azimuth cut measurement

b. Elevation cut measurement

Figure 3.17 CR response in SAR image

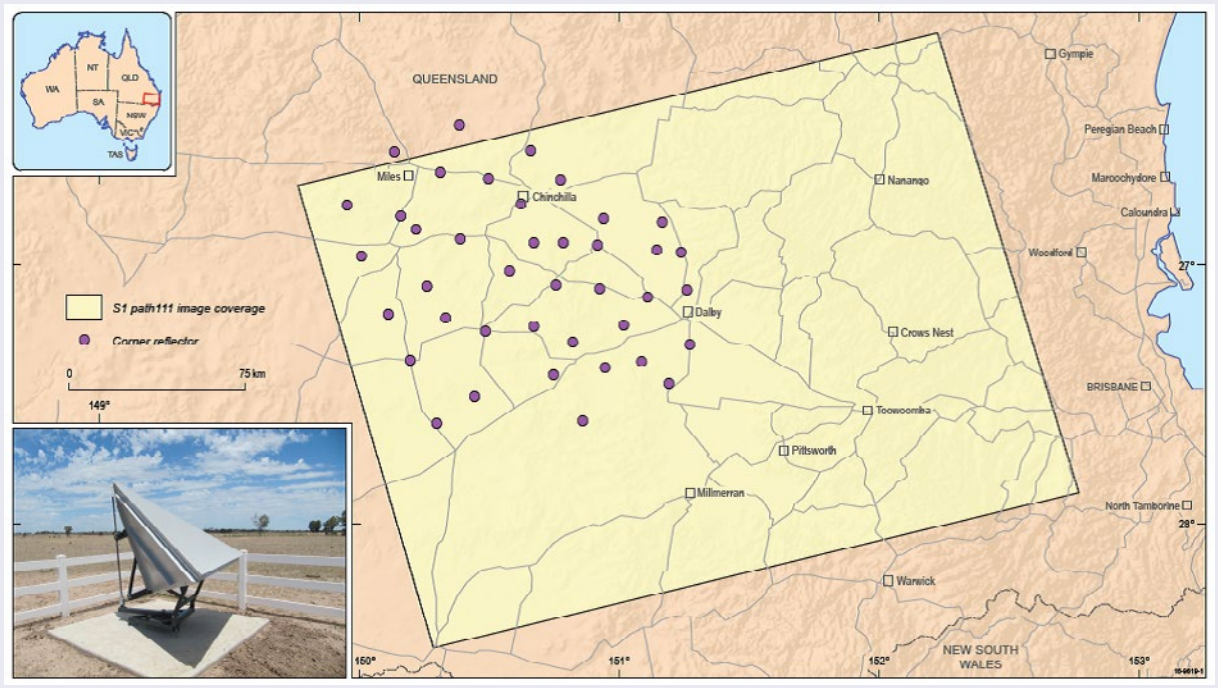
Impulse response of some of the corner reflector prototypes in TerraSAR-X images over Gunning test site on 29 December 2013, labelled by corner reflector number.



Following the performance evaluation, a total of 40 corner reflectors (CR) were deployed permanently in Queensland, Australia, as shown in Figure 3.18. The array includes 34 CR of 1.5 m, 3 CR of 2.0 m and 3 CR of 2.5 m dimension. Through the design process and the precision manufacturing techniques employed, the CR of different sizes are highly suitable for calibration and validation of spaceborne SAR instruments operating in X-, C- and L-band frequencies.

Figure 3.18 CR network

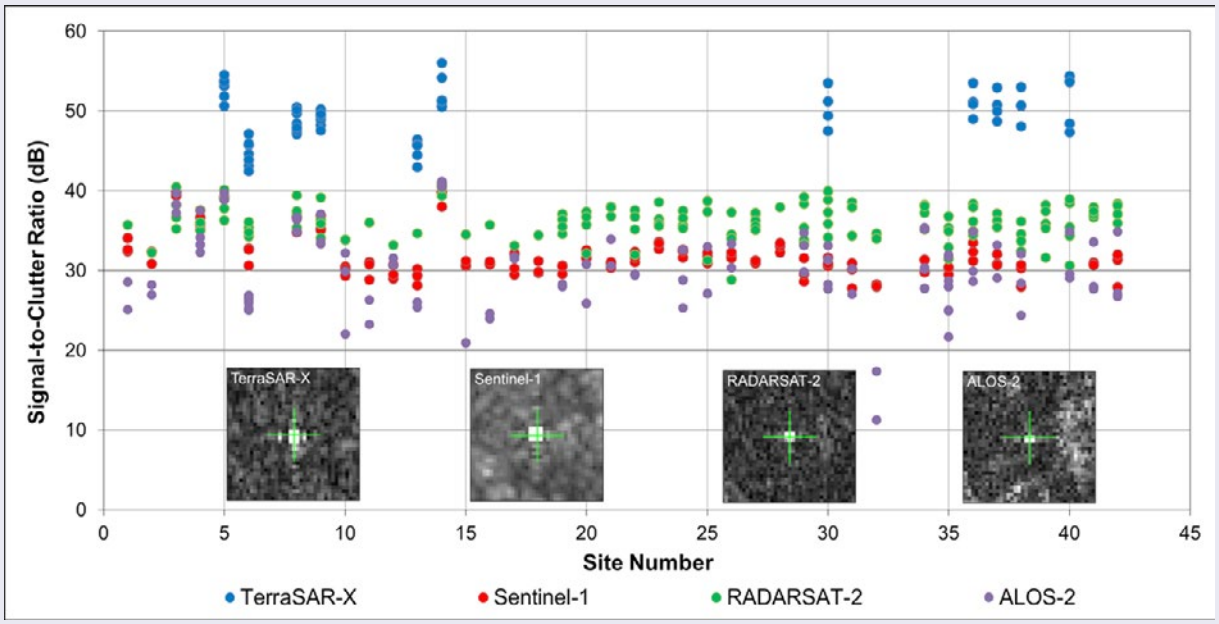
Map of the Surat Basin in Queensland, Australia showing the distribution of 40 corner reflector sites. Inset photo shows a typical 1.5 m corner reflector deployed in the network, which is designed for adjustment in azimuth and elevation.



A commonly used measure of target visibility in a SAR image is the Signal-to-Clutter Ratio (SCR). The point target response from the Queensland corner reflector array has been assessed using SCR for SAR datasets from TerraSAR-X, RADARSAT-2, Sentinel-1A and ALOS-2 satellites (see Figure 3.19). Generally, the SCR for calibration targets should be at least 20 dB (preferably ~30 dB) in order to minimise errors in computation of the calibration factor but also the signal should not be saturated. SCR for each corner reflector was calculated using the oversampled impulse response image and was found to exceed 20 dB irrespective of the frequency bands studied.

Figure 3.19 Target visibility measure

Signal-to-Clutter Ratio (SCR) for SAR datasets from TerraSAR-X, Sentinel-1A, RADARSAT-2 and ALOS-2 satellites.



The AGOS corner reflector array is the largest array of corner reflectors (40) spread over a 100 km by 100 km area that is capable of supporting calibration of SAR satellites operating in multiple frequencies including X-, C- and L-bands. It is currently supporting the European Space Agency's Fiducial Reference Measurement for SAR (FRM4SAR) initiative and being used routinely for calibration of the Sentinel-1A and Sentinel-1B SAR satellite constellation. Other international satellite operators have expressed interest in using this infrastructure for ongoing calibration of their SAR missions.

The permanently deployed corner reflector infrastructure presents an opportunity for independent calibration and comparison of SAR instruments on current and future satellite missions, and is considered an important Australian contribution to global satellite calibration/validation efforts (Garthwaite *et al.*, 2015b).

3.4 Analysis Ready Data

Traditionally, geometric and atmospheric corrections of EO imagery have been variously shared by data suppliers and users. Images were corrected individually using the best available methods, with the same image being corrected multiple times by multiple users. While this approach has enabled decades of advances in EO applications, it is inadequate for ensuring access to the expanding archive of EO data.

The volume of global EO data is clearly growing rapidly in terms of data sources, frequency and archives. Increased awareness of this trend has necessitated significant changes in the established data supply chains. Several suppliers have embraced the concept of 'Analysis Ready Data', which are calibrated and organised in a readily accessible format for immediate analysis. Such data are intended to be:

- geo-located—all pixels aligned geometrically;
- cross-calibrated—all pixels comparable through time and across sensors; and
- ordered—all pixels sequenced in time for easy retrieval.

The availability of 'Analysis Ready Data' reduces inefficiencies in data handling, calibration and processing and ensures a consistent base between users. It is anticipated that integrated data archives will extend analysis and integration opportunities for EO imagery and allow its full value to be realised. For example, CEOS Analysis Ready Data for Land (CARD4L) aims to deliver 'satellite data that have been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets' (CEOS, 2017).

One implementation of analysis ready data is the Digital Earth Australia (DEA; formerly Australian Geoscience Data Cube, AGDC). This initiative organises data into stacks of consistent, time-stamped geographic 'tiles', which can be rapidly manipulated in a high performance computing environment. Pixel flags record data quality decisions and a relational database tracks all tiles, pixels and metadata for the data cube. Data cube technology offers a streamlined approach to the varied and voluminous flow of data from the global EO satellite fleet and is expected to become the standard delivery vehicle for EO data and information products. Further information on the DEA is provided in Section 7.4.2.

3.5 Further Information

Digital Earth Australia (DEA; formerly Australian Geoscience Data Cube, AGDC): *http://www.ga.gov.au/about/projects/geographic/digital-earth-australia*

TERN AusCover Good Practice Guidelines: Held *et al.* (2015)

Jensen (2016) Chapter 6

USGS Landsat Missions: *https://landsat.usgs.gov*

Landsat-8 Calibration: *https://landsat.usgs.gov/landsat-8-l8-data-users-handbook-section-3*

CEOS Analysis Ready Data for Land (CARD4L): (CEOS, 2017)

A number of operational, commercial and research systems are already being used extensively, such as the Google Earth Engine (Gorelick *et al.*, 2017): *https://earthengine.google.com/*

3.6 References

Barsi, J. A., Markham, B. L., Helder, D. L., and Chander, G. (2007). *Radiometric calibration status of Landsat-7 and Landsat-5*. Paper presented at the SPIE 6744, Sensors, Systems, and Next-Generation Satellites XI, Florence, Italy. *http://dx.doi.org/10.1117/12.738221*.

Breaker, L. C. (1990). Estimating and removing sensor-induced correlation from advanced very high resolution radiometer satellite data. *Journal of Geophysical Research: Oceans*, 95(C6), pp. 9701-9711. doi:*http://dx.doi.org/10.1029/JC095iC06p09701*.

CEOS (2017). *CEOS Analysis Ready Data for Land (CARD4L) Description Document*. Retrieved from: *https://www.google.com.au/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiR-_OCosLXAhWLnJQKHRpUCOsQFggoMAA&url=http%3A%2F%2Fceos.org%2Fdocument_management%2FMeetings%2FPlenary%2F30%2FDocuments%2F5.5_CEOS-CARD4L-Description_v.22.docx&usg=AOvVaw2r8F89fp--1nxlALHTQsAi*.

Collings, S., Caccetta, P., Campbell, N., and Wu, X. (2010). Techniques for BRDF Correction of Hyperspectral Mosaics. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10), pp. 3733-3746. doi:*http://dx.doi.org/10.1109/TGRS.2010.2048574*.

Collings, S., Caccetta, P., Campbell, N., and Wu, X. (2011). Empirical Models for Radiometric Calibration of Digital Aerial Frame Mosaics. *IEEE Transactions on Geoscience and Remote Sensing*, 49(7), pp. 2573-2588. doi:*http://dx.doi.org/10.1109/TGRS.2011.2108301*.

Cracknell, A. P. (1998). Synergy in remote sensing - what's in a pixel? *International Journal of Remote Sensing*, 19(11), pp. 2025-2047. doi:*http://dx.doi.org/10.1080/014311698214848*.

Cudahy, T., Jones, M., Thomas, M., Laukamp, C., Caccetta, M., Hewson, R., Rodger, A., and Verrali, M. (2008). *Next Generation Mineral Mapping: Queensland Airborne HyMap and Satellite ASTER Surveys 2006–2008*. CSIRO Exploration and Mining Report P2007/34.

Drake, P., and Hatty, S. (2013). *Radar Signature Characterisation Measurements of Triangular Trihedral Calibrators Rep*. DSTO-CR-2013-0212. Defence Science and Technology Organisation, Salisbury, Australia. 132 pp.

Flood, N., Danaher, T., Gill, T., and Gillingham, S. (2013). An Operational Scheme for Deriving Standardised Surface Reflectance from Landsat TM/ETM+ and SPOT HRG Imagery for Eastern Australia. *Remote Sensing*, 5(1), pp. 83. doi:*http://dx.doi.org/10.3390/rs5010083*.

Garthwaite, M. C., Nancarrow, S., Hislop, A., Thankappan, M., Dawson, J. H., and Lawrie, S. (2015a). *Design of radar corner reflectors for the Australian geophysical observing system: a single design suitable for InSAR deformation monitoring and SAR calibration at multiple microwave frequency bands. Record 2015/03.* Geoscience Australia, Canberra. *http://dx.doi.org/10.11636/Record.2015.003*

Garthwaite, M. C., Hazelwood, M., Nancarrow, S., Hislop, A., and Dawson, J. H. (2015b) A regional geodetic network to monitor ground surface response to resource extraction in the northern Surat Basin, Queensland. *Aust. J. Earth Sci*. 62, 4.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. doi:*https://doi.org/10.1016/j.rse.2017.06.031*.

Held, A., Phinn, S. R., Soto-Berelov, M., and Jones, S. (2015). *AusCover Good Practice Guidelines: A technical handbook supporting calibration and validation activities of remotely sensed data products*. TERN AusCover. Retrieved from *http://data.auscover.org.au/xwiki/bin/view/Good+Practice+Handbook/WebHome*.

Jensen, J.R. (2016) *Introductory Digital Image Processing. A Remote Sensing Perspective*. 4th edn. Pearson Education, Inc.

Kieffer, H. H., Stone, T. C., Barnes, R. A., Bender, S., Eplee, R. E. J., Mendenhall, J., and Ong, L. (2003). *On-orbit radiometric calibration over time and between spacecraft using the Moon*. Proceedings of: Sensors, Systems and Next-Generation Satellites VI. SPIE 4.

Kieffer, H. H., and Wildey, R. L. (1985). Absolution calibration of Landsat instruments using the moon. *Photogrammetric Engineering and Remote Sensing*, 51(9), pp. 1391–1393.

Li, F., Jupp, D. L. B., Thankappan, M., Lymburner, L., Mueller, N., Lewis, A., and Held, A. (2012). A physics-based atmospheric and BRDF correction for Landsat data over mountainous terrain. *Remote Sensing of Environment*, 124, pp. 756-770. doi:*http://dx.doi.org/10.1016/j.rse.2012.06.018*.

Malthus, T. J., and Li, F. (2015). Calibration of Optical Satellite and Airborne Sensors. Chapter 4 in *'AusCover Good Practice Guidelines (A technical handbook supporting calibration and validation activities of remotely sensed data products)'* (Eds: A. Held, S. R. Phinn, M. Soto-Berelov, and S. Jones). TERN AusCover, St Lucia, Australia.

Mannstein, H., and Gesell, G. (1991). *Deconvolution of AVHRR data*. Paper presented at the 5th AVHRR Data Users' Meeting, Tromsø, Norway.

Markham, B. L., Thome, K. J., Barsi, J. A., Kaita, E., Helder, D. L., Barker, J. L., and Scaramuzza, P. L. (2004). Landsat-7 ETM+ on-orbit reflective-band radiometric stability and absolute calibration. *IEEE Transactions on Geoscience and Remote Sensing*, 42(12), pp. 2810-2820. doi:*http://dx.doi.org/10.1109/TGRS.2004.836389*.

Paolini, L., Grings, F., Sobrino, J. A., Jiménez Muñoz, J. C., and Karszenbaum, H. (2006). Radiometric correction effects in Landsat multi-date/multi-sensor change detection studies. *International Journal of Remote Sensing*, 27(4), pp. 685-704. doi:*http://dx.doi.org/10.1080/01431160500183057*.

Richter, R., Kellenberger, T., and Kaufmann, H. (2009). Comparison of Topographic Correction Methods. *Remote Sensing*, 1(3), pp. 184. doi:*http://dx.doi.org/10.3390/rs1030184*.

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N. C., Zhang, X., Jin, Y., Muller, J.-P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entremont, R. P., Hu, B., Liang, S., Privette, J. L., and Roy, D. (2002). First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment*, 83(1), pp. 135-148. doi:*http://dx.doi.org/10.1016/S0034-4257(02)00091-3*.

Slater, P. N., Biggar, S. F., Holm, R. G., Jackson, R. D., Mao, Y., Moran, M. S., Palmer, J. M., and Yuan, B. (1987). Reflectance- and radiance-based methods for the in-flight absolute calibration of multispectral sensors. *Remote Sensing of Environment*, 22(1), pp. 11-37. doi:*http://dx.doi.org/10.1016/0034-4257(87)90026-5*.

Slater, P. N., Biggar, S. F., Thome, K. J., Gellman, D. I., and Spyak, P. R. (1996). Vicarious Radiometric Calibrations of EOS Sensors. *Journal of Atmospheric and Oceanic Technology*, 13(2), pp. 349-359. doi:*http://dx.doi.org/10.1175/1520-0426(1996)013%3C0349:vrcoes%3E2.0.co;2*.

Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., and Macomber, S. A. (2001). Classification and Change Detection Using Landsat TM Data. Remote Sensing of Environment, 75(2), pp. 230-244. doi:*http://dx.doi.org/10.1016/S0034-4257(00)00169-3*.

Teillet, P. M., Barker, J. L., Markham, B. L., Irish, R. R., Fedosejevs, G., and Storey, J. C. (2001). Radiometric cross-calibration of the Landsat-7 ETM+ and Landsat-5 TM sensors based on tandem data sets. *Remote Sensing of Environment*, 78(1), pp. 39-54. doi:*http://dx.doi.org/10.1016/S0034-4257(01)00248-6*.

Teillet, P. M., Barsi, J. A., Chander, G., and Thome, K.J. (2007). *Prime candidate Earth targets for the post-launch radiometric calibration of space-based optical imaging instruments*. Paper presented at the SPIE 6677, Earth Observing Systems XII, San Diego, United States. *http://dx.doi.org/10.1117/12.733156*.

Thankappan, M., Garthwaite, M. C., and Williams, M. L., Hislop, A., and Dawson, J.H. (2013). *Characterisation of Corner Reflectors for the Australian Geophysical Observing System to Support SAR Calibration*. Proc. ESA: Living Planet Symposium, Edinburgh, 9–13 September 2013.

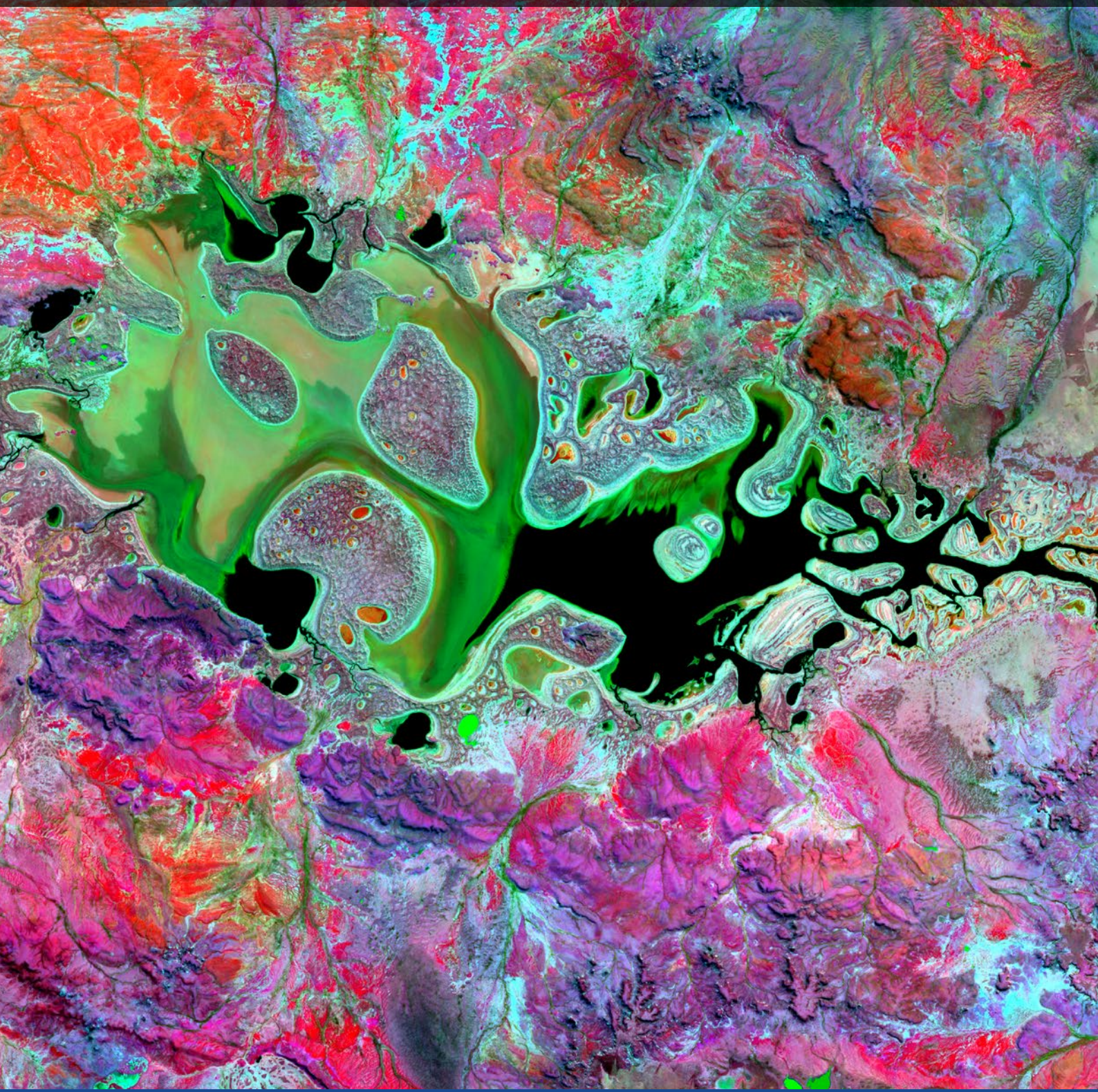
Thome, K., Markham, B., Barker, J., Slater, P. N., and Biggar, S. (1997). Radiometric calibration of Landsat. *Photogrammetric Engineering & Remote Sensing*, 63(7), pp. 853-858.

Thome, K. J., Helder, D. L., Aaron, D., and Dewald, J. D. (2004). Landsat-5 TM and Landsat-7 ETM+ absolute radiometric calibration using the reflectance-based method. *IEEE Transactions on Geoscience and Remote Sensing*, 42(12), pp. 2777-2785. doi:*http://dx.doi.org/10.1109/TGRS.2004.839085*.

USGS (2017). *SLC-off Products: Background*. Retrieved from *https://landsat.usgs.gov/slc-products-background*.

Wu, X., Collings, S., and Caccetta, P. (2010). *BRDF and illumination calibration for very high resolution imaging sensors*. Paper presented at the IEEE International Geoscience and Remote Sensing Symposium (IGARSS).

Interpretation



While image processing is largely concerned with computer manipulation of data values, a fundamental part of analysing imagery involves visual interpretation. The human eye can only differentiate a limited number of discrete colours and an even more restricted set of grey shades. Display representation of the image data therefore needs to maximise the contrast between available display colours to provide optimal differentiation of image features. Various techniques for image enhancement have been developed to highlight features of interest in displayed data.

Interpretation is generally the initial process of visually delineating and identifying features or patterns in an image. The following four sections cover the basic techniques that are relevant to visual image interpretation, namely:

- contrast—enhance contrast between image features and background (see Section 4);
- colour—using colour to highlight specific features or compare multiple channels (see Section 5);
- presentation—methods for displaying image data on monitor devices and producing hardcopy imagery, such as inkjet or laser plots or photographic products (see Section 6).
- geometry—changing image shape, scale, and orientation (see Section 7); and

Contents

Background image on previous page: Infrared colour composite over Lake Carnegie, a large ephermal lake in central WA, acquired by Landsat-8 on 8 July 2017. The vibrant colours of this non-standard colour composite derive from bands 10 (TIR 1), 5 (NIR) and 6 (SWIR 1) as RGB respectively (see also Figures 5.16 to 5.19). **Source:** Norman Mueller, Geoscience Australia.

4 Contrast

Contrast relates to the ability to differentiate a feature from its background. In image processing, contrast enhancement can be conducted in a number of ways depending on the type of features being identified in an image. These techniques can improve either the specific contrast of a selected feature or the overall image contrast.

Numeric values recorded in image or raster format may be represented as a picture by associating each image value with an intensity and/or colour on a display or hardcopy device. A simple grey-scale representation of an image channel is illustrated in Figure 4.1.

Figure 4.1 Grey-scale image display

Any spatial data can be represented pictorially by associating each data value with a grey shade or colour.

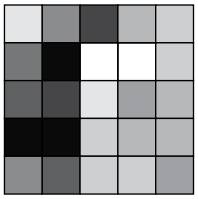
a. Image pixel values

1	5	8	3	2
6	9	0	0	2
7	8	1	4	3
9	9	2	3	3
5	7	2	2	4

b. Grey-scale lookup table

0	
1	
2	
3	
4	
5	
6	
7	
8	
9	

c. Grey-scale image



Source: Harrison and Jupp (1990) Figure 3

A finite number of grey shades or colours can be represented using display screens or hardcopy devices. Remote sensors are designed to detect reflectance/emission levels over a wider range than would occur within a single image scene, so the data range for a single remotely sensed image rarely covers the full potential range of image values (such as 0–255 for 8-bit (or 'byte') data sources—see Section 1). To maximise the contrast in a pictorial representation of image data, we need to match the range and distribution of values in the image (or selected features of interest within the image) to the available range of grey shades or colours in the display device.

4.1 Channel statistics

The range of data values in an image channel is usefully summarised by its frequency histogram (see Section 4.1.1) while the distribution of values within the data range is indicated by the variance statistic (see Section 4.1.2). These simple statistics are commonly used by image processing systems to improve image contrast.

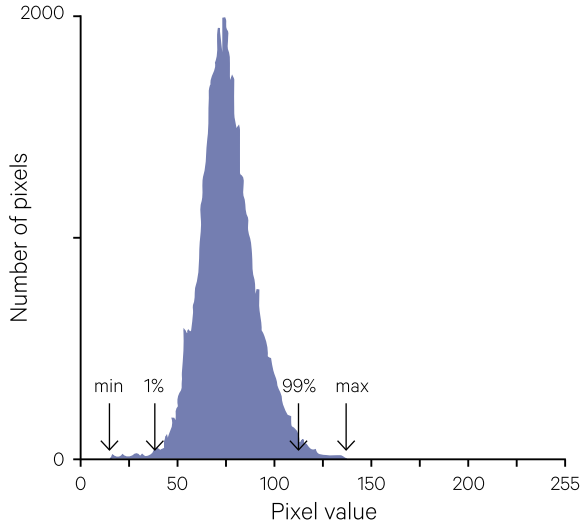
4.1.1 Histograms

The range and distribution of data values in any image channel may be represented as a histogram. The usual form for histograms in image processing is the 'frequency histogram' shown in Figure 4.2, with the horizontal axis representing the image data range and the vertical axis indicating the number of pixels. A frequency histogram is formed by plotting the number of pixels that contain each digital value in a channel against the full range of digital values. In the example in Figure 4.2, the most common image channel value (the 'mode'—maximum count) is 74 for 2,131 pixels. Ancillary statistics such as the absolute minimum, maximum and mode values in each channel are usually reported separately. The total number of pixels used to generate the histogram is also usually reported (61,440 in this example). The frequency histogram and its related statistics are used extensively in image processing to define the effective data range for image display and to transform or identify particular features.

Background image: Panchromatic image of Canberra, ACT, acquired by Landsat-8 on 10 August, 2017. **Source:** Norman Mueller, Geoscience Australia

Figure 4.2 Frequency histogram

A frequency histogram is formed by plotting the number of pixels at each image value in a channel. In this example, for a total of 61,440 pixels, the minimum channel value is 15 and the maximum value is 137. The largest number of pixels, or mode, occurs at image value 74 (2,131 pixels).

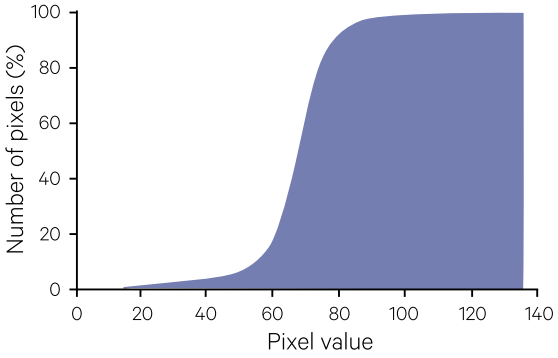


Source: Harrison and Jupp (1990) Figure 4

Percentage points (or percentiles) for the histogram summarise the histogram distribution in terms of cumulative statistics. A cumulative histogram, as illustrated in Figure 4.3, could be derived manually from the full range of percentiles if required.

Figure 4.3 Cumulative histogram

This demonstrates a (stylised) cumulative histogram for the image data used in Figure 4.2. 1% of pixels have values below or equal to 44, 20% have values below or equal to 66, while 95% of pixels have values below or equal to 89.



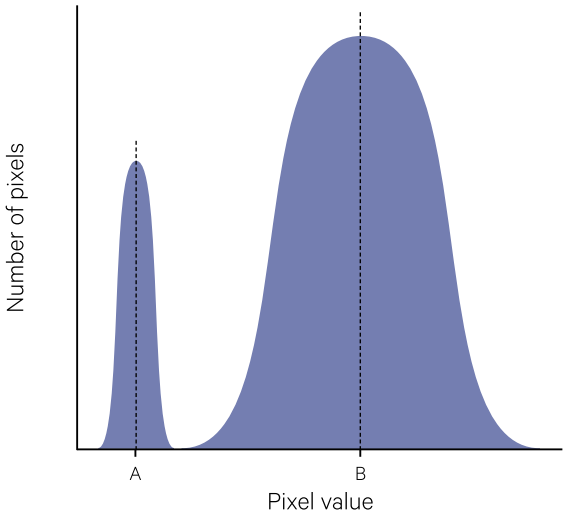
Source: Harrison and Jupp (1990) Figure 5

The shape of the frequency histogram is also significant in interpreting image data. The example histogram (see Figure 4.2) has the mode closer to the minimum than to the maximum value producing a skewed distribution. This is caused by a larger spread of values close to the maximum value, which have relatively small numbers of pixels. This feature is also reflected in the relatively large difference between the values of the 95% and 99% points (89 and 113 respectively—see Figure 4.3). The degree of 'skewness' can be readily assessed in most cases by comparing the median or modal pixel values for a channel with its mean value. Skewed data distributions commonly occur in EO imagery with the mode usually being closer to the minimum value.

Channel histograms can also be discontinuous or multi-modal in EO imagery. Discontinuities may be due to poor calibration techniques or inadequate rescaling operations. Multiple modes in a histogram can occur when an image contains several different features that have discrete reflectance levels in the wavelength range of the channel. For example, water and vegetation features have very different reflectance levels in near infrared wavelengths (approximately 700 nm–1100 nm) so an image containing water and land features could produce a near infrared channel histogram similar to the one illustrated in Figure 4.4 (see also Figure 4.10). When highly contrasting features constitute an image, it is possible for the distribution to become disjoint (that is separated) along the pixel value (horizontal) axis.

Figure 4.4 A bi-modal histogram

A bi-modal histogram occurs when an image channel contains two features that have distinctly different (non-overlapping) radiance values. For example, the near infrared reflectance for water is typically low (such as histogram peak labelled A) while vegetation reflectance is higher (such as histogram peak labelled B; see also Figure 4.10).



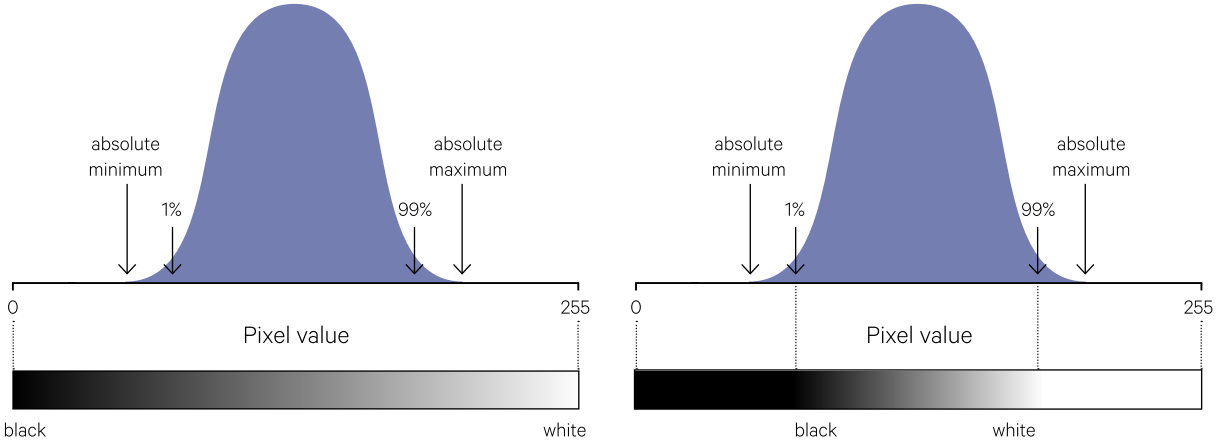
Source: Harrison and Jupp (1990) Figure 6

Various image processing operations modify image representation by altering the image histogram range or shape. When displaying an image, for example, it is desirable to achieve as much contrast as possible. Any display device has a finite number of grey levels between the darkest and brightest colours (see Section 6.1).

Figure 4.5 Histogram stretching for image display

a. 'Unstretched' histogram with the theoretical minimum being mapped to black and the theoretical maximum being mapped to white. This assignment does not consider the actual range of values that occurs in a particular image channel.

b. Stretched histogram display where minimum and maximum values of the image channel are used to define contrast limits (see also Figure 4.8).



Source: Harrison and Jupp (1990) Figure 7

In the case of the 8-bit (byte) format channel data shown in Figure 4.2 and Figure 4.3, if pixels with value 0 are displayed as black and pixels with value 255 are displayed as white, the displayed image would only contain mid-grey shades. To utilise the available contrast range of the display device, usually the minimum or 1% value is displayed as black and the 95% or 99% value as white (see Figure 4.5). This process is called image stretching and is further discussed in Section 4.2.1.

The histogram shape may also be altered to improve image contrast (see Section 4.2). This is useful in some applications to more clearly delineate particular features. Other aspects of generating and interpreting image histograms are considered in Section 8.1.1.

4.1.2 Variance

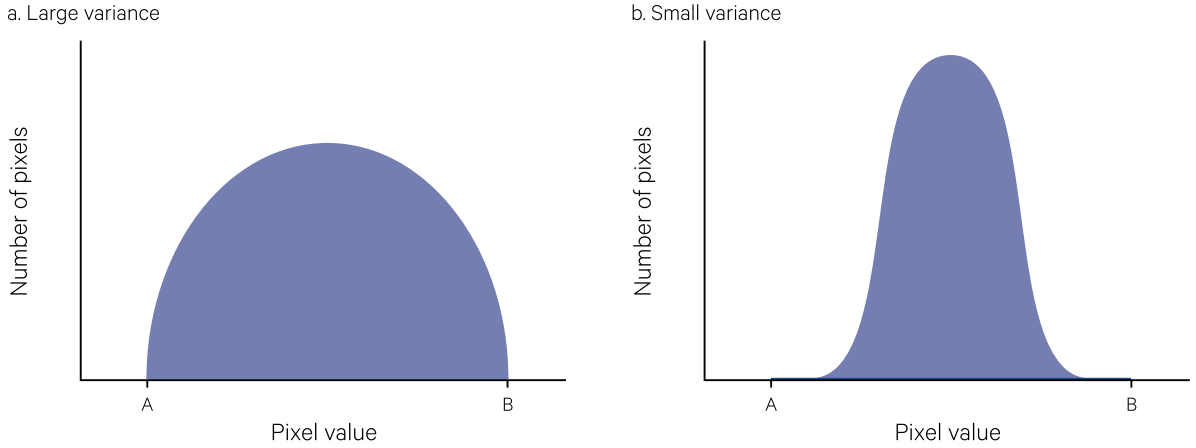
While the range of data values in a channel is an indication of the channel variability, it does not indicate the way the values are distributed between the minimum and maximum values. Figure 4.6 shows two histograms with the same range but different variance values.

Figure 4.6 Histograms with equal ranges and differing variances

Pixel value A indicates the channel minimum value, while value B is the maximum value in the channel.

a. Large variance

b. Small variance



Source: Harrison and Jupp (1990) Figure 8

A statistical measure called variance (σ^2) summarises the differences between all pixel values and the mean value for the channel. Mathematically, channel variance is the average of the squares of the deviations of the pixel values from their mean value and is calculated as:

$$\text{channel variance}\left(\sigma^2\right) = \frac{\sum\left(x-\mu\right)^2}{n}$$

where

x is pixel value in channel;
n is total number of pixels in channel (excluding null pixels); and
μ is true mean of the channel data.

Calculation of these statistics is illustrated in Excursus 4.1 for a simple example image channel. Another statistic used to indicate channel variability is equal to the square root of the variance value, and is called the 'standard deviation' (σ). Use of variance and standard deviation statistics in image analysis are further discussed in Section 8.1.2.

Excursus 4.1—Channel Statistics

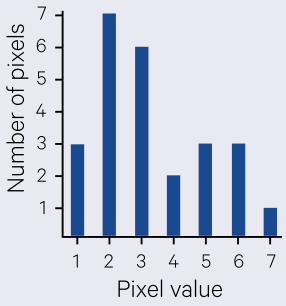
Using the variance equation given above, channel statistics are computed for an example image channel comprising 25 pixels (5 pixels in 5 rows). The example image values and the image frequency histogram are shown in Figure 4.7.

Figure 4.7 Example image

a. Example image values

1	2	3	3	4
3	3	5	6	3
2	2	2	6	5
5	4	2	2	1
7	6	3	2	1

b. Frequency histogram



Source: Harrison and Jupp (1990) Figure 9

Calculation of the variance statistic for this channel is detailed in Table 4.1. As is frequently the case in EO images, the distribution of pixel values about the mode is not symmetric.

Statistical measures of central tendency:

mode (most common value) = 2
median (midpoint of range) = 4
mean (average; μ) = 83/25 = 3.32

Statistical measure of data extent:

range (maximum - minimum) = 7-1 = 6

Statistical measures of dispersion from mean:

variance (σ^2) = 73.33/25 = 2.933

standard deviation $\left(\sigma\right) = \sqrt{2.933} = 1.713$

Table 4.1 Calculation of variance statistics

Pixel value	Frequency count	Pixel value x frequency count	(Mean value – pixel value)^2	(Mean value – pixel value)^2 x frequency count
1	3	3	5.29	15.75
2	7	14	1.69	11.83
3	6	18	0.09	0.54
4	2	8	0.49	0.98
5	3	15	2.89	8.67
6	3	18	7.29	21.87
7	1	7	13.69	13.69
Total	25	83		73.33

4.2 Contrast enhancement

Since any display (or hardcopy) device can only show a finite number of grey shades or colours, in order to optimise contrast in the viewed image we need to match the range of image values to the available range of shades/colours in the device. Various techniques can be used to match these ranges including:

- linear stretching (see Section 4.2.1);
- non-linear stretching (see Section 4.2.2);
- statistical stretching (see Section 4.2.3); and
- lookup tables (LUT; see Section 4.2.4).

4.2.1 Linear stretching

For a single image channel, if pixels equal to the potential minimum image range value are displayed as black and those near the potential maximum value in the image are displayed as white, with the values between these minimum and maximum values being linearly mapped to discrete shades of grey between black and white, the resulting image may not provide adequate contrast between features of interest. If the dynamic range of the data does not cover the potential range, the variance in the displayed image—and hence its overall contrast—may be increased by linear stretching.

A display device using an 8-bit colour space can represent 256 shades of grey in a displayed image, while a 10-bit colour space represents 1024 shades (see Section 6.1). If an image contains no pixels with values equal to the display-minimum value there will be no pixels showing as true black in the displayed image. Similarly, if the display-maximum value is set at 254 for an 8-bit colour space but there are no values above 200 in the image, the 55 'brightest' shades of grey on the display screen will not be used in displaying the image—only mid-grey and dark grey shades would be displayed.

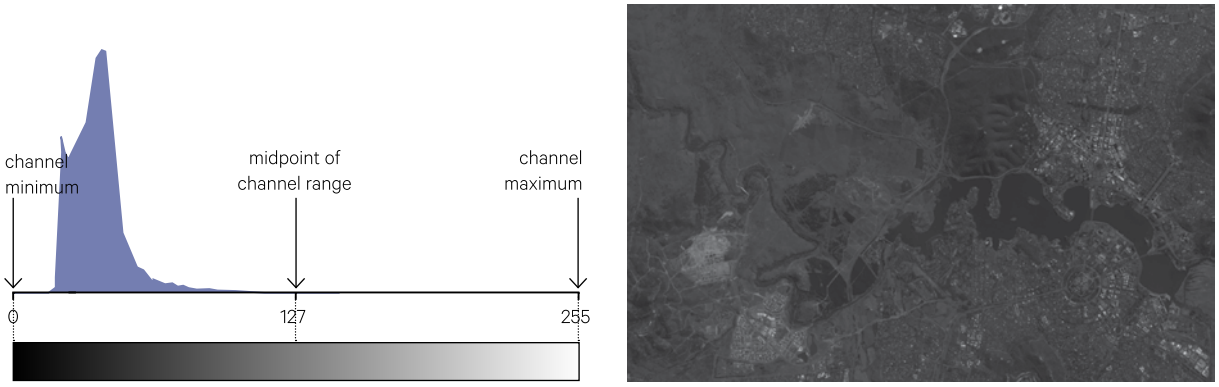
If the actual minimum and maximum values are displayed as black and white respectively, with the intermediate values again being linearly assigned to the grey shades in between, we would be utilising the full dynamic range of the display device but, due to the histogram distribution shape, most pixels would still be displayed as mid-grey shades. Alternatively, display limit values close to the 1% minimum and 99% maximum in the channel (that is, the values which are respectively above and below only 1% of all pixel values—or the central 98% of values) would reduce the range of image values that are displayed by matching this range with the major data variations in the image. This process is illustrated in Figure 4.5. In some cases the 2% and 98% values are used as the display minimum and maximum values (see Figure 4.8c).

This simple method of contrast enhancement is referred to as a linear stretch since the image values are being proportionally stretched to match the available display range. The minimum and maximum values selected for a stretch depend on the nature of the image data. In most cases the 1% and 99% (or 1% and 95%) values render displayed images with good contrast.

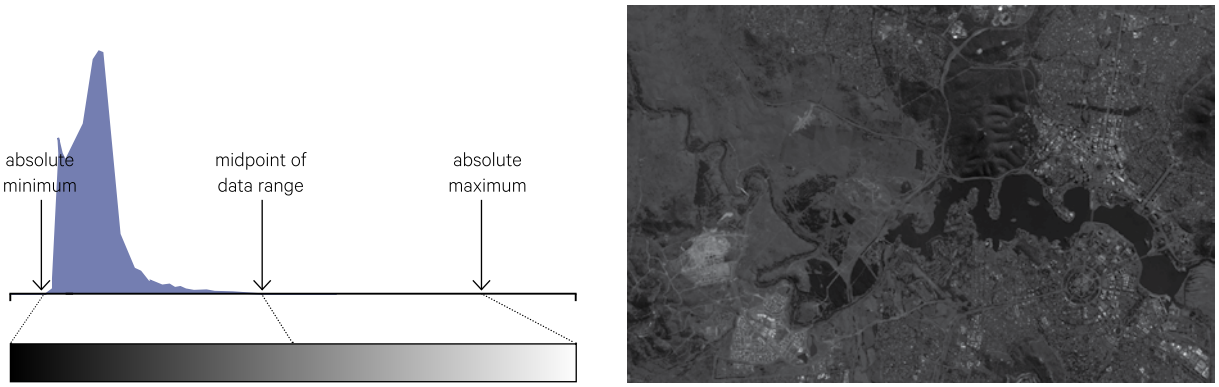
Figure 4.8 Contrast enhancement

This panchromatic image of central Canberra was acquired by Landsat-8 on 10 August 2017.

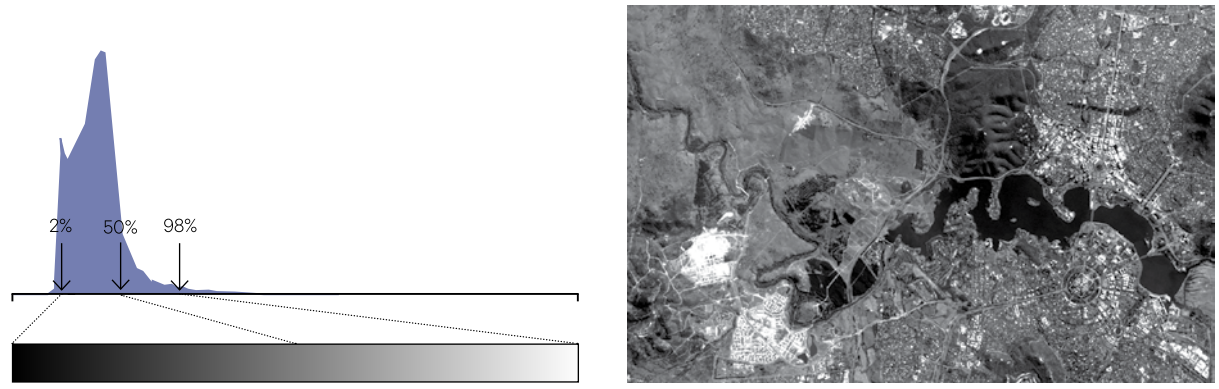
a. Unstretched image contains no true white or black pixels with most of the image being dark grey.



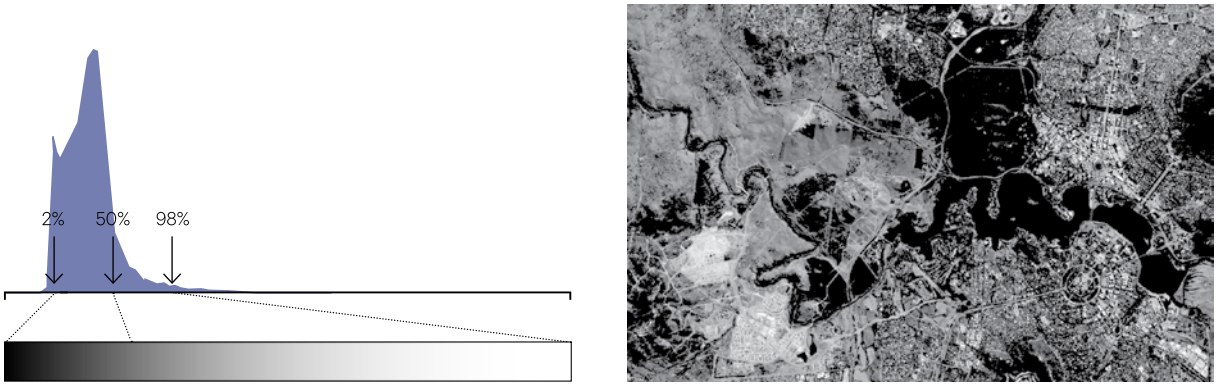
b. An image stretched between its minimum and maximum values contains some black and white pixels but most are still mid to dark grey.



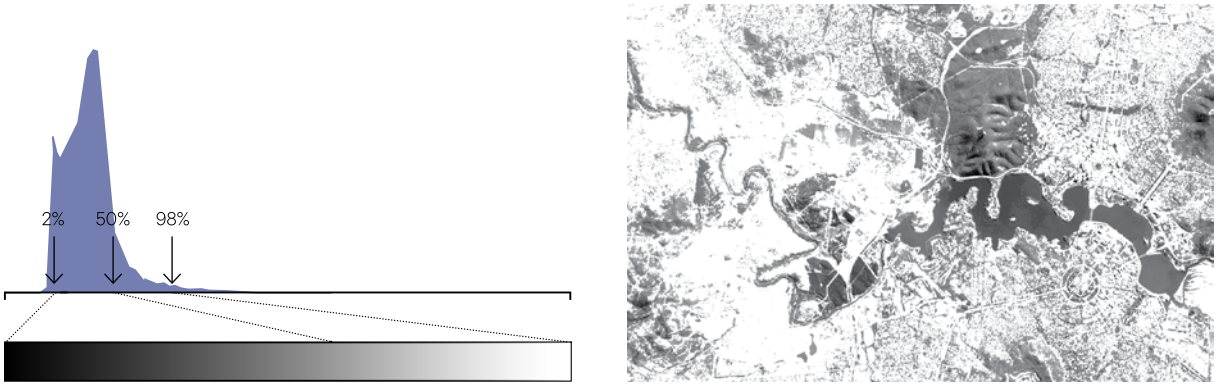
c. An image stretched between the 2% and 98% values uses the available range of grey shades more effectively.



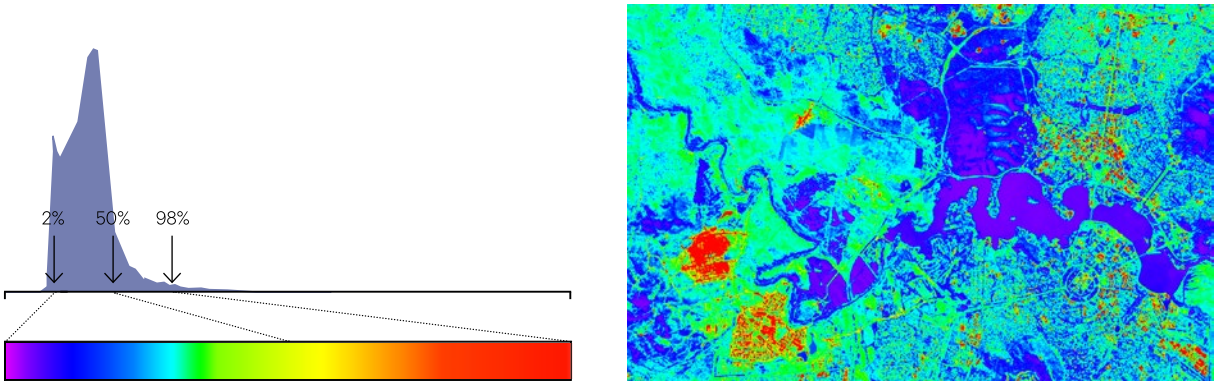
d. A non-linear (high Gaussian) stretch maps the high image values to a greater proportion of grey shades than the low image values. This improves the contrast in brighter areas at the expense of darker areas.



e. Non-linear (low Gaussian) stretch maps the low image values to a greater proportion of grey shades—this improves the contrast in darker areas but reduces contrast in lighter areas.



f. Pseudo-colour enhancement assigns different colours to different image values (see Section 5.2.1).

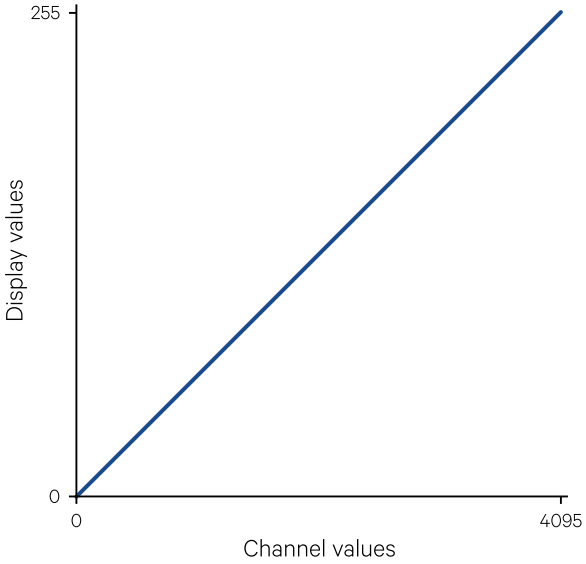


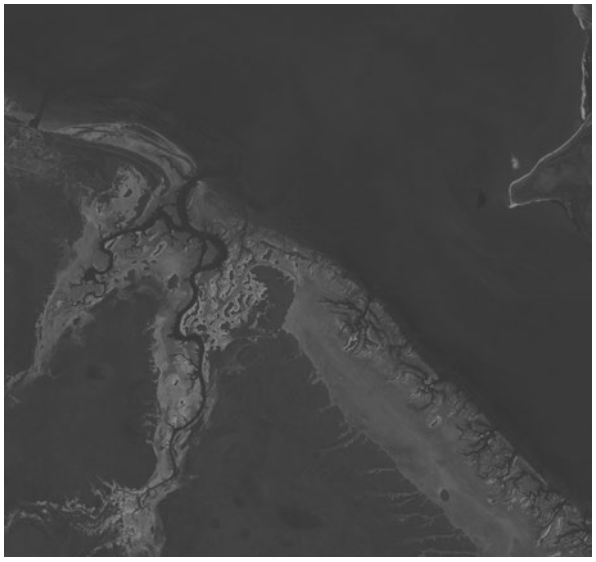
Source: Norman Mueller, Geoscience Australia

Figure 4.9 Linear stretching

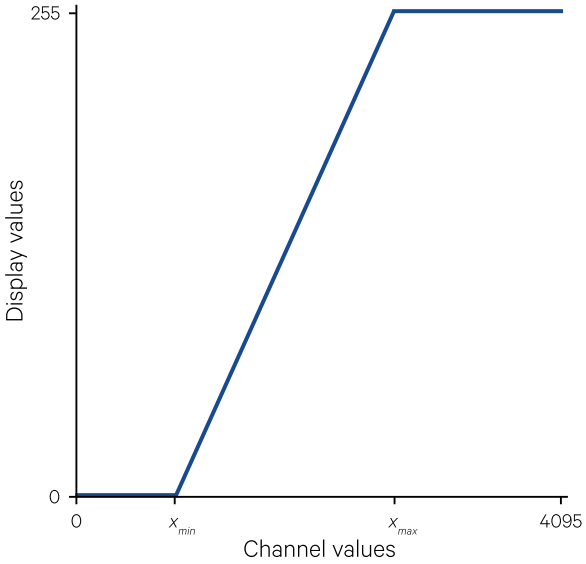
This panchromatic band was acquired by Landsat-8 on 7 July 2017 over Gulf of Carpentaria, northwest of Borroloola, NT.

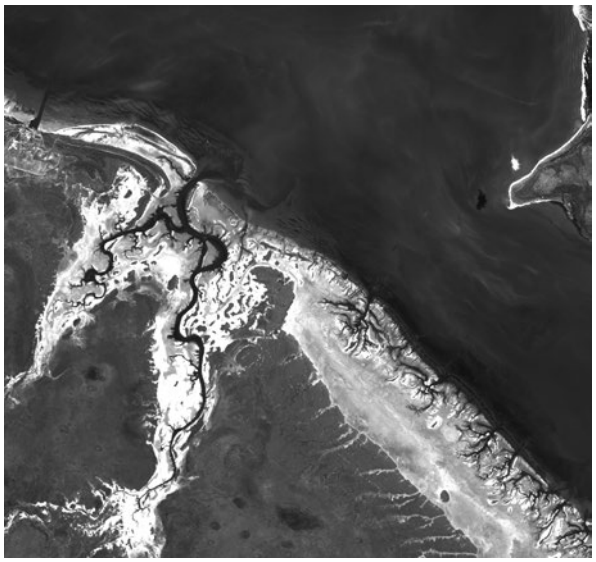
a. 12-bit Image channel displayed with value 0 as display minimum and value 4,095 as display maximum.





b. The range of image values between the nominated image minimum (x_{min}) and maximum (x_{max}) are linearly spread between the display minimum and maximum values. Image pixels with values below x_{min} are displayed as the display minimum and those with values above x_{max} are displayed as the maximum display value.





Source: Norman Mueller, Geoscience Australia

Linear stretching increases the overall contrast in a displayed image by increasing its variance. As illustrated in Figure 4.9, for an image channel being mapped to the display values 0–255, the range of displayed values, *y*, are related to the image value, *x*, as:

$$y = \begin{cases} 0 & \text{if } x < x_{min} \\ 255 \times \dfrac{x - x_{min}}{x_{max} - x_{min}} & \text{if } x_{min} \le x < x_{max} \\ 255 & \text{if } x \ge x_{max} \end{cases}$$

The variance of the displayed image values (y^2) will be approximately equal to:

$$\sigma_y^{\,2} = \left(\frac{255}{x_{max} - x_{min}} \right)^2 \times \sigma_x^{\,2}$$

so that, as the range in the image values is reduced, the variance of the displayed image increases, with values below x_{min} and above x_{max} being 'clipped' (that is, given the minimum and maximum values respectively). The contrast of a specific feature, such as a boundary, can be similarly increased by selecting minimum and maximum values that are appropriate to that feature.

The linear stretching technique assumes that the image histogram is uni-modal (see Section 4.1). Images with multi-modal histograms will require a different method for selecting these minimum and maximum values to optimise contrast in one of the value ranges. For example, Figure 4.10 shows a histogram of a thermal infrared channel for a daytime image which contains both land (warm) and water (cool) pixels. The unstretched image is shown in Figure 4.10a, and a linear stretch across both histogram peaks is shown in Figure 4.10b. The latter does not highlight features in either land or water. A stretch of values within the histogram peak with the lower values would highlight differences in the water but lose contrast in the land pixels (see Figure 4.10c), while a stretch of values spanning the histogram peak with higher values would produce good contrast in the land areas but cause water pixels to be displayed as black (see Figure 4.10d). (In this case, image segmentation methods could be used to separate the image into land and water regions, then rescale each segment separately before recombining into one image—see Sections 8, 9 and 10).

The contrast within a specific feature or region in an image may also be improved by using the minimum and maximum (or better, the 1% and 99%) values of that feature. In some cases these may be easily derived from the channel histogram as in the example in Figure 4.10. However, more commonly the appropriate minimum and maximum values would be identified using interactive training on the feature (see Section 9.1.2).

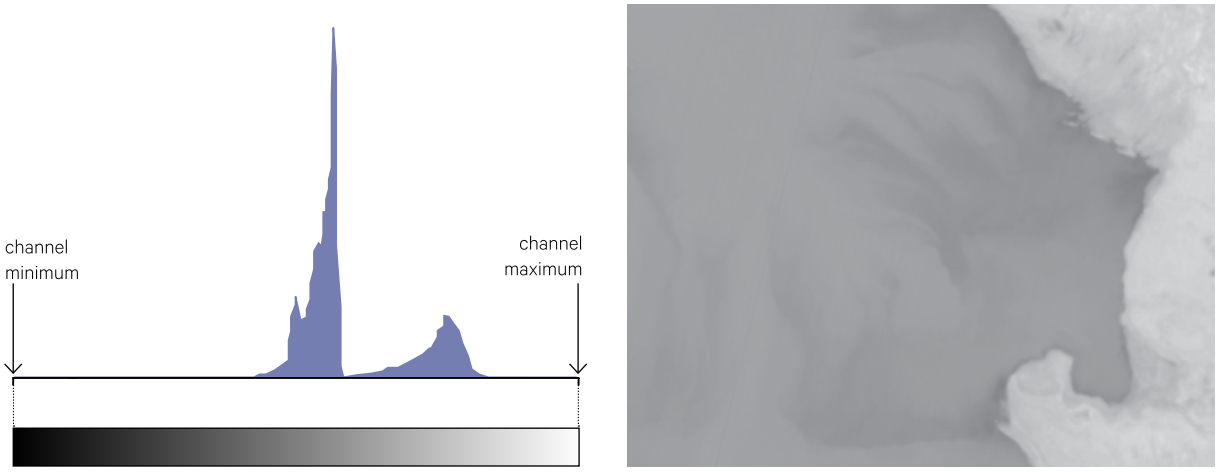
The channel data range may also be displayed as an inverse grey-scale by associating the contrast maximum value with black and the minimum with white (see Figure 4.10e). This display mode is usually referred to as an inverse or negative channel and is often represented as a negative channel number. Inverse channel display is useful for some temperature channels that are coded as increasing image numbers for decreasing temperature.

Finally, either pseudo-colouring (see Section 5.2.1 and Figure 4.10f) or density slicing (see Section 9.2.1) can also be used to visually enhance single image channels by adding colour.

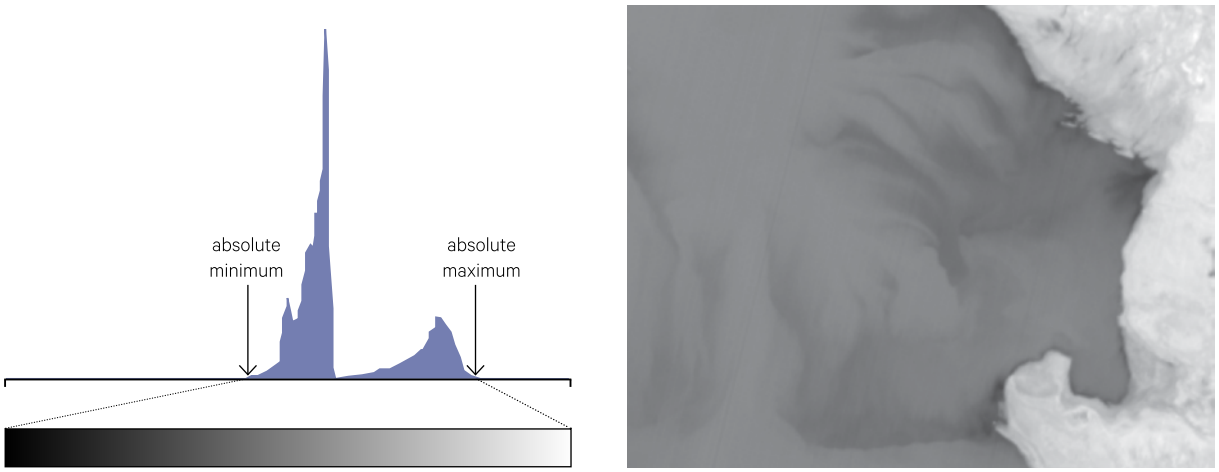
Figure 4.10 Bi-modal histogram stretching

This Landsat-8 thermal channel over Shark Bay, WA was acquired on 3 August 2017.

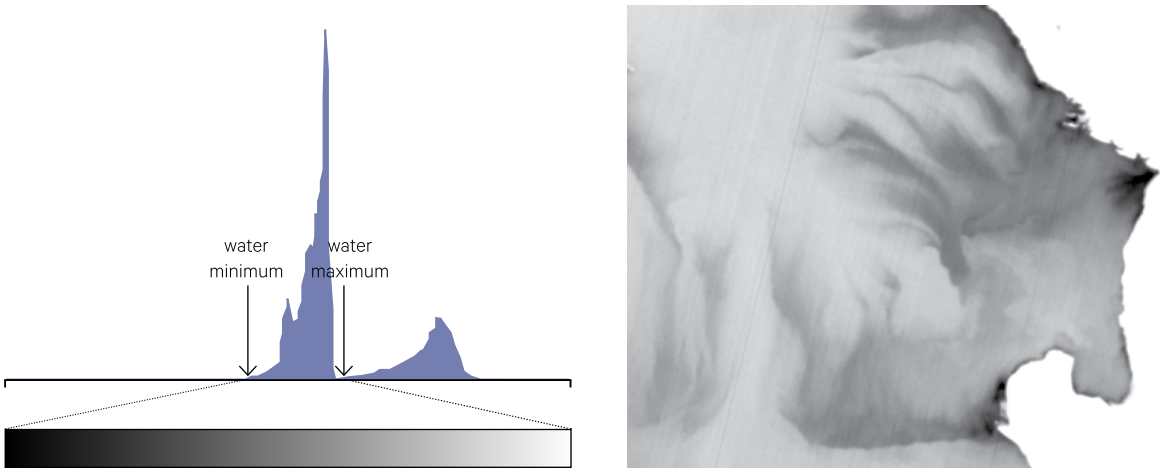
a. Original unstretched image



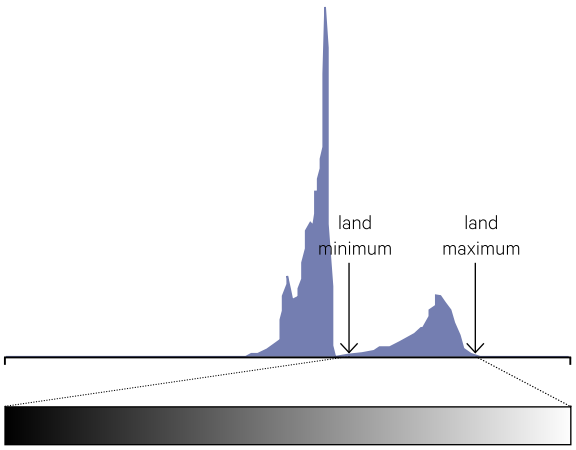
b. Linear stretch over full range of image data channel

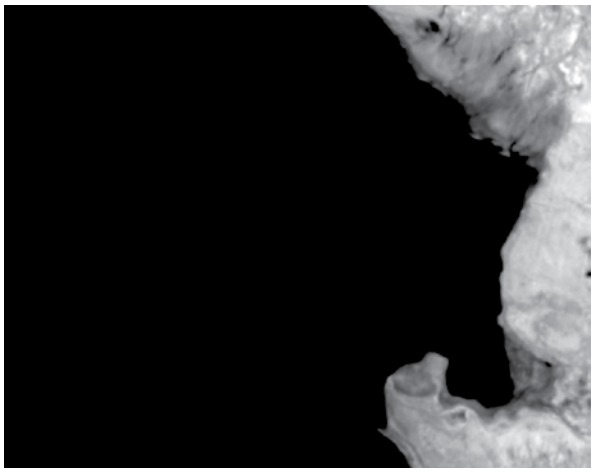


c. Stretching a range of low channel values defined by the first histogram peak improves contrast in the water but loses all contrast in the land areas.

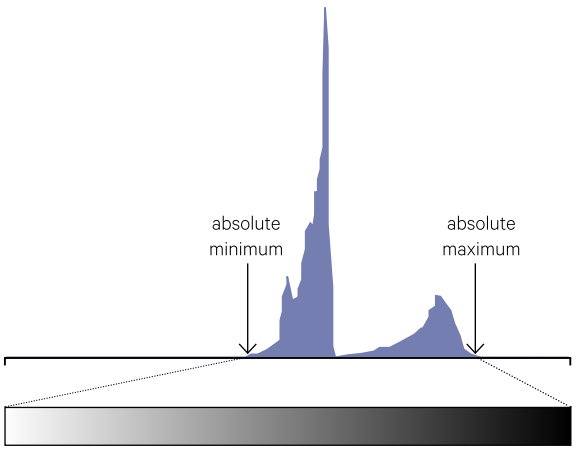


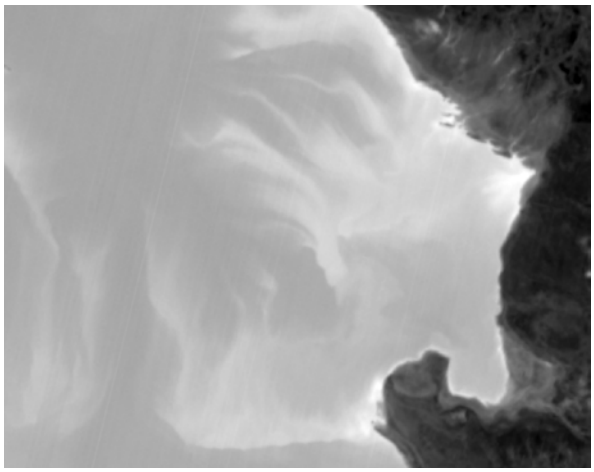
d. Linear stretching of the second histogram peak loses all water contrast to maximise contrast in land features.



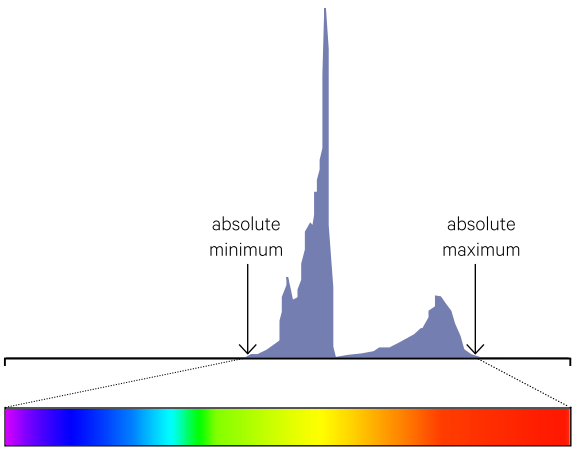


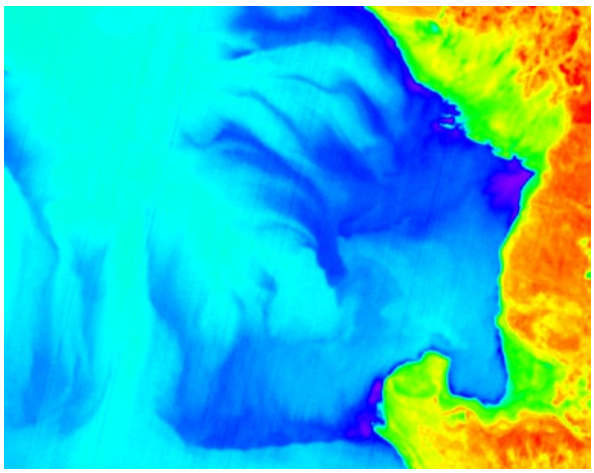
e. The image data range can also be inverse scaled so that high image values indicate low temperatures (darker) and low image values indicate high temperatures (lighter).





f. Pseudo-colouring of the full histogram range (with a different colour for each image value) improves visual discrimination over most of the image.





Source: Norman Mueller, Geoscience Australia

4.2.2 Non-linear stretching

The mapping of image values (within the defined contrast range) to the available display grey shades (or colours) may also be done in a non-linear fashion as illustrated in Figure 4.8d and Figure 4.8e. Image data are typically rescaled in a non-linear way using the logarithmic or exponential functions.

A logarithmic function is illustrated in Figure 4.11a. This graph represents the equation of:

$$y = \log_{10}(x)$$

where

$$x = 10^{y}$$

The exponential function is expressed as:

$$y = \exp_{10}(x)$$

where

$$\exp_{10}(x) = 10^{y}$$

(see Figure 4.11b).

These functions are often expressed more generally as $\log_a(x)$ or a^x where the choice of a affects the scaling and properties of the functions. When a is the natural constant 'e', the logarithmic function, $\log_e(x)$ or *ln*, is referred to as the Napierian logarithm and has special properties:

$$a^{x} = e^{\log(a^{x})} = e^{x \log(a)}$$

The base 10 number system computes these functions with the simply remembered values of:

$$\log_{10}(1) = 0 \qquad \exp_{10}(0) = 1$$
$$\log_{10}(10) = 1 \qquad \exp_{10}(1) = 10$$
$$\log_{10}(100) = 2 \qquad \exp_{10}(2) = 100$$

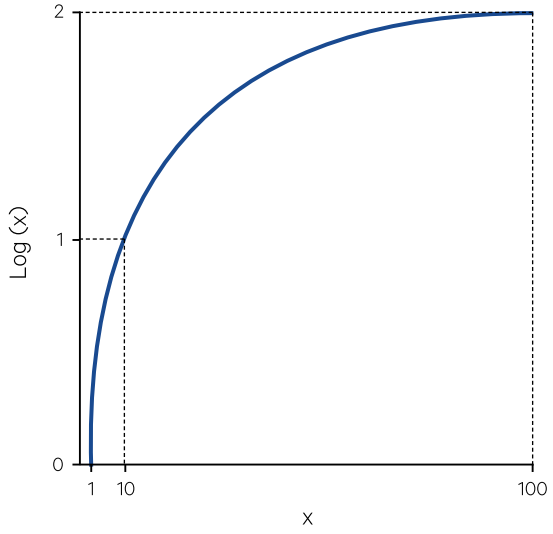
as illustrated in Figure 4.11.

As these figures and sample values show, the exponential is the inverse function of the logarithm so can be used to convert logarithmically-scaled data to a linear scale. Similarly, the log of an exponent produces linear scaling, that is:

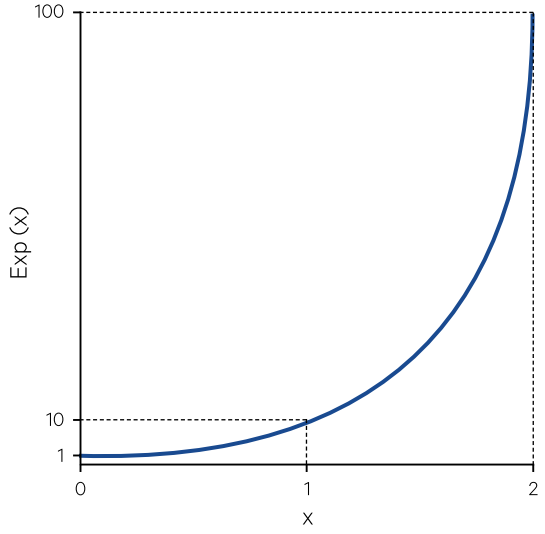
$$x = \log(\exp(x)) = \exp(\log(x))$$

Figure 4.11 Logarithmic and exponential functions

a. Logarithmic function: y = $\log_{10}$ (x)



b. Exponential function: y = $\exp_{10}$ (x)



Source: Harrison and Jupp (1990) Figure 65

Non-linear stretches are frequently implemented using an exponential function in the form:

$$y = \begin{cases} A & \text{if } x \le x_{min} \\ A + B \times \left(x - x_{min} \right)^{\gamma} & \text{if } x_{min} < x \le x_{max} \\ A + B \times \left(x_{max} - x_{min} \right)^{\gamma} & \text{if } x > x_{max} \end{cases}$$

where

x is an image pixel value;
x_{min} is the selected minimum image value for stretching;
x_{max} is the selected maximum image value for stretching;
y is the output (displayed) grey-scale value;
A is the minimum output level;
B is the scaling factor between the input and output ranges; and
γ is the gamma stretch parameter.

A linear stretch is equivalent to a gamma value of 1.0. A gamma stretch value (γ) less than one produces the result shown in Figure 4.8e, where contrast is increased in the low values (dark areas in the displayed image) but reduced in the high values (bright areas). Conversely, a gamma value greater than one results in contrast being improved in the high values (bright areas) and reduced in the low values (dark areas; see Figure 4.8d and Figure 4.12b). Gamma stretch parameters less than 1.0 are usually required during image printing to improve matching of colour intensities between the display screen and the printer (see Section 6).

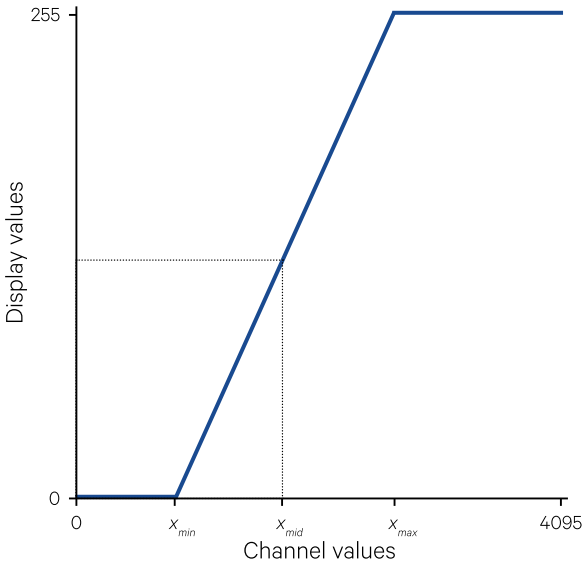
In terms of image contrast, a log transformation produces a similar result to a gamma function when the gamma value is less than one (see Figure 4.13), while the exponential transformation results are similar to a gamma function with value greater than one. Logarithmic and exponential stretches can also be implemented as image transformations as detailed in Volume 2C. Other rescaling methods, such as destriping, are also discussed in Volume 2C.

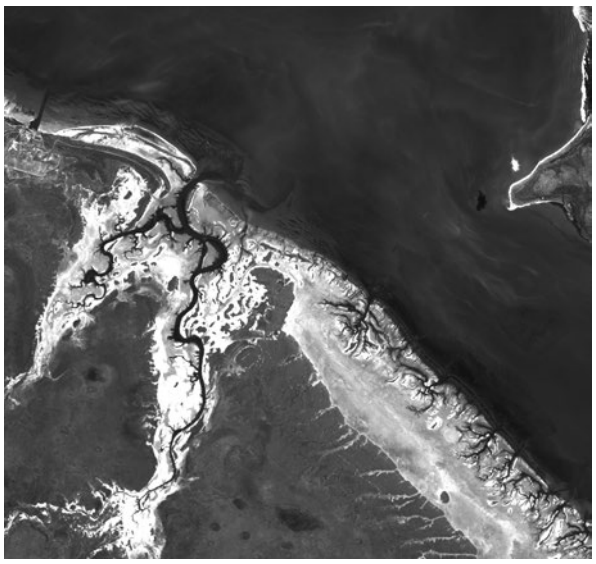
There are dark shadows on the earth,
but its lights are stronger in the contrast.
(Charles Dickens)

Figure 4.12 Enhancing high image values

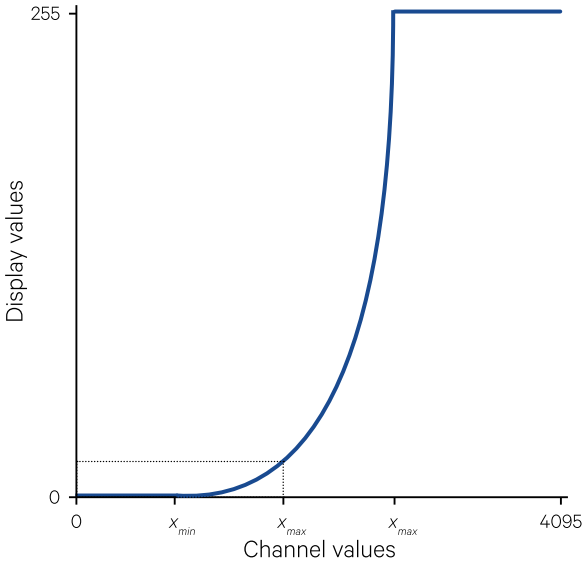
This panchromatic band was acquired by Landsat-8 on 7 July 2017 over Gulf of Carpentaria, northwest of Borroloola, NT.

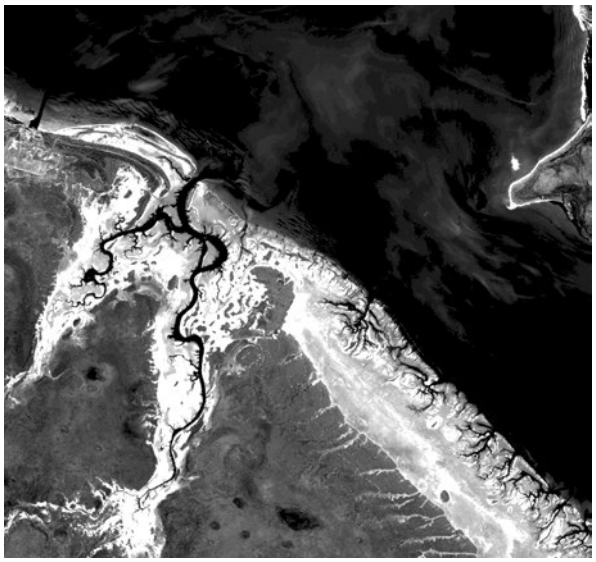
a. Linear stretch of panchromatic channel (see Figure 4.9)





b. Same image with high Gaussian stretch, which increases contrast in features with high image values.



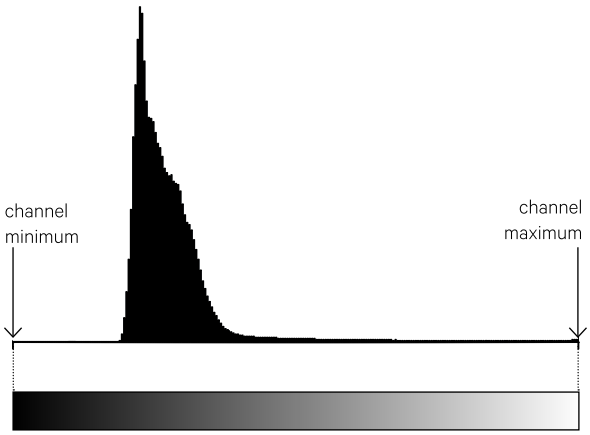


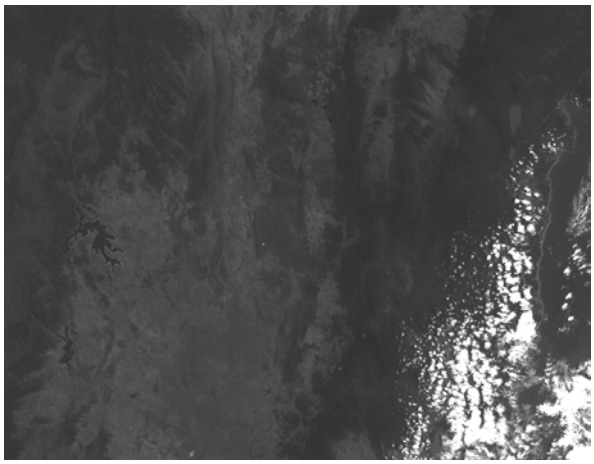
Source: Norman Mueller, Geoscience Australia

Figure 4.13 Logarithmic Stretching Example

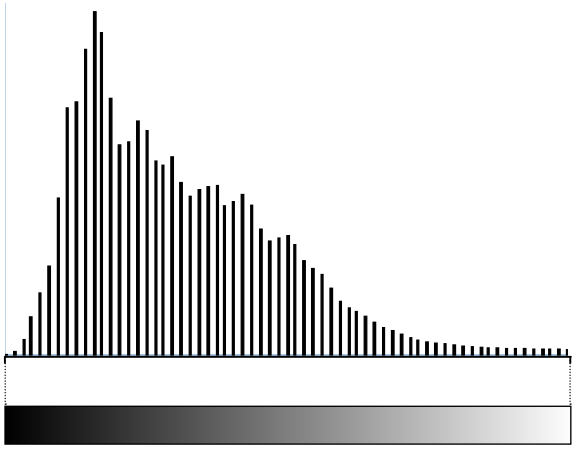
This example uses an 8-bit (byte) format image channel displayed using an 8-bit colour space. This image was acquired by Landsat-5 on 20 January 2010 over southeastern NSW. The town of Cooma is near the centre of this image, with Canberra in the northwest and the coastline visible in the east. Band 1 (visible blue) is used to differentiate between spectral features of mountainous terrain with forest cover, cleared pasture, water bodies, and cloud.

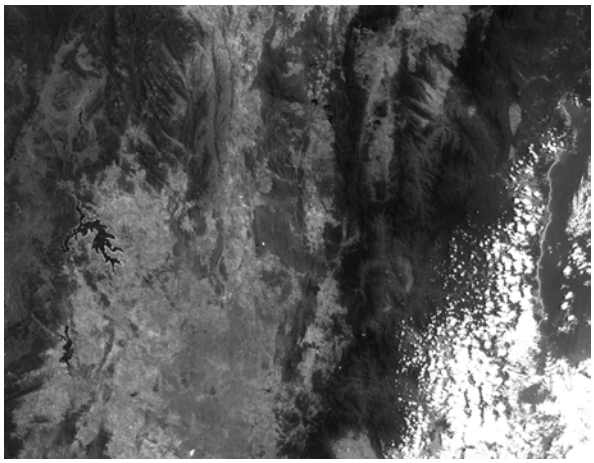
a. Unstretched image



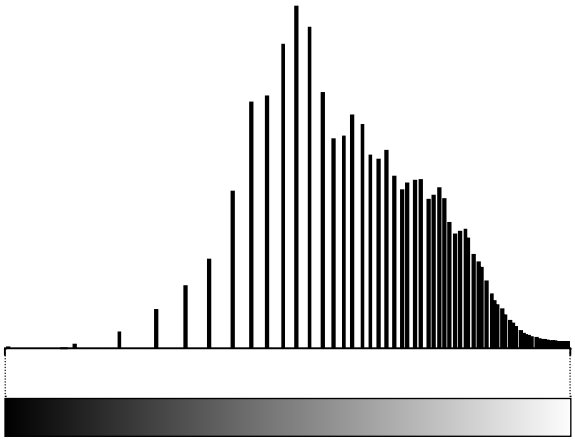


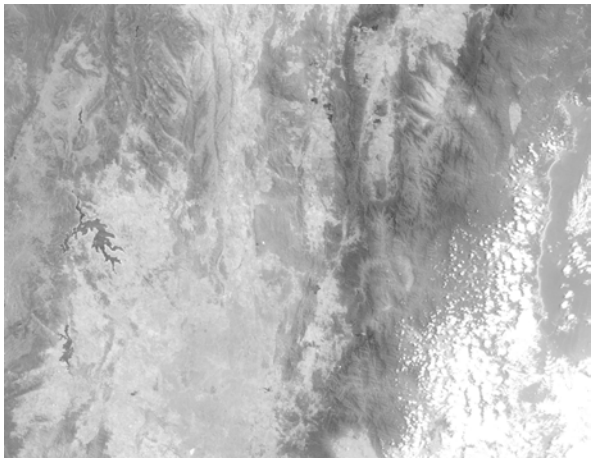
b. Linear stretch





c. Logarithmic stretch, which increases contrast in low image values.





Source: Tony Sparks, Icon Water

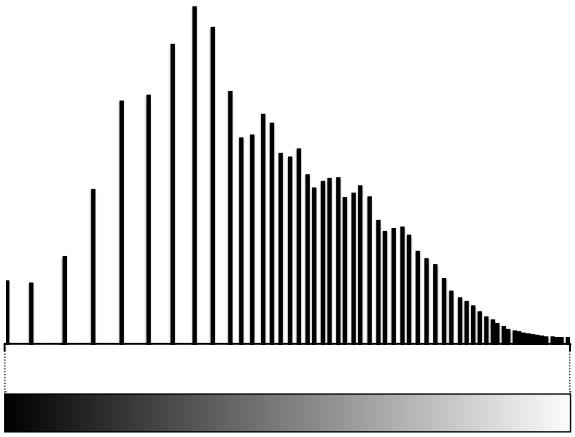
4.2.3 Statistical stretching

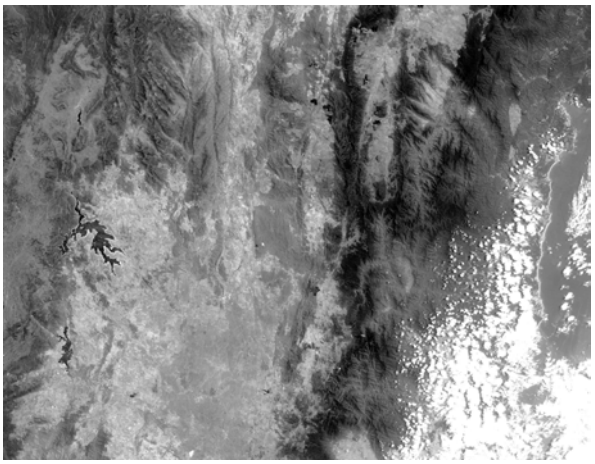
4.2.3.1 Normalisation

Some image processing systems offer an option during image display to fit each image channel to a normalised distribution based on sample channel statistics (see Figure 4.14). For normally distributed data, 95% of data lies within two standard deviations of the mean. The (sample) channel statistics can be used to determine a minimum value equivalent to the mean value minus one standard deviation and a maximum value equivalent to the mean value plus one standard deviation. These statistically-determined minimum and maximum values thus span 95% of the data range. This channel distribution can also be used to spread image values appropriately between the selected minimum and maximum values (see Figure 4.14).

Figure 4.14 Normalised stretch

Normalised stretch applied to image shown in Figure 4.13.





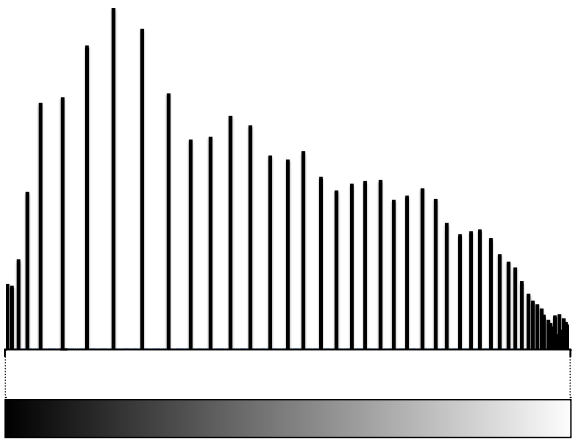
Source: Tony Sparks, Icon Water

4.2.3.2 Histogram equalisation

As the name implies, this technique attempts to equalise the number of pixels assigned to each value in the image. If values were distributed equally in the rescaled image channel it would have a flat histogram. In EO imagery, however, the initial distribution of image values over a restricted data range generally disallows this result. Histogram equalisation is a useful method for both increasing overall variance and improving contrast between those values occurring most frequently in an image (see Figure 4.15).

Figure 4.15 Histogram equalisation

Histogram equalisation applied to image shown in Figure 4.13



Source: Tony Sparks, Icon Water

This operation is based on the grey level (or bin) histogram counts, n_1, n_2, ... n_N, for each image value x_j, where j = 1 to N. For a byte image, the rescaled values x_j' are computed as:

$$x_j' = \frac{255 \times \sum_{i=1}^{j} n_i}{\sum_{i=1}^{N} n_i}$$

A linear stretch can be included in the rescaling by defining:

n_j = image minimum value for $j < x_{min}$ and
n_j = image maximum value for $j > x_{max}$.

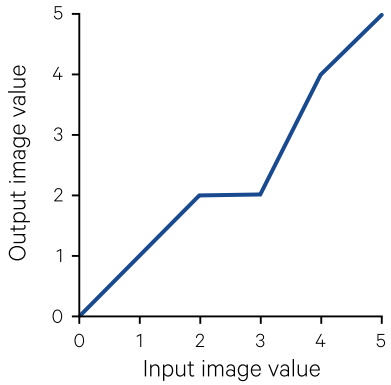
In many image processing systems, histogram equalisation can be applied as a statistical stretch during the image display process.

4.2.4 Lookup tables

A lookup table (LUT) is an expedient method for implementing a range of rescaling operations by allowing a defined output value for each possible input value in the image. A LUT is a single column table in which rows represent input data values and table entries indicate the output values to assign to each input value. The definitions can represent a wide variety of contrast enhancements such as piecewise linear stretch or sawtooth stretching (also called 'overlapping contrast enhancement') as illustrated in Figure 4.16. The impact of a user-defined LUT on a grey scale image is shown in Figure 4.17. A random LUT could also be generated to highlight small contrast differences between features by chance.

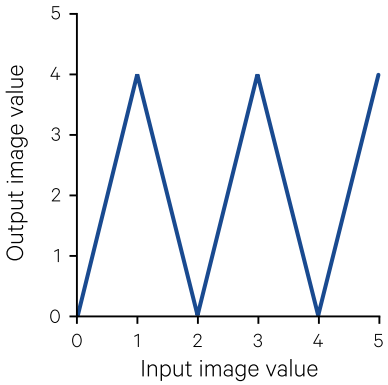
Figure 4.16 LUT operations for rescaling image data values

a. Piecewise linear rescaling



LUT entry	Output Value
0	0
1	1
2	2
3	2
4	4
5	5

b. Sawtooth stretch

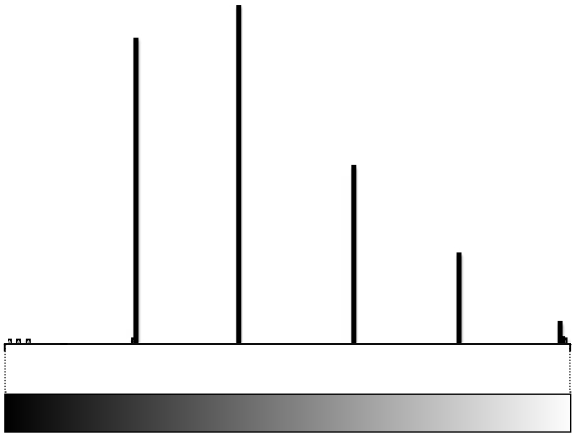


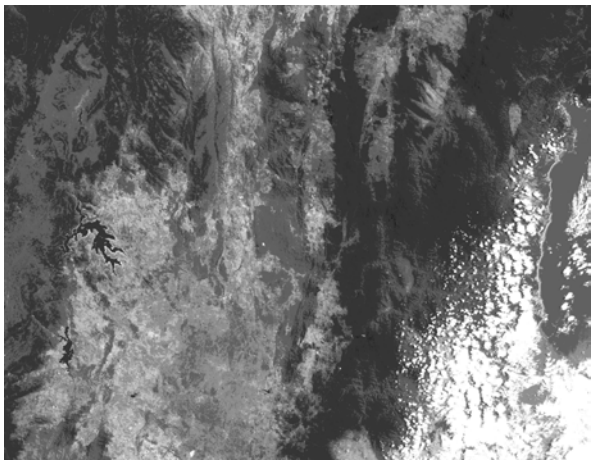
LUT entry	Output Value
0	0
1	4
2	0
3	4
4	0
5	4

Source: Harrison and Jupp (1990) Figure 10

Figure 4.17 User-defined stretch

User-defined stretch applied to image in Figure 4.13





Source: Tony Sparks, Icon Water

LUT can be used to implement statistical rescaling operations such as histogram equalisation (see Section 4.2.3.2). The use of LUT is further discussed in the context of pseudo-colouring (Section 5.2.1), image display (Section 6.1) and image transformations (Volume 2C).

4.3 Further Information

Brightness and Contrast: *https://www.tutorialspoint.com/dip/brightness_and_contrast.htm*
http://www.dspguide.com/ch23/5.htm

Castleman (1998) Chapter 5.

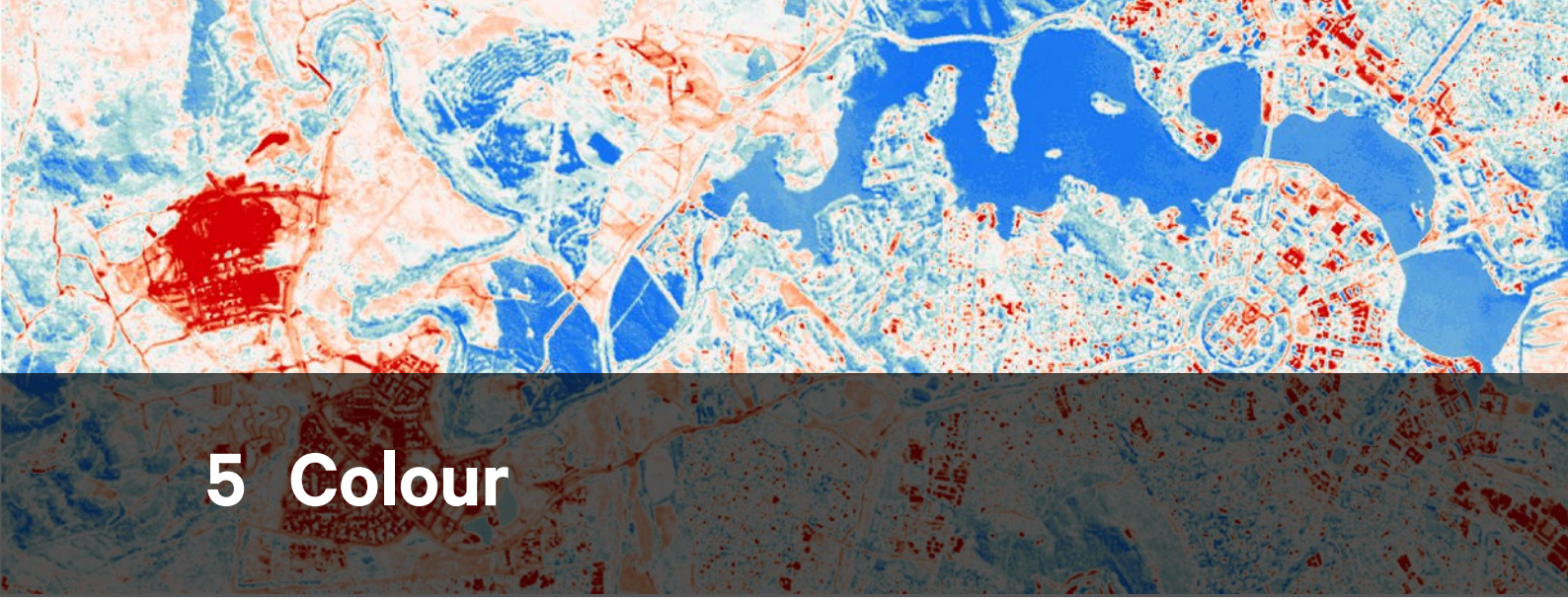
Jensen (2016) Chapter 4

4.4 References

Castleman, K.R. (1998). *Digital Image Processing*. 2nd edn. Prentice Hall. 667 pp.

Harrison, B. A., and Jupp, D. L. B. (1990). *Introduction to Image Processing. Part TWO of the microBRIAN Resource Manual* (256 pages). CSIRO Australia, Melbourne.

Jensen, J.R. (2016) *Introductory Digital Image Processing. A Remote Sensing Perspective*. 4th edn. Pearson Education, Inc.



5 Colour

Colour is an important aspect of image presentation and interpretation. Discussion so far has referred to representing image data as shades of grey between black and white. Since the human eye is not particularly good at detecting grey shades, image data can be interpreted more easily if displayed as a range of different colours. The type and range of colours used to represent data values in a colour image will directly affect our visual interpretation of the data by enhancing or reducing our perception of its patterns and features. Various contrast enhancement transformations discussed above attempt to improve this displayed colour range.

Colour look-up tables can be used to represent single channel image data using pseudo-colour techniques as described in Section 5.2.1 or by density slicing (Section 9.2.1), while multi-channel imagery can be represented as colour composites (Section 5.2.2). The broader issues of colour description, ordering and reproduction which are relevant to image representation are discussed initially in Section 5.1.

5.1 Colour basics

5.1.1 Colour description

Our common language descriptions for different colours are often vague and inconsistent making it difficult to use such descriptions to precisely define a particular colour. This situation is further exacerbated by the differences between the measured and perceived brightness of colours and the variations in visual perception within the human population. The latter factors are further discussed in Excursus 5.1.

Before considering some of the systems that attempt to quantitatively define colours, we will look at aspects of individual colours that may be used to describe our perception of them. Three attributes that are frequently used to describe colour are hue, value and chroma.

The term hue describes the attribute that is commonly referred to as colour, that is the combination of different wavelengths used to produce a colour. A 'natural' ordering of hues follows the colour spectrum (with increasing wavelength) from violet, through indigo, blue, green, yellow, and orange to red. The colours in this spectrum represent the spectral hues, that is, those produced by mixing adjacent wavelengths in the spectrum. Non-spectral hues are formed by mixing two or more spectral hues, such as purple (blue plus red) or brown.

The perceived brightness (lightness or darkness) of a colour is referred to as its value or intensity. This is different to the measured lightness (or luminance) of a colour. Intensity depends on the amount of energy present in all wavelengths of a colour.

Chroma describes the saturation or purity of a colour. This indicates the range of wavelengths around the average colour wavelength in which most energy is carried. A single wavelength would produce a spectrally pure colour and hence high saturation value, whereas a mixture of wavelengths, such as in a pastel colour, would have low value. The full spectrum of wavelengths, when combined, appears as white. This attribute is often linked with the colour value so can be difficult to describe and identify. It can also be described as the amount of grey in a colour since a high chroma (or vivid) colour can be made duller by mixing with grey.

The perceived hue, value and chroma of any colour are all affected by the patch size and shape of the colour, and the colours of surrounding patches. These factors are discussed in Excursus 5.1 and present obvious problems for colour description in systems that rely only on these perceptual attributes. Excursus 5.2 further considers the problems of colour interaction and how these may be resolved to achieve a balance of contrast and harmony in digital imagery.

Background image: Pseudo-colouring of Landsat-8 panchromatic image of Canberra, ACT, acquired on 10 August, 2017.
Source: Norman Mueller, Geoscience Australia

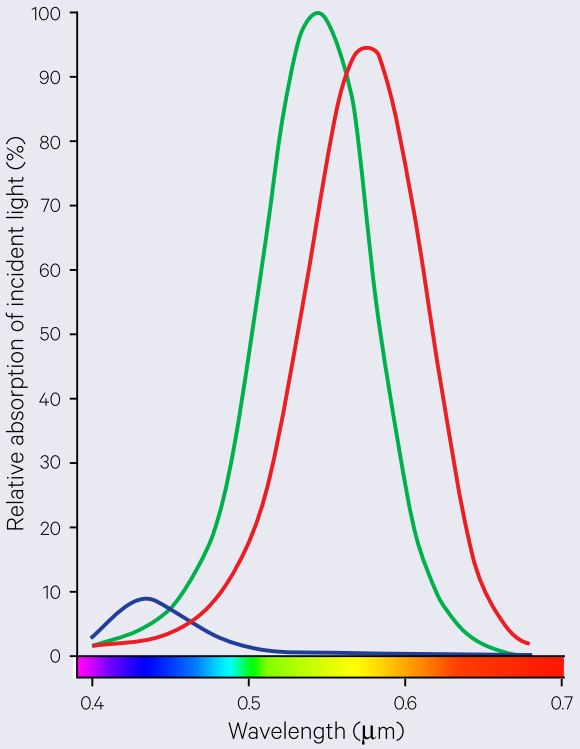
Excursus 5.1—Human Colour Perception

The precise mechanism of human colour vision is not well established. Various theories have been developed to try to explain this fundamental phenomenon. The currently accepted theory marries two previously competitive views as a two-stage process. These involve the physical, photochemical detection of a colour signal and its subsequent psychological processing.

The physical components of vision rely on two types of photochemical cells in the retina of the eye. The rods detect intensity changes independently of colour. The physical sensation of colour is believed to be due to the spectral sensitivities of three types of cones. These are commonly referred to as the blue, green and red receptors, which respond to short, medium and long wavelengths in the visible spectrum respectively (see Figure 5.1). These photochemical observations have been confirmed by various psychophysical colour mixing and matching experiments. This interpretation is often referred to as the tristimulus theory of colour vision and supports the observation that most colours can be produced from these three primary colours (see Appendix 3).

Figure 5.1 Relative light absorption of human cones

In the human eye, the relative absorption of light differs for different cones, with blue cones being significantly less responsive than green or red cones.

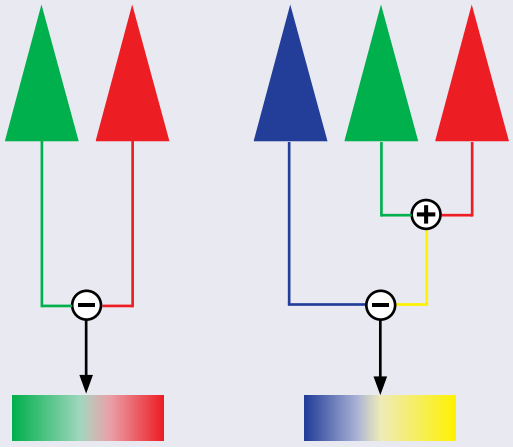


Source: Harrison and Jupp (1989) Figure 9 [Adapted from Cornsweet (1970)]

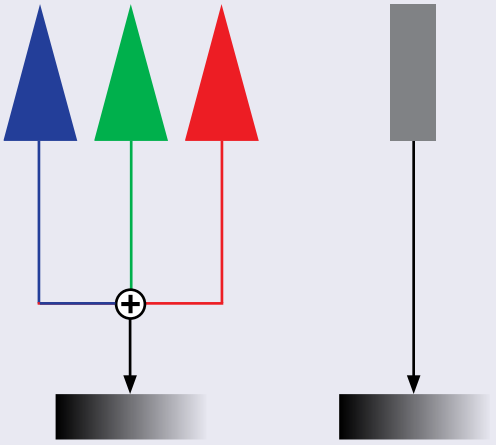
Figure 5.2 Opponent process theory of colour vision

This theory explains colour vision in terms of three colour planes. In the chromatic system, one plane (derived from the comparing the signals from red and green cones) transitions from red to green while the other creates a yellow signal from the red and green cones then compares this signal with the blue cone signal to produce a blue/yellow transition. In the achromatic system, signals from all cones are combined to produce a colour luminosity scale, which is used with the luminosity signal from the rods to perceive changes in brightness.

a. Chromatic system



b. Achromatic system



After the signals from these receptors have been coded, the brain analyses their differences to produce the psychological impression of colour. This analysis is explained by the opponent process theory (Hering, 1964), which proposes that colours are perceived in terms of three sets of 'opposing' colour pairs:

- red-green—red (R) and green (G) signals are compared to give a R versus G output signal (see Figure 5.2a);

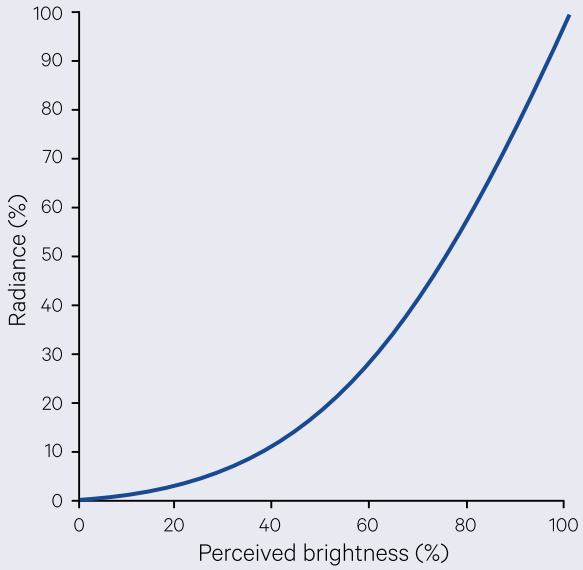
- yellow-blue—red and green are summed to produce a yellow signal (Y), which is then compared with the blue receptor signal (B) to produce a Y versus B response (see Figure 5.2a); and
- black-white—all receptor signals (R, G and B) are summed to produced a colour luminosity, which is processed in conjunction with an achromatic luminosity signal from the rod cells (see Figure 5.2b).

This recoding of the original signals reduces the transmitted colour information down to two colour difference signals and a luminance signal. A similar coding system of luminance and 'chrominance' was used in colour televisions to improve transmission efficiency (Padgham and Saunders, 1975).

Each individual may also see a particular colour differently depending on the sensitivities of his or her visual sensors. Figure 5.3 indicates the non-linear relationship between perceived and measured brightness. The context of a given colour intensity will also impact its perceived brightness (see Figure 5.4).

Figure 5.3 Perception of intensity changes

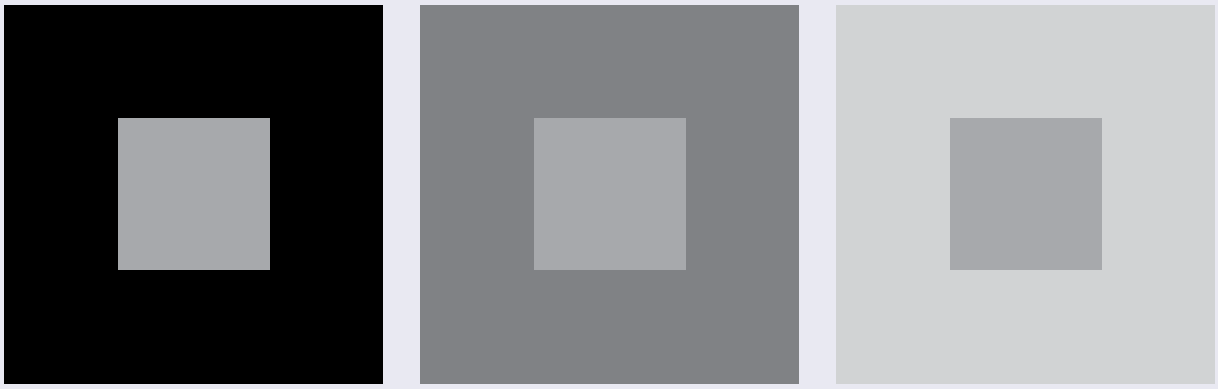
Differences in high intensities (brighter colours) are not well perceived, whereas smaller differences in low intensities (darker colours) are more easily perceived



Source: Harrison and Jupp (1990) Figure 39 [Adapted from Land (1977)]

Figure 5.4 Effect of adjacent intensities on perceived brightness

The central squares below all have the same brightness (40%) but appear as lighter and darker shades depending on the adjacent intensities (outer squares left to right: 100%, 60% and 20%).

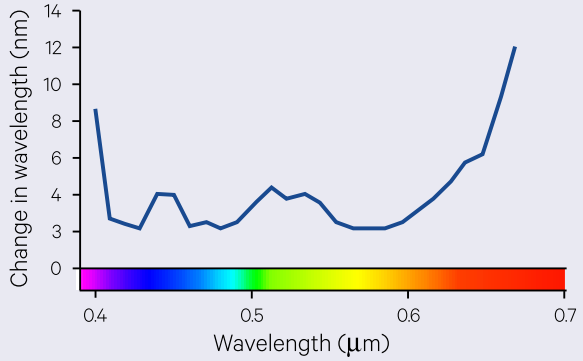


The spectral sensitivity of the human eye also varies with wavelength, being most sensitive to yellow-green colours. This means that a light radiating yellow wavelengths at a particular energy level will appear brighter than a light radiating the same amount of energy in red wavelengths. This phenomenon has obvious implications for generating imagery with apparently uniform brightness over a range of colours.

Another interesting variable in human vision is the threshold levels for perceiving chromaticity changes at different wavelengths. As shown in Figure 5.5, changes in chromaticity at wavelengths corresponding to blue/green or orange are more readily perceived than at wavelengths corresponding to blue, yellow/green or red.

Figure 5.5 Perception of chromaticity changes

The extent to which changes in chromaticity are perceived varies with wavelength.

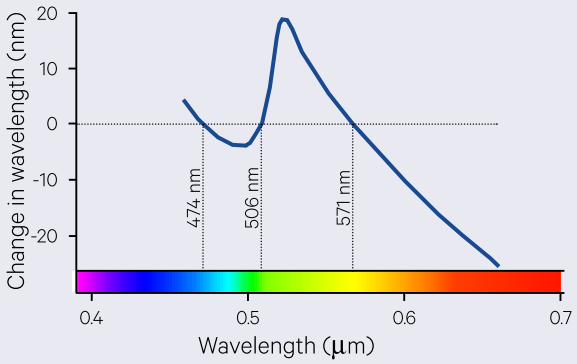


Adapted from: Cornsweet (1970)

It has been demonstrated that the human perception of spectral hue, like brightness and saturation, depends on both the wavelength and intensity of the stimulus ('Bezold-Brucke Phenomenon'). As illustrated in Figure 5.6 for a constant wavelength, hue varies with intensity except at three invariant points: 474 nm, 506 nm and 571 nm (Padgham and Saunders, 1975). Perceived hue has been observed to change with exposure time (Cornsweet, 1970) and colour saturation can also affect perceived brightness ('Helmholtz-Kohlrausch Effect'; Padgham and Saunders, 1975).

Figure 5.6 Effect of colour intensity change on perceived hue

The apparent change in wavelength at lower luminance levels is plotted against that wavelength at high luminance. Perceived hue varies with wavelength except at 474 nm, 506 nm and 571 nm.



Adapted from: Padgham and Saunders (1975)

5.1.2 Models for colour ordering

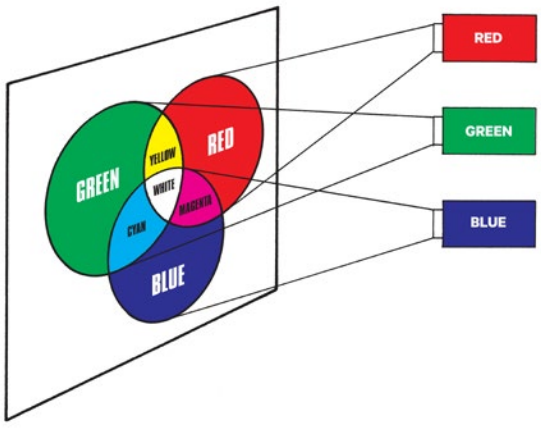
Many and various systems have been proposed to quantify or identify colours and colour differences (Niblack, 1986). These include generic models such as the simple RGB (red, green, blue) colour cube, the HSI (hue, saturation, intensity) colour space and various chromaticity diagrams as well as specialised systems such as the Munsell colour system, the CIE (Commission Internationale de l'Eclairage) models and industrial chromaticity standards.

The systems that are used to describe or identify colours rely on either technical characteristics, such as how a colour is produced or physically measured in terms of wavelengths, or perceptive qualities, that is, how we actually see it. This perception depends on both the physical colour and our visual processing of it, the latter being affected by both surrounding colours and lighting conditions. Some colour description systems attempt to use both physical and perceptual colour attributes. Colour spaces in which the perceptual difference between two colours is linearly related to the Euclidean, or linear, distance between them are termed uniform colour spaces. When image colours are selected from a uniform colour space, the visual discrimination of processing results is optimised.

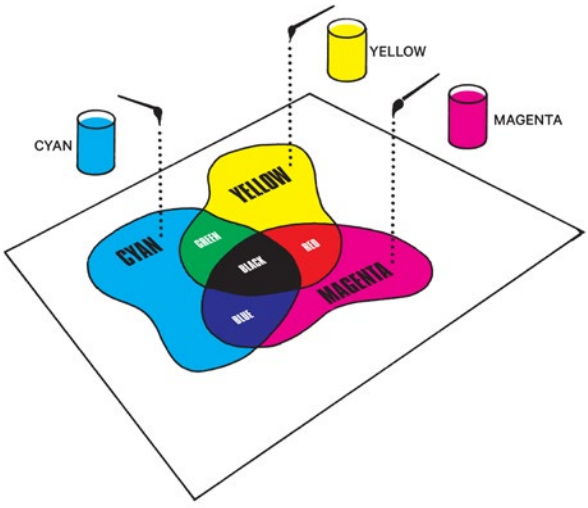
Why do two colours, put one next to the other, sing?
Can one really explain this? No.
Just as one can never learn how to paint.
(Pablo Picasso)

Figure 5.7 Additive and subtractive colour mixing

a. The additive primaries (red, green and blue) are added to a black background to give other colours.



b. The subtractive primaries (yellow, magenta and cyan) subtract from a white background to produce other colours.

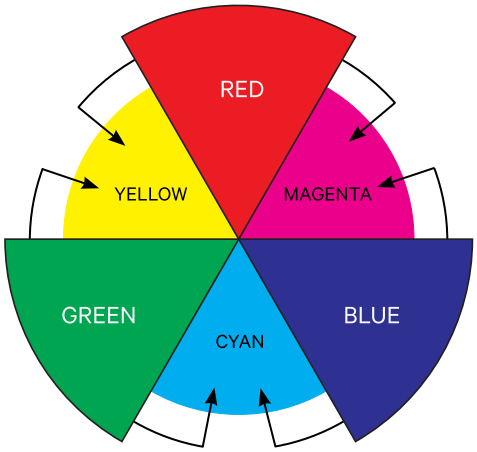


Source: Harrison and Jupp (1990) Figures 10 and 11

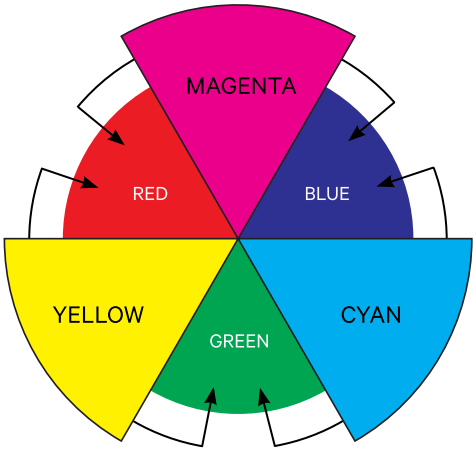
One of the simplest systems to represent colours relies on the additive primaries: red, green and blue (see Appendix 3 and Volume 1). The additive primary colours represent the basic hues that can be added together to produce white and all other colours (see Figure 5.7a and Figure 5.8a). Colour representation devices that operate by adding coloured light to a black base, such as projectors and display monitors, use additive primaries. Pairs of these primary colours can be added together to form the subtractive primaries, yellow (red plus green), magenta (red and blue) and cyan (green and blue). Film dyes and print media use the subtractive primaries to produce other colours by 'removing' their complementary colours from white (see Figure 5.7b and Figure 5.8b).

Figure 5.8 Relationship between additive and subtractive primary colours

a. Additive primaries—can be combined to form brighter, secondary colours; added all together they give white light



b. Subtractive primaries—produce secondary colours by removing their complementary colours from white; added all together they produce black



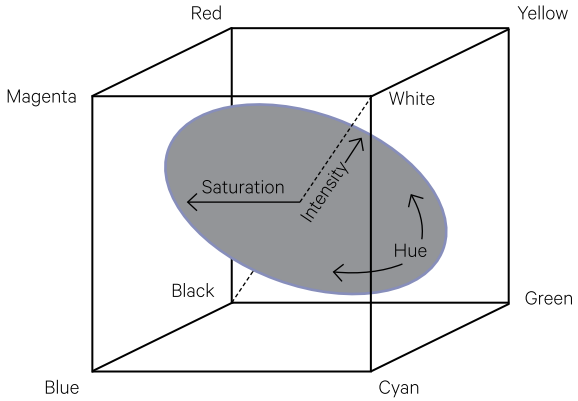
Source: Harrison and Jupp (1990) Figure 12

These colour components can be viewed as three dimensions forming a colour cube as illustrated in Figure 5.9. Due to the complementary nature of the additive and subtractive primaries, the two-dimensional 'corners' of this space represent the subtractive primaries yellow, magenta and cyan, with black and white at opposite corners of the space. Thus, the axis of grey shades is the three-dimensional line from black to white where the proportions of each primary colour are equal. A limitation of this system for defining colours is that the visual perception of colour differences is not necessarily related linearly to differences measured within the colour space.

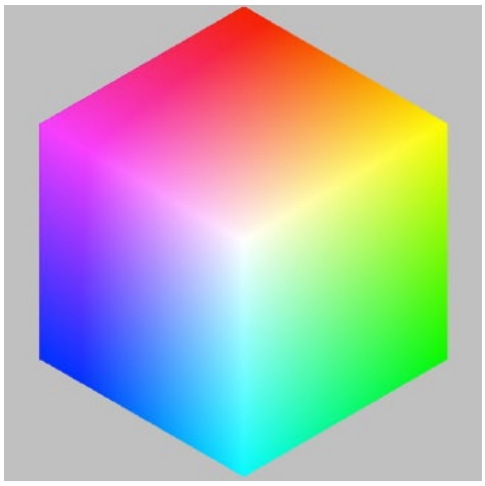
Figure 5.9 RGB colour cube

The colours formed by differing proportions of the additive primary colours, red green and blue, can be represented as a three-dimensional data space. The diagonal axis in this space, referred to as the intensity axis, represents the shades of grey from black to white that have equal proportions of each primary colour. The subtractive primaries occur at the diagonal 'corners' on each two-dimensional plane through two axes. An infinite number of planes that are orthogonal to the intensity axis can be defined. Colours on each plane have equal intensity, with saturation increasing away from the axis and hue varying around it.

a. Orientation of RGB colour cube



b. Colour planes on surface of RGB colour cube



Source: a. Harrison and Jupp (1990) Figure 11; b. Frank Clark. Retrieved from: *http://www.frank-t-clark.com/Professional/Papers/ColorHCW/ColorHCW.html)*

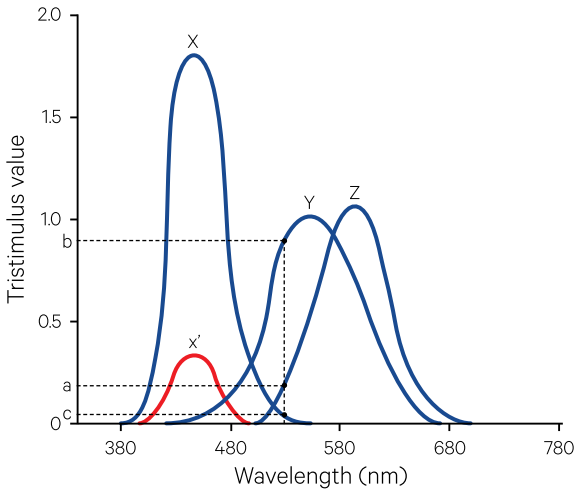
The colour coordinates defined by the RGB colour space can be expressed in terms of the perceptual attributes of hue, value (or intensity) and chroma (or saturation). The RGB axes which represent colours as Cartesian coordinates (see Volume 2B) can be mathematically rotated to cylindrical or triangular coordinates for planes of equal intensity in the cube. Since colour intensity changes along the intensity axis of the cube, as indicated in Figure 5.9, planes which are orthogonal to this axis may be defined on which the colours have theoretically equivalent intensity, with hues changing around the axis and saturation increasing radially away from it. These projections are further derived in Appendix 3.

Colour composite images are sometimes converted to the HSI colour space to manipulate the colour attributes individually. This conversion involves coding the hue, saturation and intensity coordinates as three image channels. The transformations for converting between RGB and HSI coordinates are detailed in Appendix 3. The HSI colour space is used as a basis for a number of specialised colour systems such as the Munsell System (Munsell, 1929).

A more quantitative approach to colour description was developed by the Commission Internationale de l'Eclairage (CIE). This system algebraically converts the proportions of the three primary colours required to match a given colour to the proportions of three artificial or imaginary primaries. These values are referred to as the CIE tristimulus values and are labelled X, Y and Z. The artificial primaries can be expressed as three curves as shown in Figure 5.10. For example, a colour of a single wavelength of 530 nm would be represented by the values a in X, b in Y and c in Z. Colours can be matched directly with a mixture of 1, 2 or 3 primaries, or by first adding with one or more primaries and then matching with the remaining primaries. The x' curve in the graph represents that red light needs to be added to the single wavelength colours between 400 nm and 500 nm to match the colours produced by mixing the corresponding levels of X and Y.

Figure 5.10 CIE imaginary primaries

A single wavelength colour (such as 530 nm) can be expressed in terms of its X, Y and Z component values (for example, values a, b and c for the wavelength 530 nm). The x' curve indicates the amount of red light which needs to be added to single wavelength colours between 400 nm and 500 nm to match the colours which are produced by mixing the corresponding levels of X and Y.



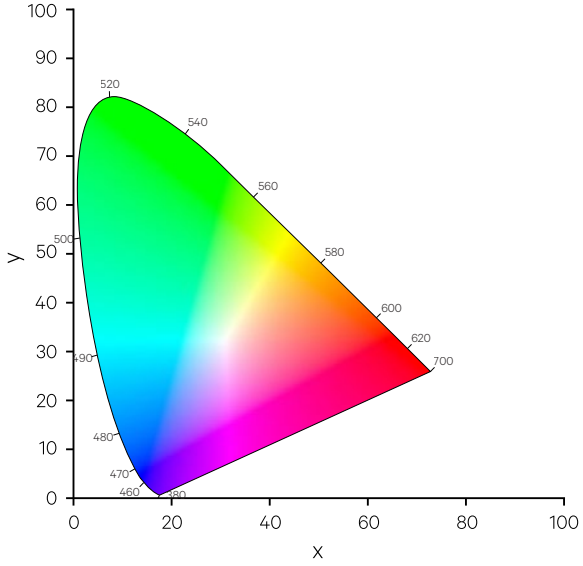
Source: Harrison and Jupp (1990) Figure 12 [Derived from CIE 1931 Standard Colorimetric Observer: CIE 1986]

This system can be used to produce a colour space that contains all perceptible colours. These are usually plotted in a normalised three-dimensional space in which two dimensions (x and y) define colour (or chromaticity) and the third defines luminance (see Appendix 3 for details). A two-dimensional plot of the colour dimensions is referred to as a chromaticity diagram or triangle, which represents changes in hue and saturation for a given intensity level. This plot forms the characteristic shape shown in Figure 5.11.

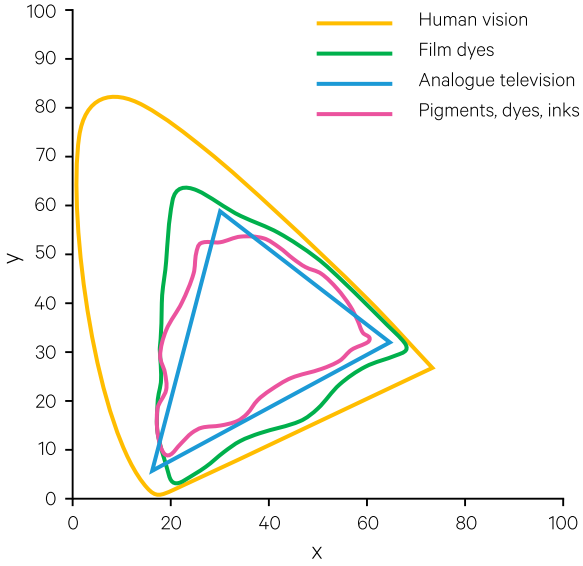
Colour is my day-long obsession, joy and torment (Claude Monet)

Figure 5.11 CIE chromaticity diagram

a. This two-dimensional representation of an intensity plane shows the ranges of colour hues that are visible to the human eye. The central white point represents the achromatic point at a given intensity level (see Appendix 3).



b. The ranges of visible colour hues are compared with those that can be reproduced by different media.



Source: a. I, Sakurambo. Retrieved from: *https://commons.wikimedia.org/wiki/File:CIExy1931.svg*. b. Harrison and Jupp (1990) Figure 13 [Adapted from Baldwin (1984) and Foley and van Dam (1984)]

This colour space is used to quantitatively measure colour differences but does not represent perceived colour differences linearly as they are not related linearly to the physical measurement of brightness. The interior and boundary of the curved region represents the visible range of colours. The pure colours form the boundary of this shape. The colour of the standard light source or illuminant, which is intended to represent sunlight, is located at the achromatic point, or white, where x, y and z have equal intensities of 1/3. (This point is defined in colorimetry as the colour of a blackbody radiator at 6504K; see Volume 1A—Section 2.11.) Again, saturation can be represented as the distance away from the achromatic point and hue is the angle of this direction.

The CIE colour space allows us to determine the purity of colours, that is, the distance of a colour from the standard white light source or illuminant (which varies along a locus with temperature) relative to its distance from the dominant wavelength at the edge of the visible region. Complementary colours can also be determined as those that lie on opposite sides of a straight line through the achromatic point. The diagonal is also used to define colour gamuts. A mixture of any two colours can produce the colours represented by a straight line between them. A mixture of any three colours produces the gamut, or range, of colours contained in the triangle bounded by the connecting lines between the colours.

The CIE chromaticity coordinates can also be converted to an approximately uniform colour space using dimensions of lightness (L*), redness-greenness (u*) and blueness-yellowness (v*) as defined in Appendix 3. The perceptual difference between two colour stimuli can then be approximated as the Euclidean distance between the points representing them in this uniform space. Other specific colour systems derived from these general models are described in Appendix 3.

5.1.3 Colour reproduction

Section 5.1.2 discussed a number of systems that have been devised to describe the range of humanly perceptible colours. However, the various display or hardcopy devices that are available for image processing can only represent a subset of these colours. The RGB technologies used to achieve a mixture of colours in different display screens are chosen to produce the widest possible range, or gamut, of colours (see Section 6).

Digital display devices typically span a greater range of colours than those achievable using colour films or inks (Baldwin, 1984). However, the use of the x' primary in colour matching (see Figure 5.10) indicates that there are some colours which cannot be readily created by adding lights. Colours outside the gamut triangle could only be produced by adding a 'negative' component of one of the primaries (see Appendix 3) so are represented by a colour on the edge of the triangle. For analogue colour television, for example, this means that the high purity blue and green colours are most restricted.

In display devices, brightness and contrast controls, as well as basic chromaticity differences between the RGB primaries used in display technologies, affect both the range of colours that can be represented and the degree of visual discrimination between them. The gamut of colours produced by hardcopy devices is affected by properties of the inks, tones and hardcopy media (paper or film) that are used.

The small size of the print or film gamut relative to display monitors means that the colours of displayed imagery cannot be fully reproduced in hardcopy. If an exact reproduction is required, a reduced gamut of colours must be used on the display device.

5.2 Colour enhancement

5.2.1 Pseudo-colouring

For a single channel of data, pseudo-colouring techniques can be used to display the image values as different colours rather than different grey shades (see Figure 4.10f). These techniques use lookup tables (LUT) to assign a colour to each image value. This operation is illustrated in Figure 5.12 for the image shown in Figure 4.1.

Figure 5.12 Pseudo-colouring of image data

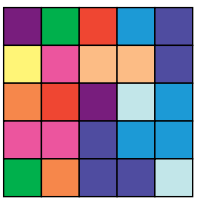
Spatial data can be represented pictorially by associating each data value with a particular colour. The selected colours may not be related to the data values but allow better visual discrimination of spatial patterns in the data than grey shade display techniques (see Figure 4.1).

a. Image pixel values

1	5	8	3	2
6	9	0	0	2
7	8	1	4	3
9	9	2	3	3
5	7	2	2	4

b. Grey scale lookup table

c. Grey scale image



Source: Harrison and Jupp (1990) Figure 14

The pseudo-colour table may be defined so that each image value has a discrete colour (see Figure 4.10f) or blocks of image values can be assigned the same colour using density slicing (see Section 9.2.1). This enhancement technique allows small changes in image values to be clearly highlighted as different colours. The pseudo-colour enhancement is most effective if applied to image data that has been contrast-stretched so that the colour differences apply over the bulk of the image data. Pseudo-colouring is also useful for interpretation and presentation of transformation results where these are recorded in a single image channel (see Volume 2C). Suggestions for selecting colours are provided in Excursus 5.2 and Excursus 9.2.

Excursus 5.2—Colour Composition

The selection of distinctive and effective compositions of colours involves an understanding of colour interactions in terms of contrast, harmony and rhythm. Both the quantity and surroundings of a colour will change its perceived hue, value and chroma.

Colour contrast can be defined in terms of perceived temperature with red, orange and yellow being considered warm hues and blues and greens as cool. The warm colours are generally more apparent in an image, especially against a neutral (grey) background, whereas cool colours are less 'active'. When colour is being used to represent a physical scale, such as temperature or yield, it should be remembered that dark colours are instinctively associated with low values and bright colours are often assumed to have high values. Strong, controlled variations in value and chroma, such as topographic shading, are usually interpreted as spatial differences in an image.

Harmony describes the sense of 'relatedness' among colours. A sequence of colours that has been selected from a natural progression in a colour model, such as a constant hue with changing value, renders an immediately perceived relationship between the image features portrayed in these colours. Traditionally, multiple hue harmonies have been established by selecting adjacent or close colours on a hue circle, or by pairing opposite or complementary colours (Greenberg *et al.*, 1982).

The final visual rhythm of an image will depend not only on the colours that have been selected but also on the relative amounts of each colour and their spatial arrangement. In image processing, this final balancing is only achievable by visual assessment of a 'painted' image.

Selection of colours should also consider our perceptive sensitivities as briefly mentioned in Appendix 3.1. Blue hues are generally not well perceived especially for thin features. The perceived colour hue of a 'patch' changes with its background and extent, with warm colours generally being too dominant for large patches. Blue on a black background, or its complementary mixture of yellow on a white background, are difficult to distinguish. Background colours that are mixtures of several spectral hues (such as brown) tend to mask fine detail. Simultaneous contrast, that is, the occurrence of one colour on two different and 'opposite' colour backgrounds, typically occurs in thematic maps and image classifications. However, such contrasting combinations reduce the discrimination between different hues so that nearly identical colours can be misread due to the immediate colour contexts in which they appear.

Colour value is a very useful attribute for distinguishing between colours but is easily forgotten so is not well suited for colours keys and is best limited to about five steps. The difference between measured and perceived intensity or value is most relevant when varying the value of a colour hue, with changes in low intensity values being less obvious than those at high intensities (see Excursus 5.1). A sub-conscious association is usually made between the perceived brightness difference of colours and the relationship between the values they represent. The value of background colours should not be too high as these will tend to draw attention away from the feature colours. Different hues also appear to have different intensities even in pure colours; for example, yellow always appears less saturated than other hues (see Figure 5.1).

The perceived chroma of a colour tends to increase with patch size so that selecting a colour from a small sample area may be misleading, especially for background colours. This can also lead to confusion between two similar hues or two values of the same hue that have important differences of chroma, since large areas of a lower chroma colour may appear similar or identical to small areas of the higher chroma colour.

5.2.2 Colour composite imagery

Multi-channel image data provide another opportunity for colour image display. In the same way that the human eye perceives colours as the proportions of the blue, green and red wavelengths it detects (see Excursus 5.1), colour monitors use blue, green and red colour primaries to create a full gamut, or range, of colours on a display screen. A colour composite image is formed by associating three different image channels with these three colour primaries: one channel is displayed as a blue scale, another as a green scale and a third as a red scale (see Figure 5.13). The display device then effectively registers these three colour scales as though they were three transparencies being viewed in front of a uniform light source. The resulting composite colours of a pixel depend on its shades of red, green and blue in the individual channels. Usually each channel of a colour composite image is individually contrast-enhanced to improve contrast in the displayed image.

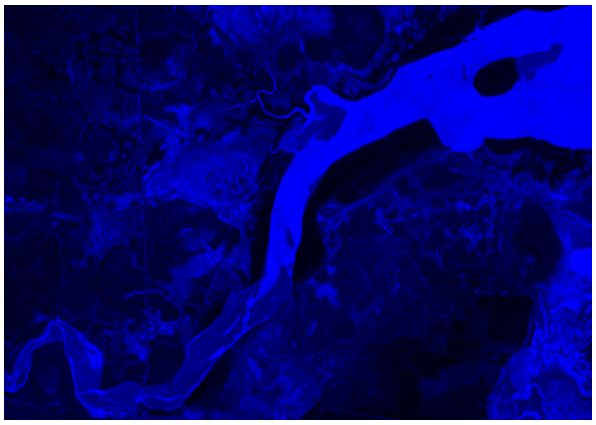
Figure 5.13 Formation of a colour composite image

Three image channels can be viewed simultaneously by associating each channel with a different primary colour in a colour composite image. Additive colour devices, such as colour monitors, use the primary colours blue, green and red while subtractive colour devices, such as plotters and photographic media, combine their inverse primaries yellow, magenta and cyan. This example uses channels 3 (visible green), 8 (NIR) and 11 (SWIR) from a Sentinel-2 image over the coastal town of St Lawrence, in central Queensland, that was acquired on 28 August 2015.

a. In the additive colour system (used by devices such as colour monitors), one channel is displayed as a blue scale, a second channel is shown as a green scale, while the third is displayed as a red scale (see second column below). These primary colours are effectively treated as registered transparencies over a light source so add together to form the lighter composite colours in Figure 5.13c below.

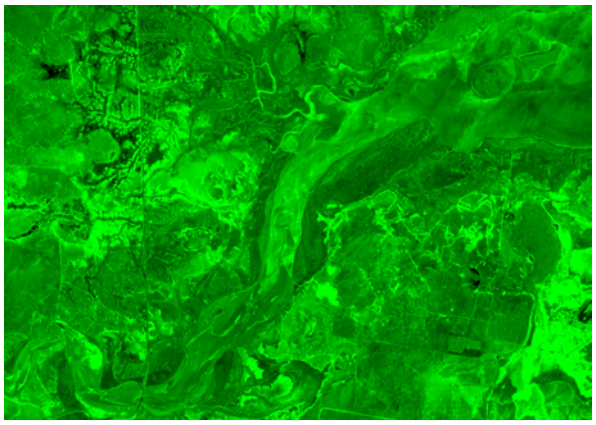
Band 3 grey scale

Represented as blue

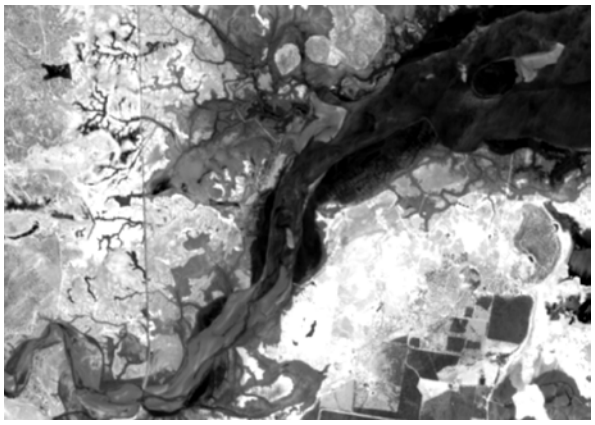


Band 8 grey scale

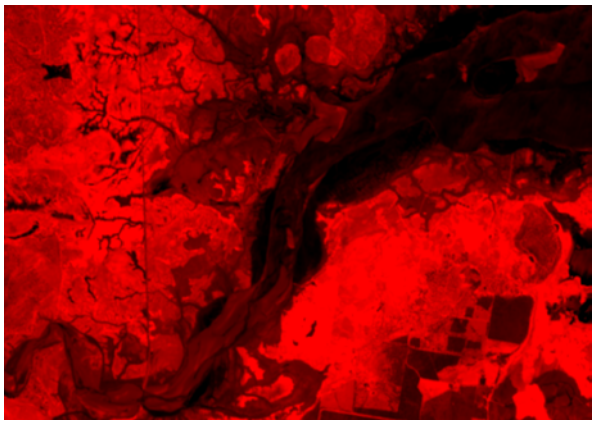
Represented as green



Band 11 grey scale



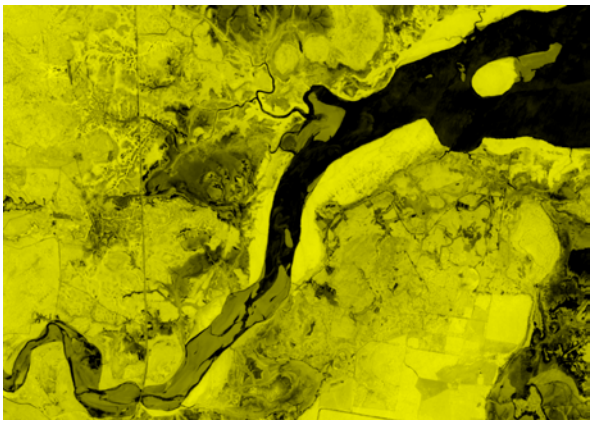
Represented as red



b. In the subtractive colour system (used by plotters and photographic media), the equivalent colour composite is formed from a yellow scale (inverse of the 'blue' channel), a magenta scale (inverse of the 'green' channel) and cyan scale (inverse of the 'red' channel) as shown in second column below. The inverse of a channel is commonly indicated as the negative channel number.

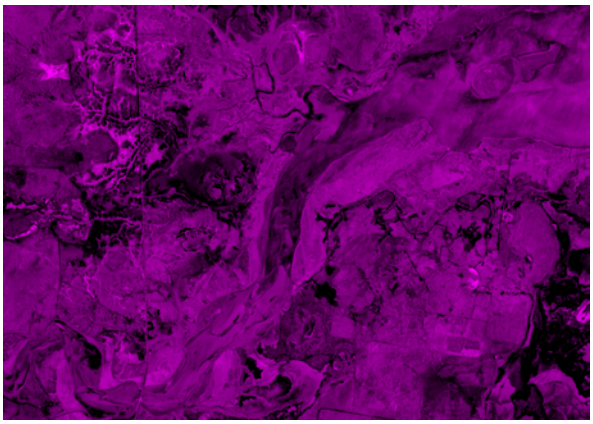
Inverse of band 3 as greyscale

Represented as yellow

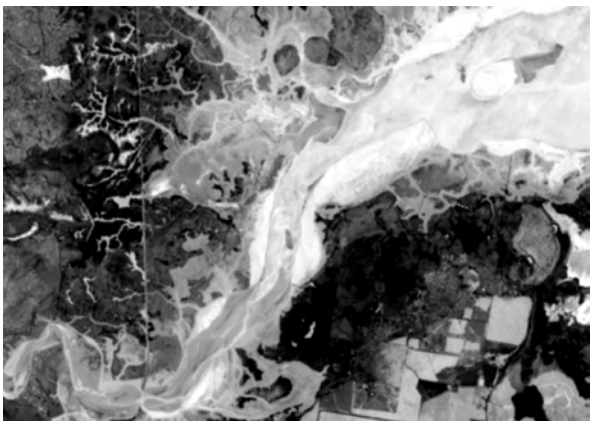


Inverse of band 8 as greyscale

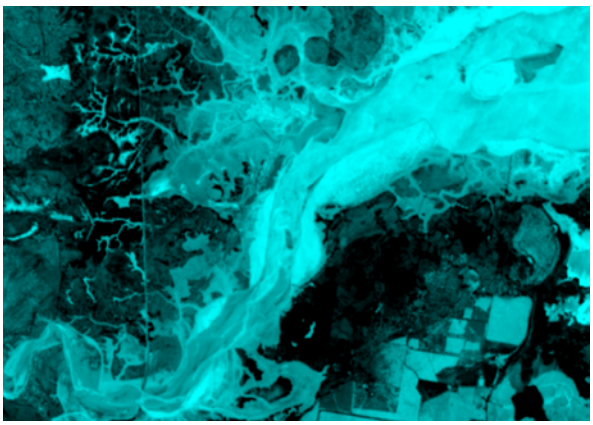
Represented as magenta



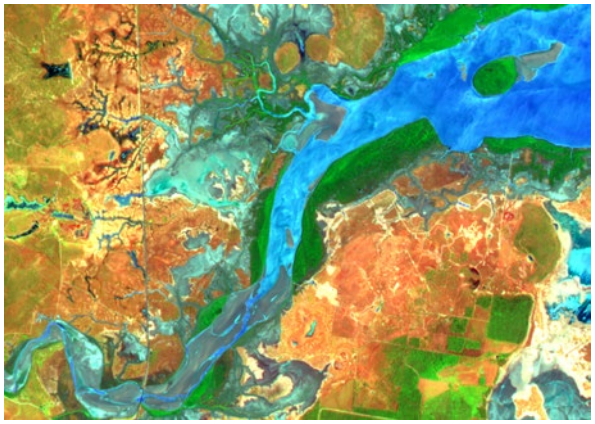
Inverse of band 11 as greyscale



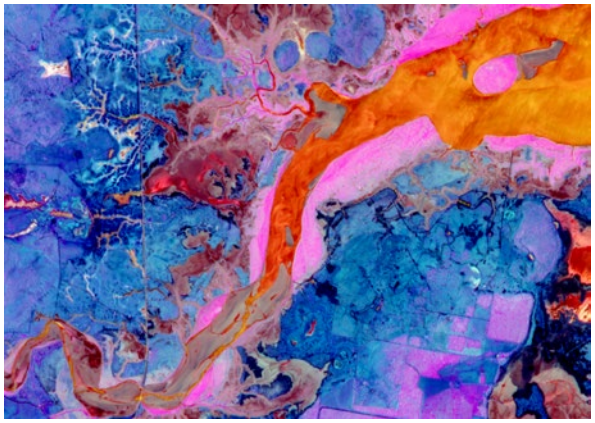
Represented as cyan



c. Colour composite formed as described in Figure 5.13a and Figure 5.13b.



d. An inverse colour composite is formed In the additive system by displaying the inverse of each channel (shown in first column of Figure 5.13b above) to the blue, green and red primaries. In the subtractive system, the same result is achieved by mapping the intensities of each channel (shown in the first column of Figure 5.13a) directly to the subtractive primaries: yellow, magenta and cyan.



Source: Norman Mueller, Geoscience Australia

A 'true', 'real' or 'natural' colour composite is formed by simultaneously displaying:

- a blue data channel (that is, recorded reflectances were in the wavelengths of blue light) as a 'blue scale' (with shades of blue, instead of grey, representing the channel values);
- a green data channel as a green scale; and
- a red channel as a red scale (see Figure 5.14d).

This should resemble the actual colours of the scanned object. With EO data, the colours of a 'true' colour composite image would be similar to those of the Earth's surface viewed from a high altitude aircraft (see Figure 5.15).

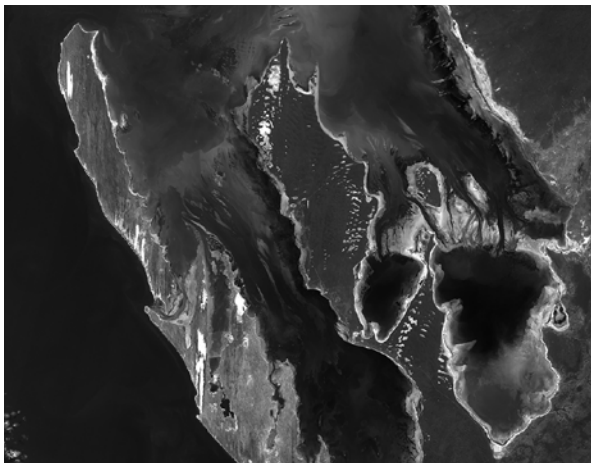
In nature, light creates the colour.
In the picture, colour creates the light.
(Hans Hofmann)

Figure 5.14 Shark Bay colour composite example

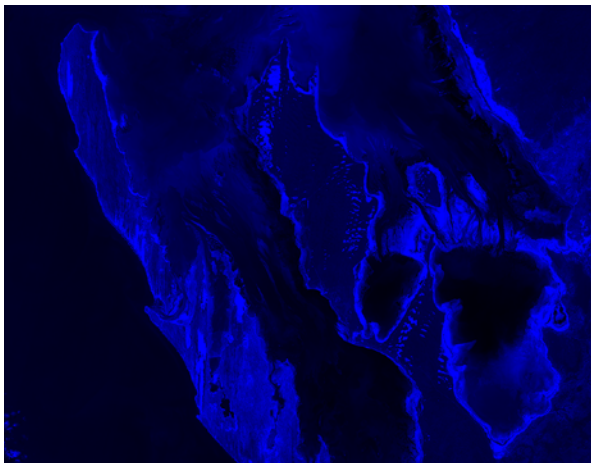
Three image channels can be viewed simultaneously on an additive colour device by associating each channel with a different additive primary colour (red, green and blue) in a colour composite image. These primary colours (a, b and c) are effectively treated as registered transparencies over a light source, so add together to form lighter composite colours (see Figure 5.14d overpage).

This example shows selected visible channels from a Landsat-8 image over Shark Bay, WA, acquired on 3 August 2017. Since blue wavelengths reflect from greater water depths than green or red wavelengths, most water features are coloured blue. Green wavelengths are also reflected from shallow water resulting in aqua colours for water adjacent to land. The red soils of this region have the highest reflectance in the red band. Cloud in the lower left corner of the image is bright in all visible channels so combines to appear white. These bands combine in a natural colour composite to produce feature colours that are comparable to an aerial photograph (see Figure 5.15 overpage).

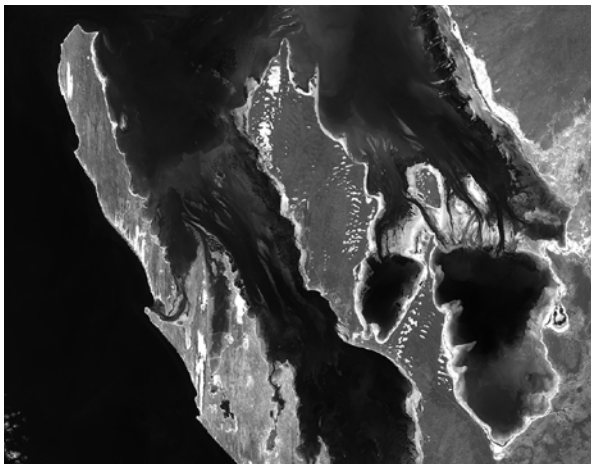
a. Greyscale of band 2 (visible blue)



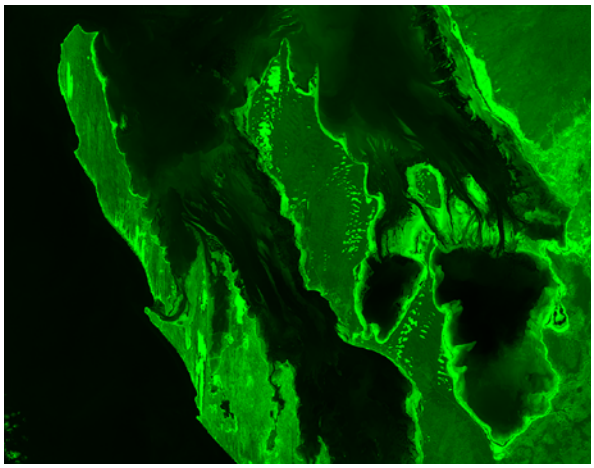
Represented as blue



b. Greyscale of band 3 (visible green)

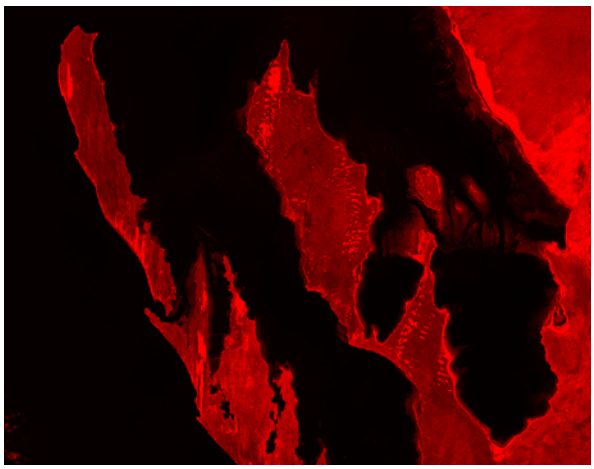


Represented as green

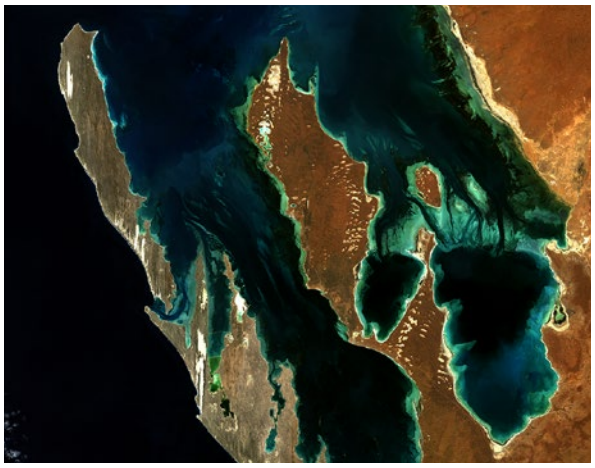


c. Greyscale of band 4 (visible red)

Represented as red



d. Colour composite of bands 2, 3 and 4 as RGB respectively



Source: Norman Mueller, Geoscience Australia

Figure 5.15 Aerial photograph of Shark Bay

This oblique aerial photograph over Shark Bay shows the true colours of ground features.

Source: Jason Mazur ©Mirage Digital. Used with permission.

Any other mapping of channels to the blue, green and red primaries is called a false colour composite since features in the resultant colour image do not have their true colours. Since visible bands tend to be highly correlated in EO imagery, including one or more non-visible bands in a colour composite increases the information content of the displayed image. This increased information content improves contrast between image features and thus enables better visual discrimination of those features.

Many EO sensors record multiple bands in visible and infrared wavelengths. For example, eight (of the 11) bands acquired by Landsat-8 are shown in Figure 5.16 over a large ephemeral lake, Lake Carnegie, in the arid centre of Western Australia. While some bands are more similar than others, each provides different information about the ground features. Different combinations of these bands can be used to create a range of colour composites (see Figure 5.17). When selected with knowledge of the reflectance characteristics of particular ground features, specific band combinations highlight similarities or differences between features of interest.

A standard false colour composite image for EO data displays a green data channel as blue, a red data channel as green, and a near infrared data channel as red as illustrated in Figure 5.17b. This combination of data channels produces a colour composite with similar colour representation to a colour infrared photograph with a characteristic set of colours for different land cover features:

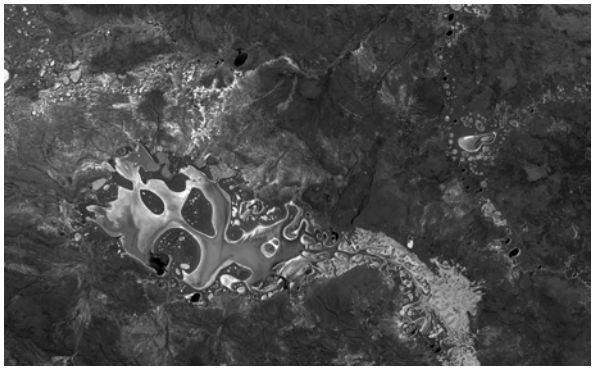
- vegetation appears red due to high reflectance in a near infrared reflectance channel, which is displayed as a red-scale;
- exposed red-brown soils appear green since they have highest reflectance in the red channel, which is displayed as a green-scale;
- water appears blue since its highest reflectance is in the green channel, which is displayed as a blue-scale; and
- bare areas appear white, since they have high reflectance in all of these channels.

Another informative colour composite displays SWIR as red, near infrared as green and visible green or visible blue as blue; see Figure 5.17c). For comparison, a colour composite of Lake Carnegie using Sentinel-2 image bands 11 (SWIR), 5 (red edge) and 3 (green) as RGB is shown in Figure 5.18. With imagery containing thermal data, an attractive and useful colour composite is formed by assigning TIR to red, NIR to green, and a visible channel to blue (see Figure 5.17d).

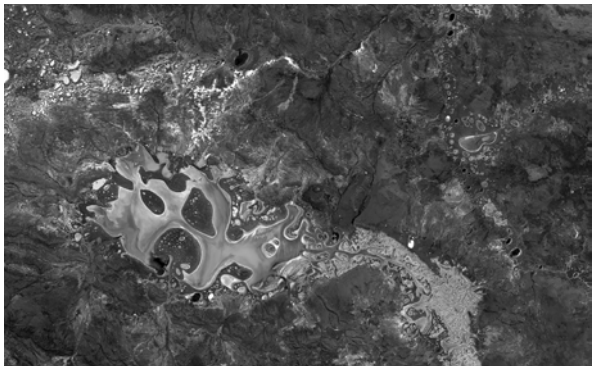
Figure 5.16 Example image channels for colour composites

Individual channels from Landsat-8 image over Lake Carnegie, WA, acquired on 8 July 2017.

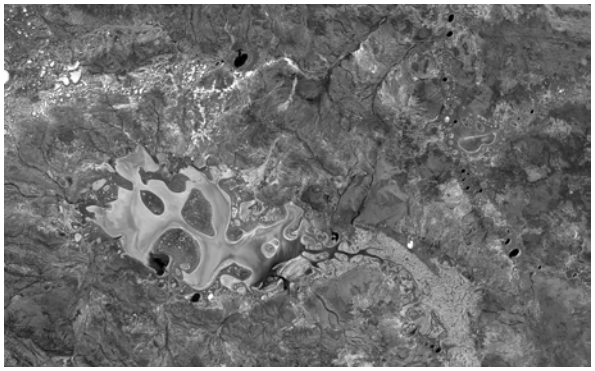
a. Blue channel (band 2: 0.452–0.512 μm)



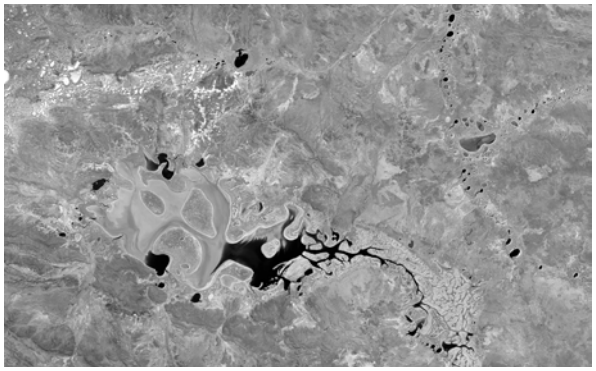
b. Green channel (band 3: 0.533–0.590 μm)



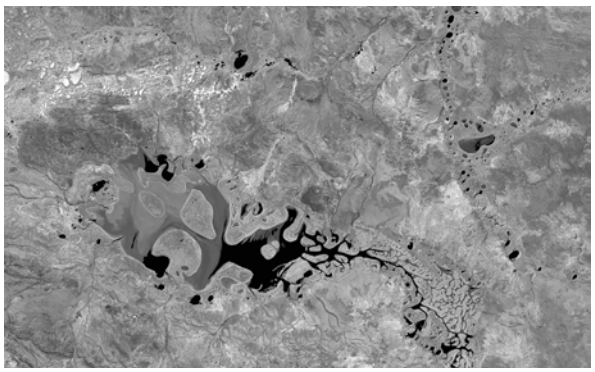
c. Red channel (band 4: 0.636–0.673 μm)



d. NIR channel (band 5: 0.851–0.879 μm)

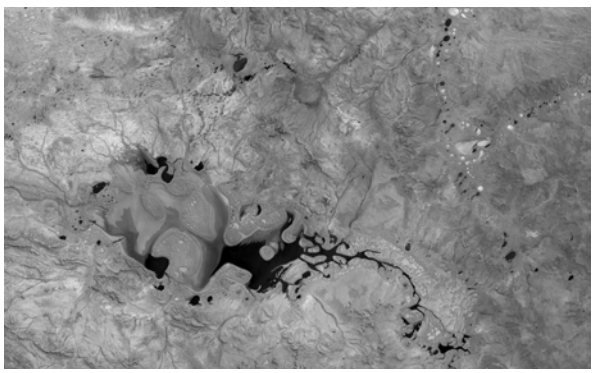


e. SWIR1 channel (band 6: 1.566–1.651 μm)

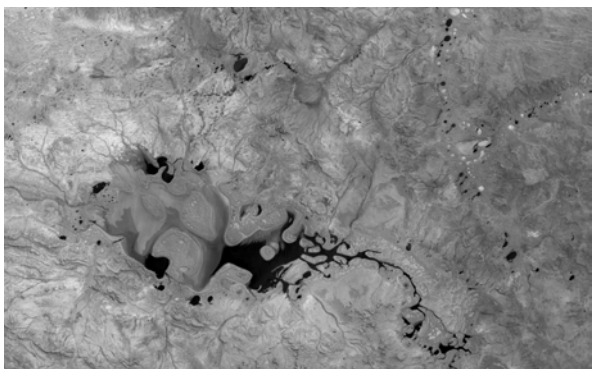


f. SWIR2 channel (band 7: 2.107–2.294 μm)

g. TIR1 channel (band 10: 10.60–11.19 μm)



h. TIR2 channel (band 11: 11.5–12.51 μm)



Source: Norman Mueller, Geoscience Australia

Figure 5.17 Standard colour composite images

These colour composite images were formed from selected triplets of the Landsat-8 channels shown in Figure 5.16.

a. A natural colour composite has colours that are similar to the colours of the original image scene since:

- a channel of red radiance values (OLI band 4) displays as red;
- a channel of green radiance values (OLI band 3) displays as a green scale; and
- a channel of blue radiance values (OLI band 2) displays as a blue scale.

As the name implies, natural colour composites represent the true colours of ground features. In this image, vegetation is sparse so land areas are dominated by the red soil background. Water in the salt lake varies in colour depending on sediment levels, with muddy water appearing green or brown.

b. A standard false colour composite for EO data associates:

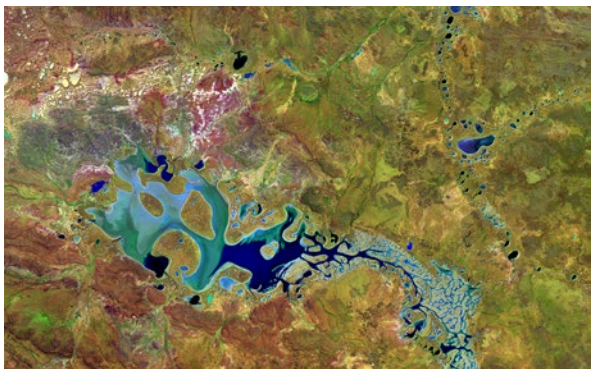
- a NIR channel (OLI band 5) with a red scale;
- a red channel (OLI band 4) with a green scale; and
- a green channel (OLI band 3) with a blue scale.

The colours of features in the resulting colour composite are similar to those of a colour infrared photograph. For example, vegetation appears as shades of red, deep clear water appears dark blue, shallow or muddy water is light blue and bare red soil appears green.

c. A SWIR false colour composites can be formed from a wide range of channel/colour combinations. In this example:

- SWIR 1 channel (OLI band 6) is displayed as a red scale;
- NIR channel (OLI band 5) is displayed as a green scale; and
- visible green channel (OLI band 3) is displayed as a blue scale.

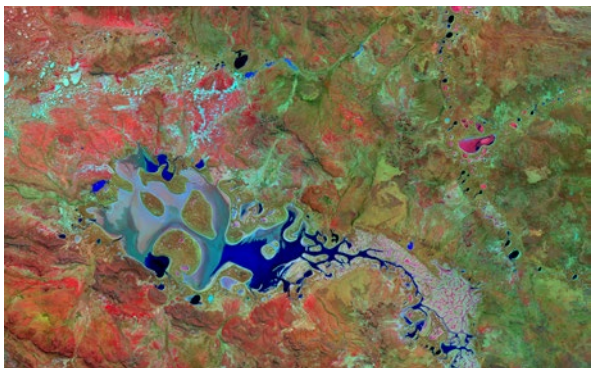
This colour composite renders 'intuitive' colours for ground features, with barer soils showing as red (due to higher SWIR reflectance), vegetation as green (due to high NIR reflectance) and muddy water as blue (due to high green reflectance). Compared with the natural colour composite (see Figure 5.17a above), this image highlights the differences in vegetation to the north and south of the lake.



d. In this thermal false colour composite:

- TIR 1 channel (OLI band 10) is displayed as a red scale;
- NIR channel (OLI band 5) is displayed as a green scale; and
- visible green channel (OLI band 3) is displayed as a blue scale.

Since this colour composite displays a thermal channel as red, hot bare, land areas are clearly visible as red tones and 'hot' lakes appear pink. Vegetation shows as green (due to high NIR reflectance) and most lakes appear as shades of blue (due to high green reflectance).



Source: Norman Mueller, Geoscience Australia

Figure 5.18 Thermal colour composite using Sentinel-2 imagery

This Sentinel-2 image over Lake Carnegie, WA, was acquired on 12 March 2017 and is displayed using bands 11, 5, 3 as RGB. These bands render a thermal colour composite that is comparable to Figure 5.17d.

Source: Norman Mueller, Geoscience Australia

Different combinations of bands can be used to enhance certain features by forcing them to appear as a more prominent colour. Non-standard colour composites may be formed by displaying different combinations of image channels as different primary colours, as illustrated in Figure 5.19. These composites can be valuable to highlight particular features or emphasise the differences between a pair of similar image bands. Since the various Landsat sensors have differing channel numbers, a summary of equivalent bands is included for reference in Excursus 5.3.

A number of statistical methods for selecting the most appropriate image bands for colour composites on the basis of their correlation have been proposed. For example, Sheffield (1985) suggested an algorithm based on image covariance statistics for selecting band combinations for colour composites from imagery with more than three channels. Another approach, the 'Optimum Index Factor' (OIF; Chavez *et al.*, 1982), selects the triplet of bands with the maximum sum of standard deviations and the minimum correlation between pairs of bands (see Section 8.1). Other methods for selecting sets of bands with minimum correlation are discussed in Volume 2C (in the context of image transformations) and Volume 2E (in the context of image classification).

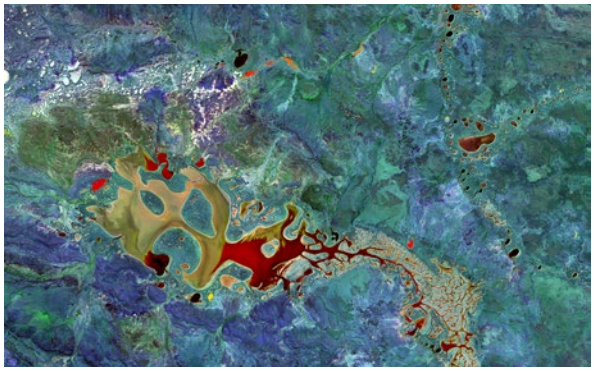
Figure 5.19 Non-standard colour composite images

These unusual colour composite images were formed from selected triplets of the Landsat-8 channels shown in Figure 5.16.

a. In this non-standard colour composite:

- green channel (OLI band 3) is displayed as a red scale;
- NIR channel (OLI band 5) is displayed as a green scale; and
- SWIR 1 channel (OLI band 6) is displayed as a blue scale.

This composite image uses the same triplet of bands as the SWIR false colour composite shown in Figure 5.17c, but displays the bands as non-intuitive colours so that vegetation appears as blue shades and water appears as red shades. Contrasting colours can enhance discrimination of some features.



b. In this non-standard colour composite:

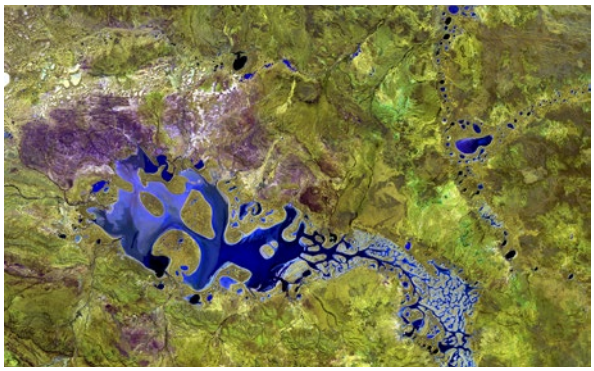
- NIR channel (OLI band 5) is displayed as a red scale;
- SWIR 1 channel (OLI band 6) is displayed as a green scale; and
- a blue channel (OLI band 2) is displayed as a blue scale.

This combination of bands can be used to highlight land versus water features.

c. In this non-standard colour composite:

- SWIR 2 channel (OLI band 7) is displayed as a red scale;
- SWIR 1 channel (OLI band 6) is displayed as a green scale; and
- blue (OLI band 2) is displayed as a blue scale.

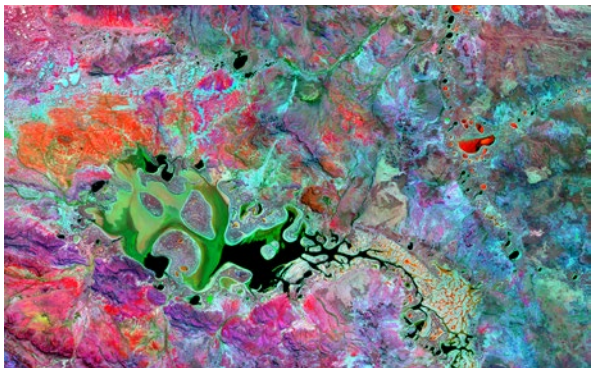
By using both SWIR bands, this composite highlights the differences in their information content. Since the SWIR bands are displayed as red and green, a range of yellow hues show changes in SWIR reflectances, possibly due to changes in mineralogy. Water clearly contrasts with land by appearing as blue tones.



d. In this non-standard colour composite:

- TIR 1 channel (OLI band 10) is displayed as a red scale;
- NIR channel (OLI band 5) is displayed as a green scale; and
- SWIR 1 (OLI band 6) is displayed as a blue scale.

This vibrant combination of bands differentiates between ground features on the basis of their reflectance properties in different infrared wavelengths.



Source: Norman Mueller, Geoscience Australia

Excursus 5.3—Landsat band numbers

Source: Norman Mueller, Geoscience Australia

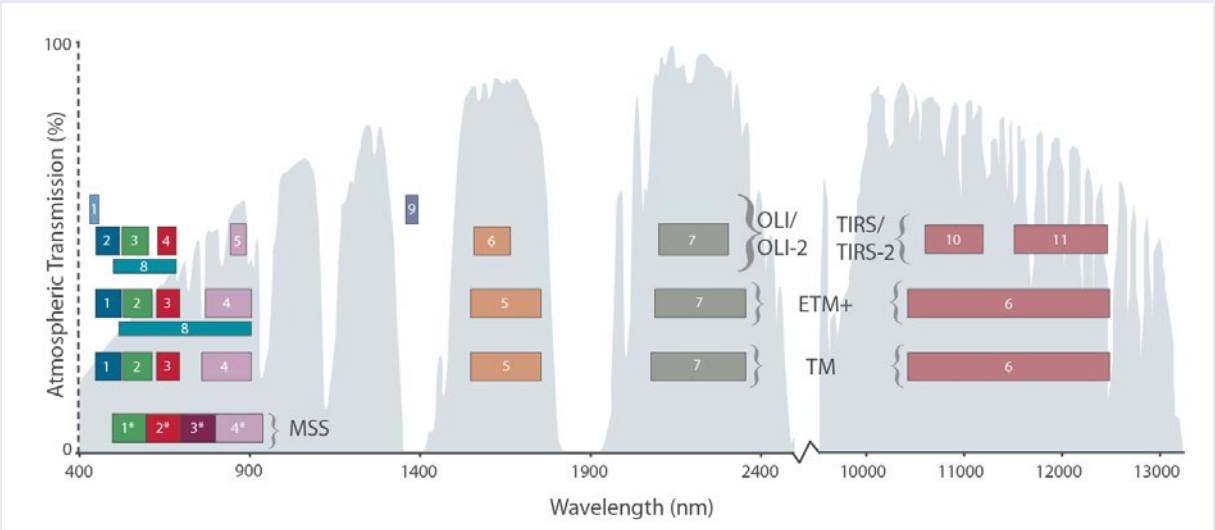
The Landsat-8 OLI sensor has additional spectral bands compared to Landsat-4/5 TM and Landsat-7 ETM+. A coastal blue band was added as Band 1 for Landsat-8 OLI, resulting in a change to the band numbering order. The spectral bands used in various Landsat sensors are shown in Figure 5.20.

Table 5.1 summarises the equivalence of spectral band numbers in these sensors. This summary should be useful for producing comparable colour composites from imagery acquired by different Landsat sensors.

Table 5.1 Equivalent Landsat band numbers

Spectral band	Sensor		
	Landsat-4/5 TM	Landsat-7 ETM+	Landsat-8 OLI
Coastal blue	-	-	1
Visible blue	1	1	2
Viaible green	2	2	3
Visible red	3	3	4
Near infrared	4	4	5
Shortwave infrared 1	5	5	6
Shortwave infrared 2	7	7	7
Thermal infrared 1	6	6	10
Thermal infrared 2	-	-	11
Panchromatic	-	8	8
Cirrus (1.363–1.384 mm)	-	-	9

Figure 5.20 Spectral bands in Landsat sensors



Source: NASA. Retrieved from: *https://landsat.gsfc.nasa.gov/wp-content/uploads/2016/10/all_Landsat_bands.png*

5.3 Further Information

Colour Society of Australia: *http://www.coloursociety.org.au*

Optimum Index Factor: *http://spatial-analyst.net/ILWIS/htm/ilwisapp/optimum_index_factor_functionality_algorithm.htm*

Gonzalez and Woods (2017) Chapter 7

Hunt (1987, 2004)

Levine (1985)

5.4 References

Baldwin, L. (1984). Color Considerations. *BYTE Magazine*, 227-249.

Chavez, P. S., Berlin, G. L., and Sowers, L. B. (1982). Statistical method for selecting Landsat MSS ratios. *Journal of Applied Photographic Engineering*, 8, pp. 23–30.

Cornsweet, T. N. (1970). *Visual Perception*. Academic Press, New York.

Foley, J. D., and van Dam, A. (1984). *Fundamentals of Interactive Computer Graphics*. Addison-Wesley.

Gondalez and Woods (2017). *Digital Image Processing*. Pearson Education, Inc.

Harrison, B. A., and Jupp, D. L. B. (1989). *Introduction to Remotely Sensed Data. Part ONE of the microBRIAN Resource Manual* (156 pages). CSIRO Australia, Melbourne.

Harrison, B. A., and Jupp, D. L. B. (1990). *Introduction to Image Processing. Part TWO of the microBRIAN Resource Manual* (256 pages). CSIRO Australia, Melbourne.

Hunt, R.W.G. (1987). *Measuring Colour*. Ellis Horwood Ltd. .

Hunt R.W.G (2004). *The Reproduction of Colour*. 6th edn. John Wiley and Sons, West Sussex.

Land, E. (1977). The retinex theory of colour vision. *Scientific American*, 237(6), pp. 108-128.

Levine, M.D. (1985) *Vision in Man and Machine*. McGraw-Hill, Inc. 574 pp.

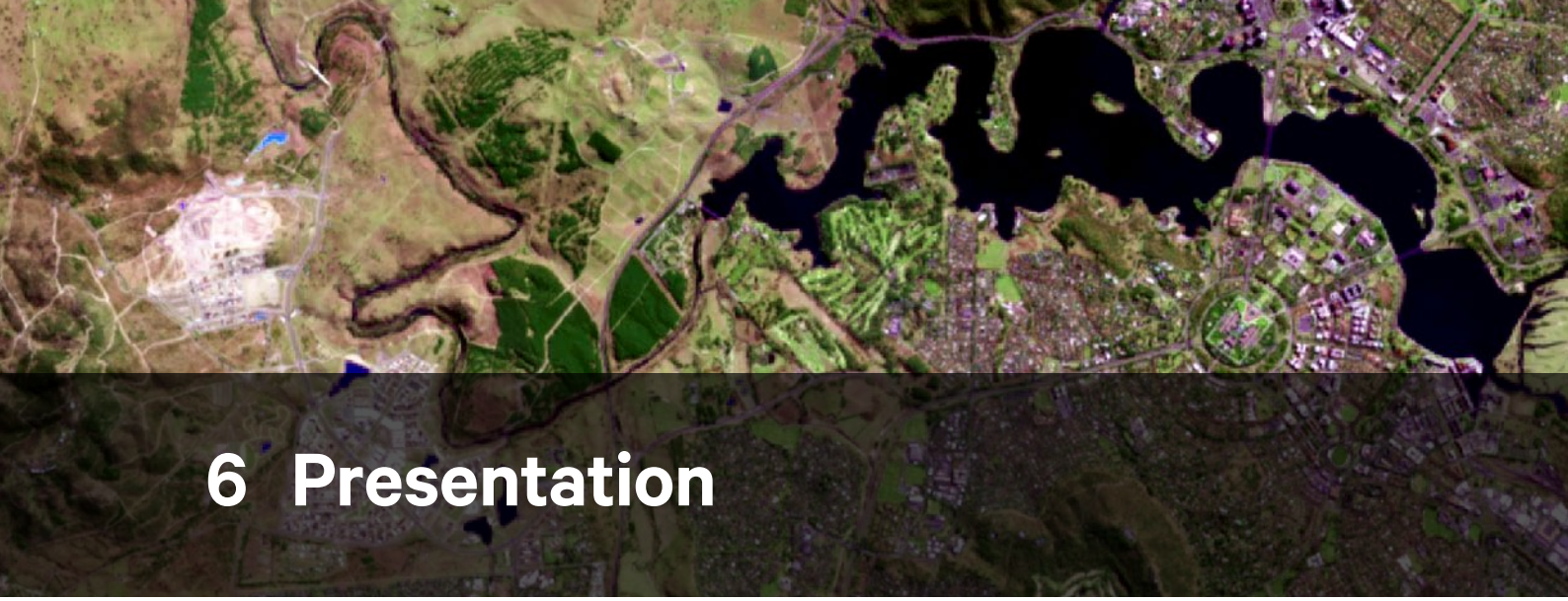
Munsell, A. H. (1929). *Munsell Book of Color*. Munsell Colour Co., Inc.

Niblack, W. (1986). *An Introduction to Digital Image Processing*. Prentice-Hall International, New Jersey.

Padgham, C. A., and Saunders, J. E. (1975). *The Perception of Light and Colour*. Academic Press, New York.

Sheffield, C. (1985). Selecting band combinations from multispectral d. *Photogrammetric Engineering and Remote Sensing*, 51(6), pp. 681–687.

Stone, M. C., Cowan, W. B., and Beatty, J. C. (1988). Color gamut mapping and the printing of digital color images. *ACM Transactions on Graphics*, 7(4), pp. 249-292. doi:*http://dx.doi.org/10.1145/46165.48045*.



6 Presentation

The techniques commonly used to represent spatial data as a picture are discussed in Sections 4 and 5, while mechanisms to alter image geometry are presented in Section 7. This Section considers options for viewing imagery, namely:

- image display (see Section 6.1); and
- image hardcopy (see Section 6.2).

6.1 Image Display

EO images are typically displayed on computer screens, either for visual interpretation, interactive enhancements and interrogations, or assessment of other processing procedures. This section introduces common features of image processing systems that are relevant to image display:

- using display hardware (see Section 6.1.1);
- displaying grey scale imagery (see Section 6.1.2); and
- displaying colour composite imagery (see Section 6.1.3).

Other commonly available display features include zooming and annotation of the displayed image.

6.1.1 Display Hardware

The technology behind computer display screens has changed significantly in recent decades from monitors using cathode ray tubes (CRT) to liquid crystal display (LCD) to light emitting diodes (LED). We will not venture into engineering aspects of these different technologies, but simply observe that this development path has resulted in display hardware with significantly higher resolution and flexibility for much lower cost.

For image display, any display screen is treated as a grid, which can be activated at discrete points to represent individual image pixels. All colour screens effectively involve activation of blue, green and red 'sub-pixels' to create colour images. These three primary colours are integrated by the eye and interpreted as a single colour (see Section 5). Specific hardware features of image display screens determine the number of colours that may be displayed on the screen at any one time.

EO imagery is generally represented on display screens in terms of four display modes:

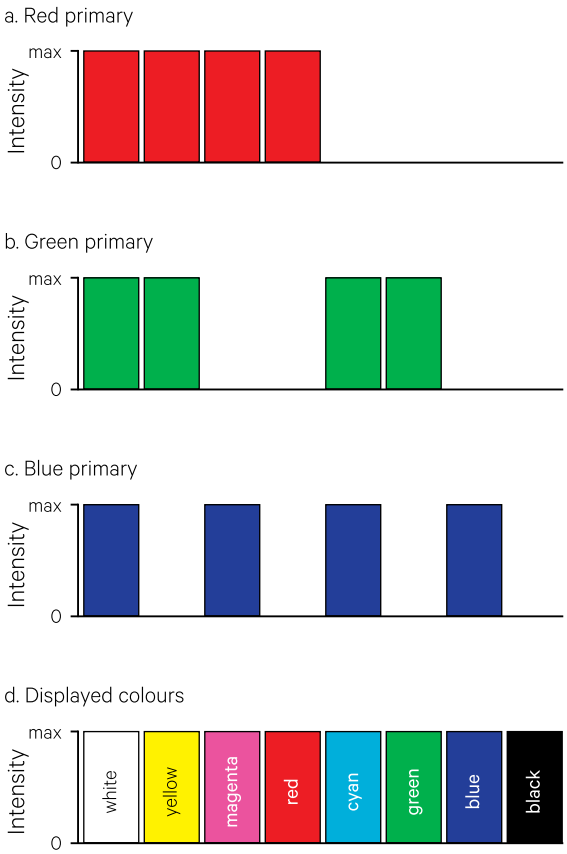
- bit-map—all pixels are presented as either black or white; this mode is used to define image masks (see Section 10);
- grey scale—represents a mapping of values in one image channel to shades of grey on the display screen (see Section 4);
- pseudo-colour—shows a grey scale image (one image channel) as range of colours rather than grey tones (see Section 5.2.1); and
- colour composite—operates by associating three image channels with the three additive primary colours (see Section 5).

Where the three channels being displayed in colour mode represent the same data values (that is, they are the same data channel), the proportions of blue, green and red at each screen pixel are identical, so add together to give a shade of grey. Some simple examples of colour mixing are illustrated in Figure 6.1.

Background image: Colour composite of Landsat-8 image of Canberra, ACT, acquired on 10 August, 2017, with bands 6, 5, 3 displayed as RGB, and sharpened using band 8 (panchromatic). **Source:** Norman Mueller, Geoscience Australia

Figure 6.1 Additive colour mixing

A range of colours can be produced on the display monitor using the additive primary colours, red green and blue. This example only shows mixing of the additive primaries to create all additive and subtractive primary colours, plus black and white.



Source: Harrison and Jupp (1990) Figure 27 [Adapted from Richards (1986)]

An intermediate display operation can occur in most image processing systems whereby each image value is assigned a different output value or colour using a colour lookup table (LUT; see Section 4.2.4 and Volume 2C). Thus, by re-defining the values in the LUT, we can interactively change the colours of the displayed image without altering the image values. Various display programs in image processing software make extensive use of this colour LUT feature for interactive image enhancements, such as pseudo-colouring (see Sections 5 and 9).

To maintain a displayed image involves continually refreshing the display screen for every pixel of every line in the display. Early CRT monitors used screen phosphors with relatively low persistence (that is, they faded easily) so needed to be refreshed (or rewritten) frequently to avoid flickering. While the standard refresh rate for CRT screens was 30 frames per second, newer screen technologies have superior persistence so require less frequent refresh cycles.

6.1.2 Grey-scale Imagery

6.1.2.1 Contrast stretching

Minimum and maximum values for contrast stretching may be determined using:

- image histogram statistics:
  - over the whole image (see Section 4); or
  - interactive training on a particular feature (see Section 9.1.1.1); or
- interactive screen enhancements (see Section 6.1.2.2).

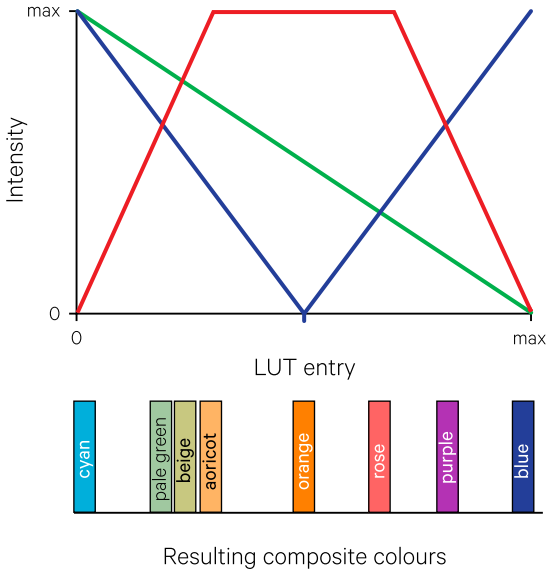
Interactive enhancements use the hardware LUT feature to modify the displayed grey shades of the image values in display memory. In some image processing systems, interactive contrast stretching also allows a non-linear stretch to be applied to the displayed image, for example via gamma function values. The interactive stretching operations are generally monitored using a grey scale bar showing current minimum and maximum values and/or a two-dimensional graph of image versus display values.

6.1.2.2 LUT manipulation

Lookup tables (LUT) allow pseudo-colouring of a grey scale image via a LUT editing facility. For example, a LUT editor could permit a 256 entry LUT to be defined graphically by specifying the intensities of the blue, green and red primaries for each entry value. An example LUT generated this way is shown in Figure 6.2.

Figure 6.2 Generating a colour LUT from RGB intensities

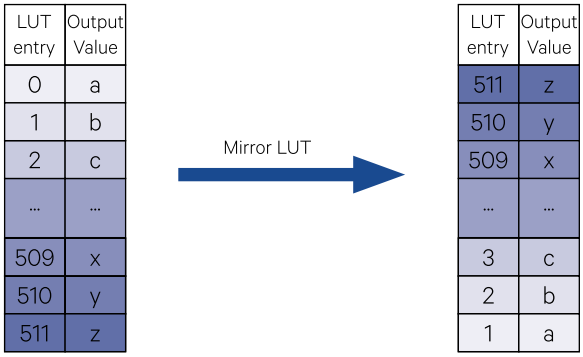
In the additive colour system, varying intensities of red, green and blue can be graphically defined for each LUT entry. These intensities are summed for each entry to produce a set of graduating colours.



Source: Harrison and Jupp (1990) Figure 30

Figure 6.3 Mirroring the LUT

The order of entries can be reversed to alter colour assignment in the displayed image. This type of operation can be applied to all or part of the LUT. Other operations such as rotation of entries can also be used to systematically alter the LUT.



Source: Harrison and Jupp (1990) Figure 31

Other manipulations which may be applied to LUTs include mirroring the full set of entries, that is, reversing the order of colour values between entries 0–511 to apply to entries in the order 511–0 as illustrated in Figure 6.3. The LUT entries may also be rotated in variable step sizes. These operations are most easily understood when applied to a standard colour table (see Section 9.2.1).

6.1.3 Colour Composites

Colour composite imagery may be formed by assigning any three image channels to the blue, green and red colour primaries (see Section 5.2.2). Most image processing systems allow the inverse of a channel (that is, showing low values as bright, and high values as dark) to be displayed (for both colour and grey scale images). The inverse of a channel is commonly indicated by specifying the negative of that channel (see Figure 5.13b in Section 5). One or two colour primaries may also be de-activated in a colour composite display (commonly by entering '0' rather than a channel number). Generally, menu-defined minimum and maximum values for appropriate channels of the image entry being displayed are automatically used unless local stretch values are active. In most image processing systems, a non-linear stretch can also be defined before displaying the image, such as by specifying an appropriate gamma function values for each channel.

Some image processing systems allow spatial data to be displayed as 'RGBI' composites, associating three image bands with the red, green and blue colour primaries and a fourth data source to be shown as intensity. This function is convenient for embedding other spatial information (such as topography) into a colour composite (see Volume 2C).

6.2 Hardcopy

Hardcopy of image data during or after processing is valuable for various stages of interpretation, field checking and final presentation. This usually involves writing the image as a colour or grey scale picture to a hardcopy device. To overlay a hardcopy image onto a map requires that the image data be appropriatedly scaled and often rectified (that is, geometrically modified) in some way before presentation (see Section 7 and Volume 2B).

The basis for colour representation in hardcopy devices is different from that used in display devices (see Sections 5 and 6.1, plus Appendix 3). Whereas a display monitor derives its full gamut, or range, of colours from the additive primaries (blue, green, and red), hardcopy devices use the subtractive primaries (yellow, magenta and cyan) to create a large range of colours as illustrated in Figure 5.13. Section 6.2.1 discusses some features of printing technology that can be used to modify the range of output colours, while Section 6.2.2 briefly introduces photographic output options.

6.2.1 Printer Output

Printers use a fine grid of discrete, coloured dots in the subtractive primaries to form imagery in much the same way display devices activate the additive primaries associated with screen pixels (see Sections 5 and 6.1. Various technologies have been developed to apply colour to a plotting medium (paper or transparent film). For example, to represent individual image pixels, ink jet printers 'squirt' fine drops of ink, whereas laser printers use laser technology to apply toner to the output media.

Most printers create their colour gamuts using consumables of the three subtractive primaries: yellow, magenta and cyan. As illustrated in Figure 5.7b and Figure 5.8b, by mixing combinations of these three colours in any printed dot we can create the eight different colours listed in Table 6.1. Due to the variable qualities of consumables and the printing mechanism used, mixing yellow, magenta and cyan often creates a dark brown colour, rather than true black, so most printers provide a black option in addition to the three primary colours. Occasionally other consumable colours are also used to enable a greater variety of output colours.

Table 6.1 Subtractive colour mixing

Ink Used in Dot			Resulting Dot Colour
Yellow	Magenta	Cyan	
off	off	off	White
on	off	off	Yellow
off	on	off	Magenta
off	off	on	Cyan
on	on	off	Red
on	off	on	Green
off	on	on	Blue
on	on	on	Black

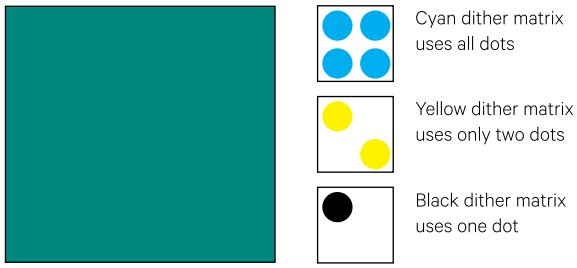
Printers are sometimes classified in terms of whether the amount of colour applied to each dot can be varied:

- binary or bilevel printers (such as inkjet printers and early laser printers) only have two levels of intensity for each primary colour—on or off—resulting in the eight colours listed in Table 6.1. To create a larger apparent range of colours, these printers use techniques such as dithering (see Figure 6.4). Dithering, however, necessarily reduces the effective resolution of the printer (see Appendix 4).
- multi-level printers (such as modern laser printers) have multiple levels of intensity for each primary colour so can directly print a greater range of colours—from 8 to 64 tones for each primary colour, depending on printer technology. These printers require less dithering, so can retain higher output resolution.
- continous tone printers (such as dye sublimation printers or film recorders) offer the largest range of intensities.

Continuous tone images (contone) comprise a seemingly infinite range of grey tones or colours and are only truly rendered by continuous tone printers. However, many laser printers now blend the applied primary colour within printer dots to create exact shades, rather than rely on the eye to integrate them. This blending of primaries within dots results in smoother colour gradations and reduced graininess. Other differences between printers relate to the variation that can be achieved in dot size, shape or placement, which primarily impact image sharpness.

Figure 6.4 Dithering

The dithering technique is implemented by grouping together blocks of dots (or nibs), then using the blended colour of each group of dots to represent a single pixel in the image. The size of a block can vary and affects both the number of colours that can be represented and the effective resolution of the printed output (see Appendix 4). In this example, a teal sample colour would be rendered using 100% cyan, 50% yellow, 25% black and 0% magenta.



Colour prints may be formed from EO imagery using pseudo-colouring or colour composites. As discussed in Section 5.2.2, to display a colour composite image using the additive primaries of blue, green and red, one image channel is represented as a blue scale, another as a green scale and a third as a red scale. To print the same image we need to map our image data to the subtractive primaries yellow, magenta and cyan. That is:

- the channel displayed as blue is printed as an inverse yellow scale (with low image values being plotted as high intensity and vice versa);
- the channel displayed as green is printed as an inverse magenta scale; and
- the channel displayed as red is printed as an inverse cyan scale.

The combinations of intensities in our additive primary colours can thus be approximated in subtractive colour media by representing their inverse intensities in the complementary subtractive primary. This inverse mapping is handled automatically in most image processing systems so colour composites only need to be specified in terms of the blue, green and red primaries for printing. To ensure differentiation of colours in a printed output, some image processing systems also allow multiple user-defined colour maps to be defined for particular printer configurations. A given image, for example, could have one LUT for printing to a laser printer and another LUT for printing to an inkjet printer.

A final consideration in printing imagery is the differences between the additive and subtractive colour spaces. The way colours are formed and perceived in these two systems is physically different so it is often difficult to print an image using exactly the same colours as appear on the display screen (see Section 5.1.3). Our perception of the brightness of an object is non-linearly related to the measured amount of light, or luminance, from it, with discrimination being greater between darker light levels (see Figure 5.3). Similarly, the lightness which is perceived as the mid-tone in a linear sequence of varying grey levels between black and white is darker than the real mid-tone (Drury, 1987). Other aspects of colour perception are introduced in Appendix 3.

The additive display system starts with a black screen and creates colours by adding coloured light to this black background (see Figure 5.13). Small changes in intensity in the dark shades are perceived more readily than in the light shades.

In subtractive colour systems, inks or dyes are added to a white background to produce various colours, that is they are 'subtracted' from white. On the white background, small additions of colour have more visual impact in the bright colours than in darker colours. Thus, while the two renditions of colour may be theoretically comparable, our perception makes them appear different. To compensate for this, a non-linear contrast stretch is usually employed when printing imagery (see Section 4.2.2).

The entertainment is in the presentation.
(John McTiernan)

6.2.2 Photographic Output

Photographic output of image data can be produced using colourwriting (or photowriting) techniques. This output can be in the form of positive or negative transparencies or prints. Film writers use a point of variable intensity to expose a small area of colour (or black and white) film. The size of the area exposed and the number of pixels and lines that can be written to a negative both depend on the selected aperture size. Image pixel intensity information is used with a film-response calibration function to determine the intensity of exposure at each point.

6.3 Further Information

Dithering: *http://webstyleguide.com/wsg2/graphics/dither.html*
https://www.tutorialspoint.com/dip/concept_of_dithering.htm

Hunt (2004)

6.4 References

Harrison, B. A., and Jupp, D. L. B. (1990). *Introduction to Image Processing. Part TWO of the microBRIAN Resource Manual* (256 pages). CSIRO Australia, Melbourne.

Hunt R.W.G (2004). *The Reproduction of Colour*. 6th edn. John Wiley and Sons, West Sussex.

Richards, J. A. (1986). *Remote Sensing Digital Image Analysis: An Introduction*. Springer-Verlag, Berlin, Germany.

7 Geometry

Image data have properties relating to both colour and geometry. EO imagery can display a variety of geometric distortions due to characteristics of the scanning mechanism used to acquire the data (see Volume 1A). These distortions are summarised in Section 3.2.2 above.

The basic concepts and simple manipulations relating to image geometry, which are relevant to a range of image processing operations, are introduced in Sections 7.1 and 7.2. Image rectification and registration may be required in remote sensing exercises to locate image features on maps, locate field sites in images, or create output imagery that are scaled to overlay a map or some other image (see Sections 7.3 and 7.4). In particular, image registration is essential for multi-temporal studies, where multiple images need to be geometrically registered for change detection analyses, or for integrating map data (such as cadastral boundaries) with an image (see Volume 2D). Mosaicking of multiple scenes for large area studies also requires accurate image rectification and registration.

The detailed methodologies required to rectify the geometry of EO images or register such imagery with other spatial data are described in Volume 2B. Registered or integrated image datasets are required for multi-date analyses (see Volume 2D.

7.1 Image Coordinate System

Figure 1.1 (see Section 1 above) represents an image as a n-dimensional table or matrix of numbers, with image lines forming rows in the table, pixels along a line forming columns in the table, and channels in the image corresponding to different data attributes or variables. Entries in the table can be identified by specifying their corresponding column and row numbers, counting from column 1 at the leftmost column and row 1 as the top row. Similarly, the location of a particular pixel in an image is traditionally recorded as the number of pixel columns or samples from the leftmost pixel and the number of lines from the top line as illustrated in Figure 7.1a.

This coordinate system differs from the traditional planar system used for two-dimensional graphs where all positive locations are above and to the right of a central origin (see Figure 7.1b). The image coordinate system effectively places the origin at the top-left corner of the image with the centre of the top-left pixel being location (1,1). Other locations are then counted to the right of, and down from, this origin.

In the context of image locations, a 'pixel' position corresponds to a column in the image matrix and a 'line' position corresponds to a row. Typically, both the size of an image, and locations within it, are given as a coordinate pair in the order '(*pixels across, lines down*)' in the same way as the planar system specifies '(*x, y*)'. The definition of the image coordinate system is further discussed in Volume 2B.

Image geometry may be altered by the simple operations of modifying image size and orientation (see Section 7.2) or the more complex procedures of rectification and registration (see Section 7.3, and Volume 2B).

Geometry is not true, it is advantageous.
(Henri Poincare)

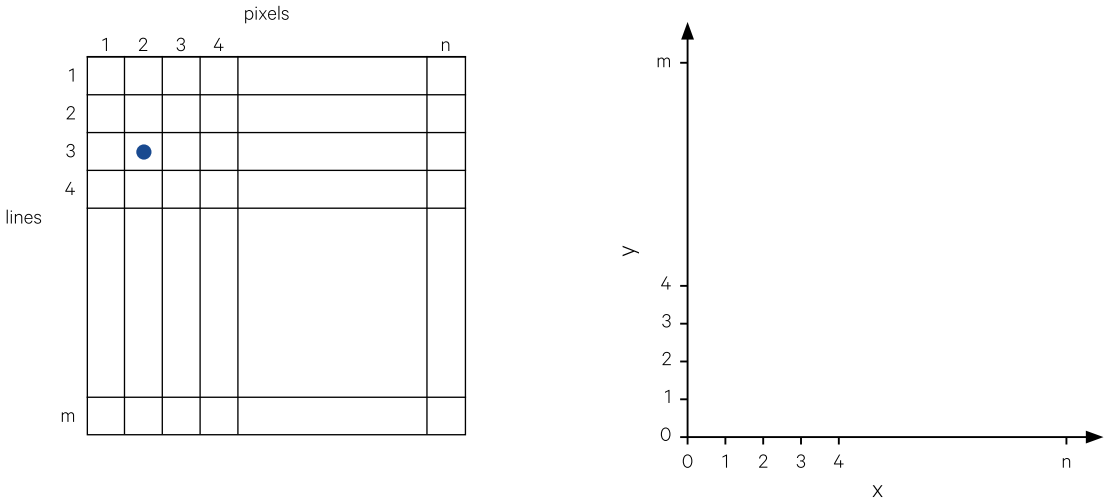
Background image: Colour composite of Landsat-8 image of Canberra, ACT, acquired on 10 August, 2017, displayed using bands 6, 5, 3 displayed as RGB, and sharpened using band 8 (panchromatic). This image has been overlaid with map information showing roads and water features.
Source: Norman Mueller, Geoscience Australia

Figure 7.1 Standard image coordinate system

A standard image coordinate system is compared with the traditional planar coordinate system.

a. Image coordinates are relative to an origin in the top left corner of the image. Location ● is in pixel 2 and line 3 of the image or '(2,3)'. The same coordinate referencing system is used for all channels in the image.

b. Most map coordinates are based on a planar system with the origin in the lower left corner of the map.



Source: Harrison and Jupp (1990) Figure 15

7.2 Shape and Orientation

7.2.1 Modifying image size

The size of image files may be altered in a number of ways:

- subsetting a large image into one or more smaller image(s) (see Section 7.2.1.1);
- joining several images into a larger image (see Section 7.2.1.2);
- blocking an image into a smaller image by:
  - averaging the values of several pixels in the input image to a single value in the output image or
  - sub-sampling the input image to represent a larger pixel (see Section 7.2.1.3); and
- enlarging the scale of an image by duplicating single pixel values in the input image to become multiple pixels in the output image (see Section 7.2.1.4).

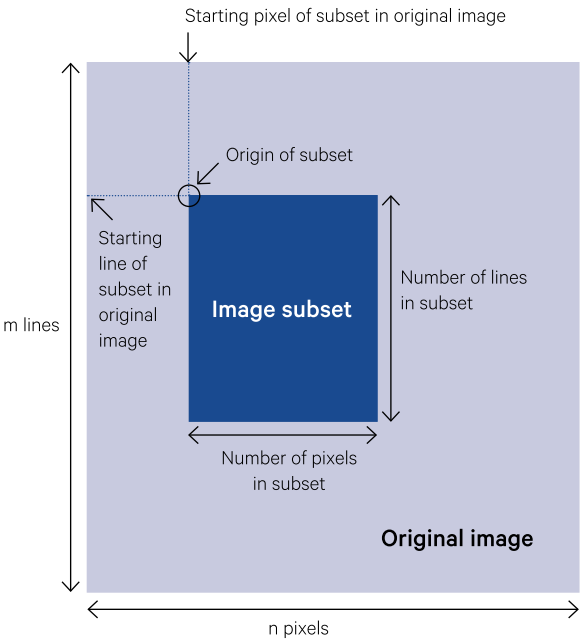
Modifying image scale potentially introduces artefacts into the output image. The impact of upscaling (creating larger pixels by blocking) and downscaling (creating smaller pixels by enlarging) EO imagery is considered in Volume 1B—Section 2.4.

7.2.1.1 Subsetting

Image subsetting simply involves extracting one or more portions of an image file to become separate image files or 'subsets'. This portion is usually specified as a starting location, that is the origin in terms of the original image coordinates (see Figure 7.1), and the size of the new image subset (the number of pixels and number of lines). This operation generally assumes that the number of channels is not modified between images, although the output image could contain fewer channels if required. Figure 7.2 illustrates the parameters required for image subsetting.

Figure 7.2 Subsetting an image

A sub-area within an original image can be defined as a given number of pixels and lines from an origin pixel location.



Source: Harrison and Jupp (1990) Figure 16

7.2.1.2 Sewing and mosaicking

A reciprocal action to image subsetting is image sewing or mosaicking where several smaller images are joined together to form a larger image. A mosaic can be viewed as a picture or design formed by cementing coloured objects onto a surface (Ahuja and Schachter, 1983). Multiple input images, possibly with different geometries, can be mosaicked onto a new grid, or base map, using image resampling (see Section 7.3). This process requires that the relationship between each input image and the base map be defined precisely, generally by applying a multi-stage modelling approach as detailed in Volume 2B.

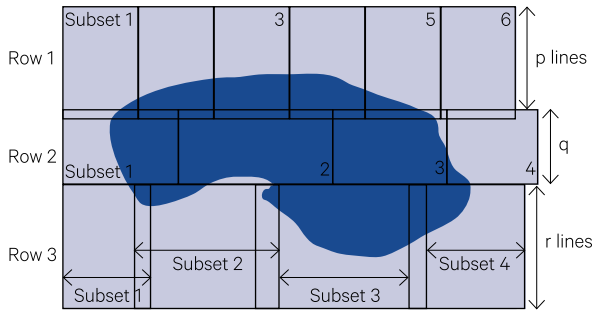
Image sewing usually refers to the operation of patching subsets together assuming that all subsets are compatible in terms of scale and orientation. This operation does not alter the shape or scale of the (sub)image in any way (see Figure 7.3). By contrast, image mosaicking allows the subsets that are being joined to be altered geometrically, in terms of shape, scale and orientation, before joining (see Section 7.3).

Image sewing simply requires that the origins of subsets being joined be specified relative to a consistent image coordinate system. The operation of image mosaicking can additionally use a range of sophisticated models to account for geometric differences between the image subsets being joined and/or modify the scale of the output image relative to the input images. This process is especially relevant to EO data where the along-line and along-track distortions can vary significantly over a complete image scene and hence differ between subsets of any given scene (see Section 7.3 and Volume 2B).

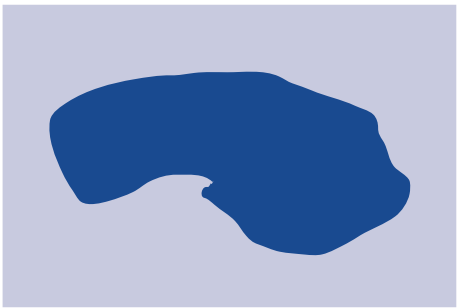
Figure 7.3 Joining multiple images

Within a row of subsets, the number of pixels in, and the number of pixels being copied from, each image can be different, but the same number of lines is copied from each input image. This ensures that areas of overlap in the input images are represented only once in the output image.

a. Input images



b. Output image



Source: Harrison and Jupp (1990) Figure 17

7.2.1.3 Reducing

Image size and scale may be reduced (without subsetting) by the operations of blocking or sub-sampling. Reducing image size is sometimes called image 'minification'.

Image blocking reduces image size by averaging the values of multiple pixels in the input image into a single pixel in the output image. The selected block size determines the number of pixels that are averaged. The effect of a 2 pixel by 3 line block on an input image of 10 pixels and 10 lines is shown in Figure 7.4 In this case, since the blocking factor is not evenly divisible into the number of lines in the input image, any partial blocks are not included in the output image. By default, blocking generally starts from the first pixel in an image.

Figure 7.4 Blocking an image

This example uses a single channel image of 10 pixels and 10 lines with a 2x3 blocking factor to produce an output image with 4 pixels and 3 lines.

a	b	5	6	7	6	6	5	4	3
c	d	4	5	6	6	5	5	5	4
e	f	3	7	6	7	6	5	6	7
3	4	4	5	5	7	7	6	6	7
4	4	4	4	5	6	6	7	7	7
5	5	4	5	6	6	5	6	6	5
5	4	4	4	5	6	7	8	4	3
4	3	3	4	5	7	9	6	2	1
4	2	3	8	7	9	9	7	3	1
2	3	3	9	9	8	7	6	2	2

where (a+b+c+d+e+f)/6 = A

A	5	6	5	5
4	4	6	6	6
4	4	7	8	2

Source: Harrison and Jupp (1990) Figure 18

Figure 7.5 Sub-sampling an image

a. An input image with 20 m pixel size is sub-sampled to simulate an output image with 40 m pixels by dropping alternate pixels and lines in the image

Input image pixel size = 20 m

a	b	c	d	e	f	g	h
i	j	k	l	m	n	o	p
q	r	s	t	u	v	w	x
y	z	A	B	C	D	E	F
G	H	I	J	K	L	M	N
O	P	Q	R	S	T	U	V
W	X	Y	Z	1	2	3	4
5	6	7	8	9	10	11	12

Output image pixel size = 40 m

a	c	e	g
q	s	u	w
G	I	K	M
W	Y	1	3

b. An input image with 20 m pixel size is sub-sampled to simulate an output image with 30 m pixels.

Input image pixel size = 20 m

a	b	c	d	e	f	g	h
i	j	k	l	m	n	o	p
q	r	s	t	u	v	w	x
y	z	A	B	C	D	E	F
G	H	I	J	K	L	M	N
O	P	Q	R	S	T	U	V
W	X	Y	Z	1	2	3	4
5	6	7	8	9	10	11	12

Output image pixel size = 30 m

a	c	d	f	g
q	s	t	v	w
y	A	B	D	E
O	Q	R	T	U
W	Y	Z	2	3

Source: Harrison and Jupp (1990) Figure 19

In image analysis, blocking can be used to develop spatial statistics by simulating the effect of a larger pixel size (see Section 8.2). The process may also be applied after resampling to a very small pixel size to produce a more continuous image (see Volume 2B).

The size and 'density' of an image may also be reduced by sub-sampling. This process effectively represents a larger pixel size by selecting the value of a single pixel in the input image rather than using the average of a block of pixels as is the case for image blocking. Figure 7.5a and Figure 7.5b demonstrate this process for converting from a 20 m pixel size to output pixel sizes of 40 m and 30 m respectively.

7.2.1.4 Enlarging

The reverse process to reducing image size is image enlargement, where a single pixel in the input image is duplicated to represent multiple pixels in the output image. This process is sometimes called image 'magnification' and has the effect of changing image scale by spatially sub-dividing the input pixels as shown in Figure 7.6a. This enlargement process may be performed in conjunction with precise scaling as described in Volume 2B, or more simply by using a consistent multiplicative factor. Figure 7.6b shows the result of enlarging an image by converting from a 50 m pixel to a 20 m pixel using along-line resampling (see Volume 2B). Similarly, this process can be used to convert a single line and/or pixel image into a larger image to match the extent of another image as shown in Figure 7.6c.

Figure 7.6 Enlarging an image

A single input pixel could be expanded to a larger number of output pixels by specifying an output pixel width that is smaller than the specified input width in the same proportion as the required expansion.

a. Duplicate each pixel to represent a smaller pixel size

Input image pixel size = 40 m

a	b	c	d
e	f	g	h
i	j	k	l
m	n	o	p

Output image pixel size = 20 m

a	a	b	b	c	c	d	d
a	a	b	b	c	c	d	d
e	e	f	f	g	g	h	h
e	e	f	f	g	g	h	h
i	i	j	j	k	k	l	l
i	i	j	j	k	k	l	l
m	m	n	n	o	o	p	p
m	m	n	n	o	o	p	p

b. Enlarge image by changing relative pixel size from 50 m to 20 m

Input image pixel size = 50 m

a	b	c	d
e	f	g	h
i	j	k	l
m	n	o	p

Output image pixel size = 20 m

a	a	b	b	b	c	c	d	d	d
a	a	b	b	b	c	c	d	d	d
e	e	f	f	f	g	g	h	h	h
e	e	f	f	f	g	g	h	h	h
e	e	f	f	f	g	g	h	h	h
i	i	j	j	j	k	k	l	l	l
i	i	j	j	j	k	k	l	l	l
m	m	n	n	n	o	o	p	p	p
m	m	n	n	n	o	o	p	p	p
m	m	n	n	n	o	o	p	p	p

c. Matching single pixel, or line of pixels, to a larger image

Input pixel depth = 60 m (1 line)

a	b	c	d	e

Output pixel size = 120 m (6 lines)

a	b	c	d	e
a	b	c	d	e
a	b	c	d	e
a	b	c	d	e
a	b	c	d	e
a	b	c	d	e

Source: Harrison and Jupp (1990) Figure 20

7.2.2 Changing image orientation

Image orientation may be modified in a number of ways:

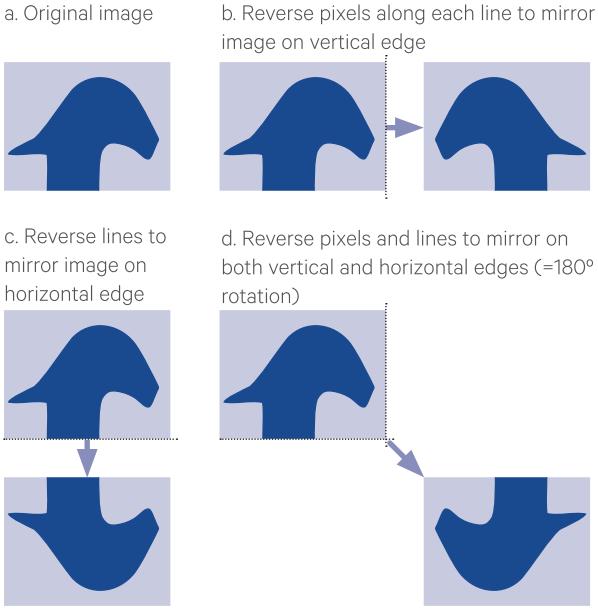
- reversing pixels and/or lines (see Section 7.2.2.1);
- transposing image along a diagonal (see Section 7.2.2.2); and
- rotating the image origin (see Section 7.2.2.3).

As noted in the sub-sections below, different combinations of these operations can be used interchangeably to achieve a particular result. The development of polynomial models that involve image rotation, and the process of resampling imagery, are introduced in Section 7.3, and detailed in Volume 2B.

7.2.2.1 Reversing pixels and/or lines

The reversal of pixel order along an image line produces a mirror image of the original image representation at a vertical edge. Similarly, the order of lines down an image can be reversed by simply mirroring the original image along the first or last line. Such processes are illustrated in Figure 7.7.

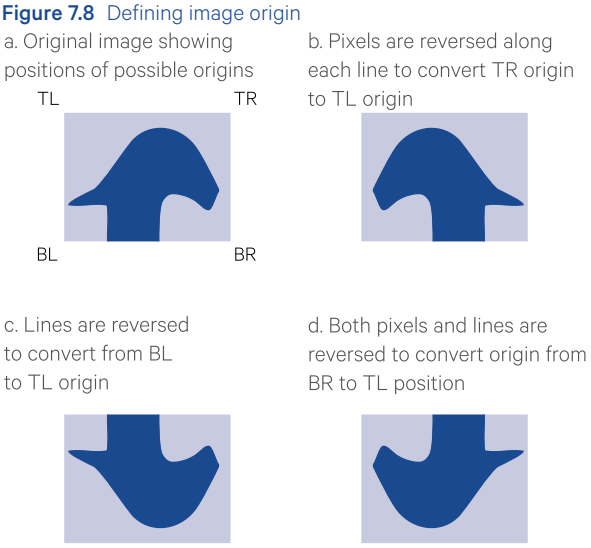
Figure 7.7 Reversing pixels and/or lines in an image



Source: Harrison and Jupp (1990) Figure 21

Since these operations do not involve image rotation, the required reversals can generally be specified by indicating which corner of the original image is to be the top-left corner in the processed image. The corner positions are typically labelled relative to the original image orientation, that is, top-left (TL), top-right (TR), bottom-left (BL) and bottom-right (BR) (see Figure 7.8). Lines and/or pixels are then reversed and/or swapped as required to produce an output image with the defined origin in the top-left position.

Figure 7.8 Defining image origin

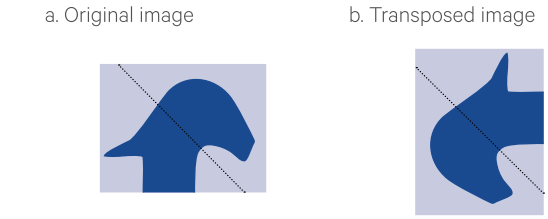


Source: Harrison and Jupp (1990) Figure 22

7.2.2.2 Transposing an image

This process involves rewriting the image with the rows of the input image being columns in the output image and columns of the input image being output rows (see Figure 7.9). The same result could be achieved by rotating the image 90° (see Figure 7.10b) then reversing pixels along each line (see Figure 7.7). For square images, the effect of this operation is to flip the image about the diagonal axis between the top-left corner and the bottom-right corner.

Figure 7.9 Transposing an image

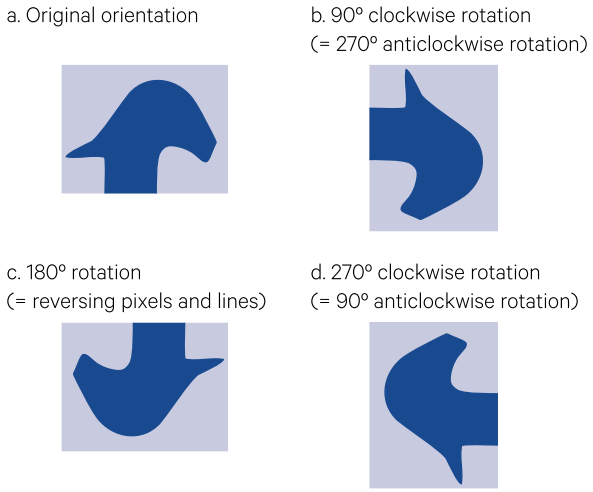


Source: Harrison and Jupp (1990) Figure 23

7.2.2.3 Rotating image origin

The image origin can also be altered by rotating an image by 90°, 180° or 270° in a clockwise or anti-clockwise direction (see Figure 7.10). These simple rotation angles can be applied to an image using combinations of pixel and/or line reversals (see Section 7.2.2.1) and image transpositions (see Section 7.2.2.2) as illustrated in Figure 7.8, Figure 7.9 and Figure 7.10. A rotation of 90° in a clockwise direction is equivalent to transposing the image, then reversing pixels (see Figure 7.10b). Rotating by 180° (see Figure 7.10c) is the same as reversing both lines and pixels (see Figure 7.7d and Figure 7.8d above), while a 270° rotation in a clockwise direction (see Figure 7.10d) gives the same results as reversing the lines after transposing an image.

Figure 7.10 Rotating an image



Source: Harrison and Jupp (1990) Figure 24

7.3 Rectification and Resampling

Previous sub-sections introduce the concepts of geometric manipulation of imagery and describe operations which may be required to modify the scale or orientation of imagery before processing. Those operations described thus far, however, do not actually change the 'shape' of the image nor correct for geometric distortions.

EO imagery contains a variety of geometric distortions which need to be accounted for before registering with different image sources or map data (see Volume 2B—Section 2). Image processing software can implement a range of models to account for known geometric distortions in satellite and aircraft scanner imagery as well as map projection models which compensate for the spatial inconsistencies between different map coordinate systems (see Volume 2B—Section 3). These models can be used in conjunction with control point modelling techniques to achieve accurate registration between different coordinate systems (see Volume 2B—Section 4). The fundamentals of map projections are introduced in Volume 2B—Section 1.

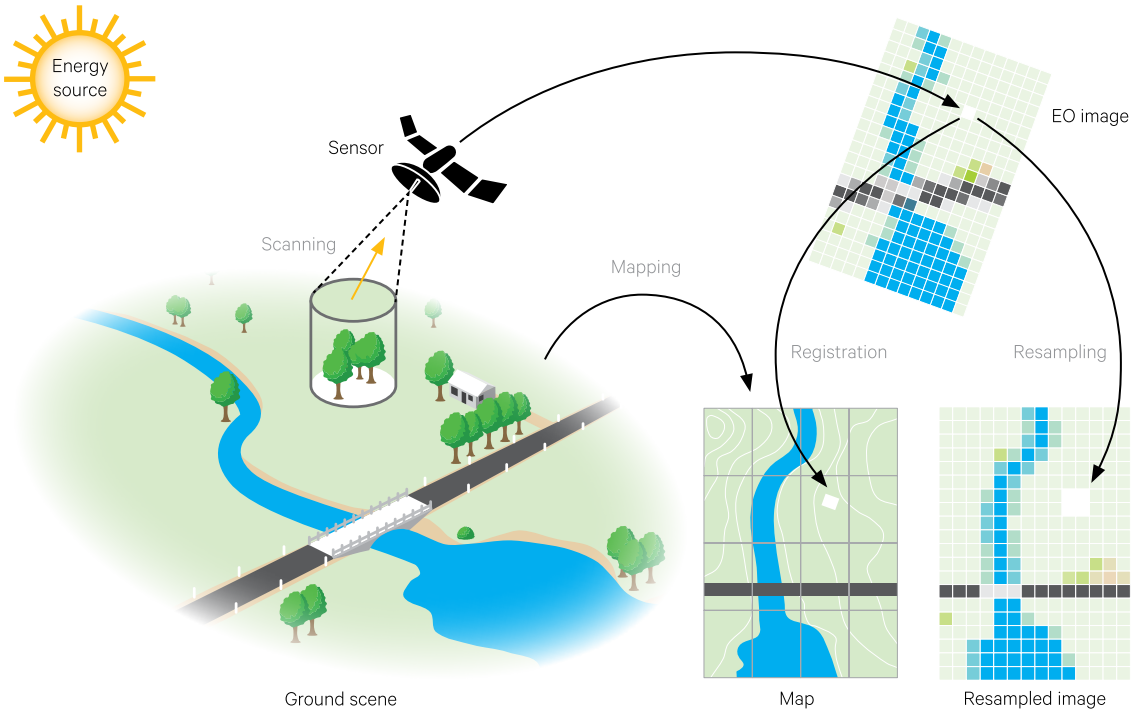
Image rectification involves modelling the geometry and topology of an image. Image registration models the relationship between the geometry of an image and the geometry of some other spatial representation of the imaged scene (such as a map or another image). Image resampling uses the modelled relationship between two 'geometries' to rewrite the geometry of one dataset to match the other, for example, to overlay an image onto another image or a map (see Volume 2B—Section 5). This process changes the spatial relationship between objects in the image, and may be compared to selectively stretching sections of the image as if it were a rubber sheet. Resampling can also be applied to multiple input images, to create a mosaic (see Section 7.2.1.2).

The rationale for image rectification and registration in the context of EO images is illustrated in Figure 7.11. The left side of this figure shows a ground scene being imaged by an EO sensor. Each optical pixel—a remote measurement that is integrated over a 'volume' on the surface being imaged—relates to one geometric pixel in the EO image (see Volume 1B—Section 1.2). In Figure 7.11, one optical pixel is shown as a cylinder. The relationship between the geometry of the resulting image and the geometry of the scene being imaged may be described by modelling the optical and physical relationship between the scene and the sensor. As illustrated on the right side of Figure 7.11, this relationship is commonly defined relative to a map of the scene.

I think the universe is pure geometry—
basically a beautiful shape twisting around
and dancing over space-time.
(Antony Garrett Lisi)

Figure 7.11 Earth Observation scene geometry

This illustration shows a ground scene being imaged by a remote sensor. A 'volume' on the surface that is observed as one optical pixel is highlighted by the cylinder. The corresponding geometric pixel for this location is shown as a white square in the EO image. While the EO image contains recognisable features from the ground scene, its geometry rarely matches a standard map projection. To rectify the image to map coordinates, the locations of specific features needs to be identified on a map. For example, the location of the highlighted optical pixel is shown here as a white square on the map. To allow EO imagery to be digitally or physically registered to a map, it then needs to be resampled to match the map geometry. As shown in this example, the original optical pixel may be represented by multiple geometic pixels in the resampled image.



Adapted from Harrison and Jupp (1992) Figure 1

As detailed in Volume 2B, transformations between two different geometries, such as an EO image and a map, usually involve three stages:

- a nominal coordinate transformation which corrects for known image distortions;
- a nominal map transformation, which accounts for inconsistencies in map coordinates; and
- a polynomial transformation to relate the two nominally transformed geometries.

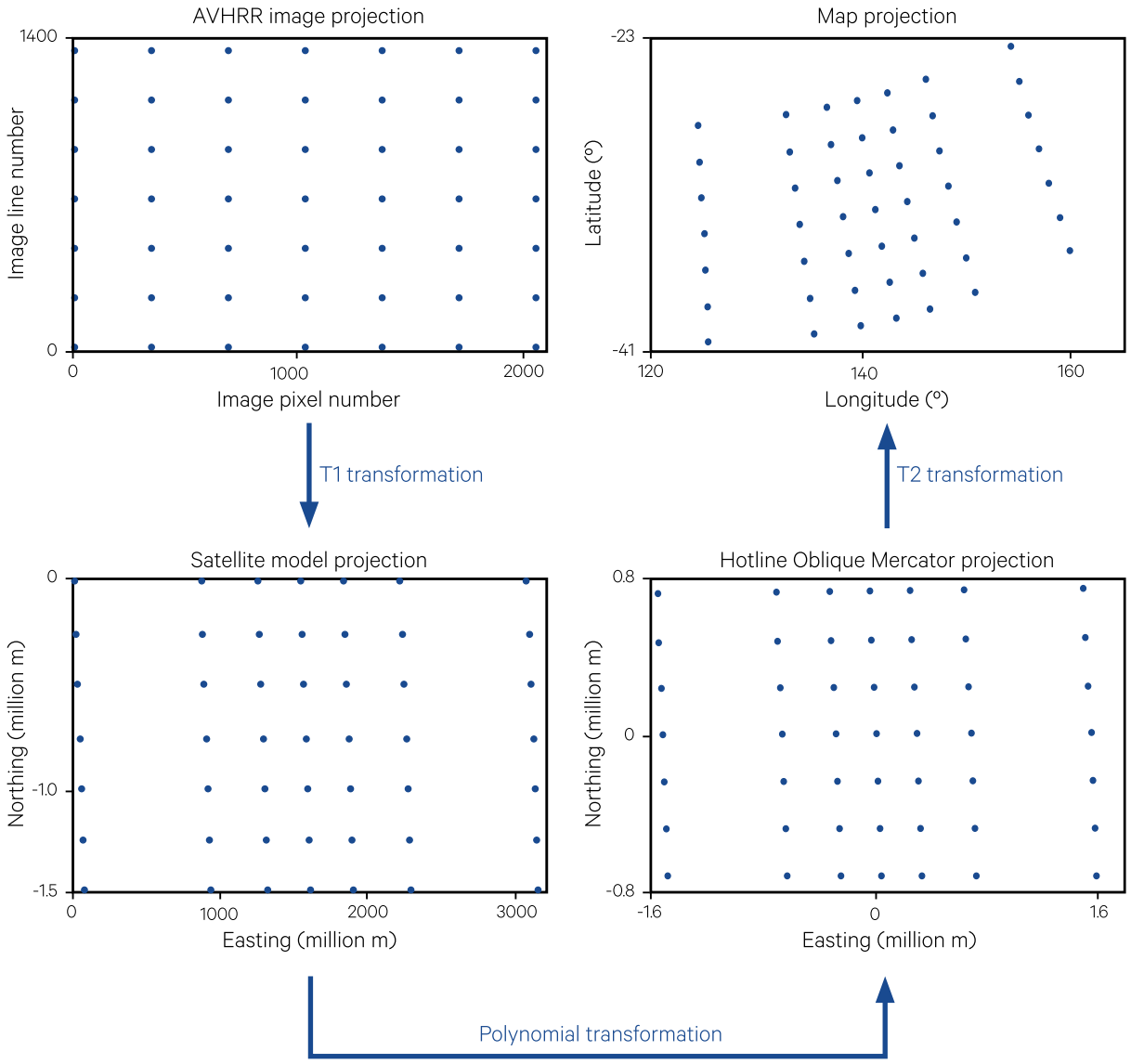
To convert from coordinates in an image to a map, for example, the modelling sequence would be:

- convert from EO image coordinates to nominal image coordinates using a pre-defined image model;
- convert from nominal image coordinates to nominal map coordinates using a polynomial model derived from ground control points; and
- convert from nominal map coordinates to coordinates of selected map projection using a pre-defined map model.

This three-stage approach allows the relationship between the two nominally transformed geometries to be modelled by a lower order (and thus more stable) polynomial model. Figure 7.12 illustrates this process for an AVHRR image being registered to geographical coordinates. Development of appropriate rectification and registration models for EO imagery is described in Volume 2B.

Figure 7.12 Three-stage rectification process

This schematic shows the three stages of modelling from evenly-spaced pixel locations in an AVHRR image to geographical coordinates. The T1 transformation uses a pre-defined satellite model to convert from image coordinates to a grid based on metres Similarly, the T2 transformation uses pre-defined cartographic models to convert between map coordinates and a metre grid. The relationship between the two modelled grids can then be computed as a polynomial model of lower order than would have been required to map directly from image to map coordinates. A corresponding set of inverse models are used to convert from map to image coordinates.



Source: Harrison and Jupp (1992) Figure 31

Generally, the scene geometry is represented by an independent mapping exercise so that the image is viewed as being registered with the scene when it is registered to a map of the scene. To sensibly overlay the image onto the map, or map data onto an image, both the map and image data need to be viewed using the same projection and datum (see Figure 7.13). This can be done either by resampling the image data to the same projection as the map or, for visual presentation only, converting the map and/or image data to a matching projection during the image display process. The latter option assumes that the map and image data are represented in projections that can be 'interconverted' during image display. The resampling process for EO imagery is detailed in Volume 2B.

Figure 7.13 Overlaying map data

Aerial imagery acquired in 2016 and orthorectified to the ACT (Stromlo) Grid, which is based on the Australian Geodetic Datum 1966 (AGD66)[3]. In this example, a portion of this imagery over central Canberra is overlain with a GIS layer that shows the actual extent of lakes. The GIS layer, however, is stored in MGA 55 coordinates relative to the Geodetic Datum of Australia 1994 (GDA94). MGA coordinates based on GDA94 are shifted approximately 200 m northeast of MGA coordinates based on AGD66.

a. Overlay of Lakes GIS layer without matching the image coordinate system (in this case due to a different datum) results in an overlay shift to the northeast of approximately 200 m relative to image features.

b. Overlay of Lakes GIS layer after transformation to the same coordinate reference system used in the image.

Source: Tony Sparks, Icon Water, based on ACT Government orthorectified imagery supplied by AAM (Ref. No.: 26121A) CC BY 4.0.

7.4 Image grids

Resampled imagery implicitly conforms either to a:

- custom grid system—whose scale and geometry is customised for a specific EO sensor (see Section 7.4.1); or a
- nested grid configuration—which is generic for a range of sensors and products (see Section 7.4.2).

7.4.1 Sensor-specific custom grids

To allow EO data to be overlaid with map data or other imagery, the original geometry of the data is modified to conform to a selected grid. The image grid is based on an appropriate map coordinate system and a grid cell size that suits the spatial resolution of the original data.

Custom grids used by many EO data suppliers differ for each data source and do not necessarily conform to a common underlying grid cell size or map coordinate system. Consequently, when applications require that imagery from different EO sources be compared, one or both images need to be further resampled to a common map projection and/or custom grid. If not conducted with care and consistency, this process may introduce additional artefacts, especially when multiple images are being analysed. Accordingly, a new range of EO image products is being offered, which attempt to standardise image grids between different data sources (see Section 7.4.2).

7.4.2 Sensor-agnostic nested grids

Traditionally EO imagery has been reprojected onto a grid whose scale and geometry is customised for a specific EO sensor and conforms to a relevant map projection (see Section 7.4.1). Customised grids are not necessarily compatible between different sensors, however, thus complicating comparison of imagery from different sources. To address this problem, the Australian and New Zealand Land Information Council (ANZLIC) have developed standards for a National Nested Grid (NNG), with scalable grid cell sizes (ANZLIC, 2012). The grid cell sizes of products comprising this suite are designed to be scalable (such as 25 m, 250 m and 1000 m) and thus facilitate easy data comparison between different products. This format will greatly simplify comparison of image datasets acquired by different sensors at varying spatial scales.

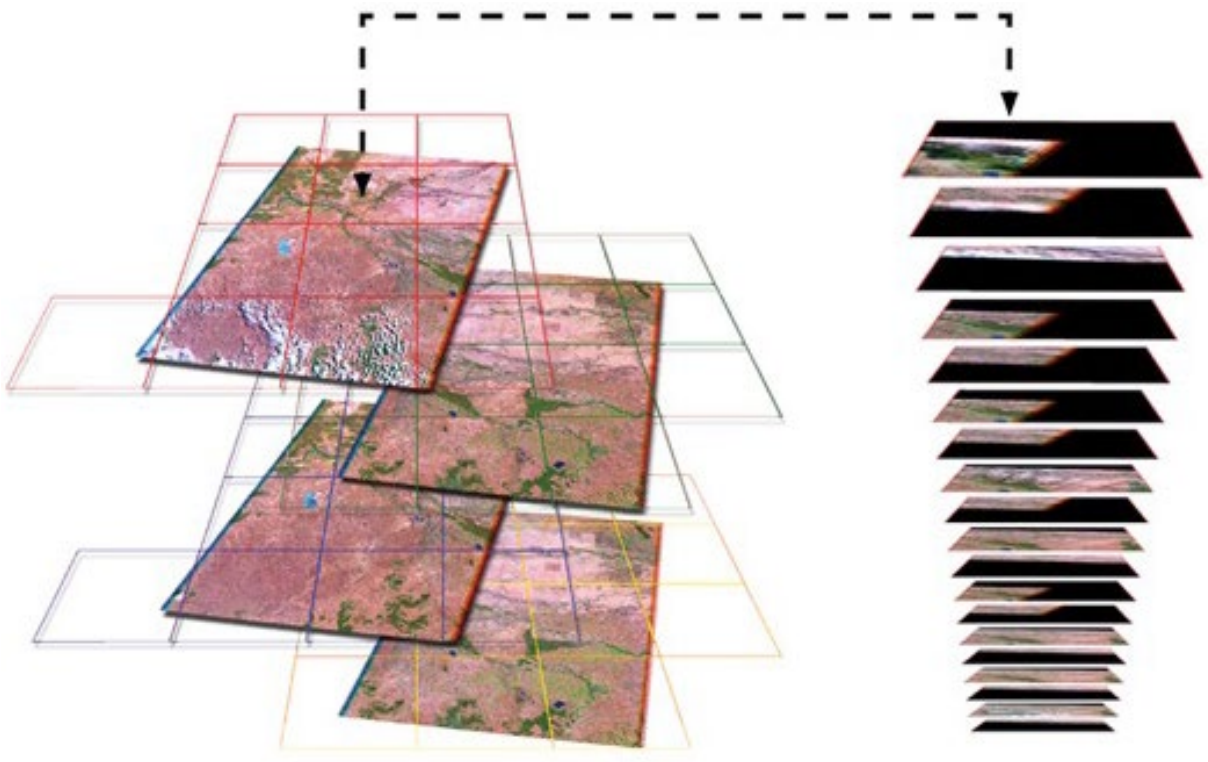
Accordingly, major EO product suppliers are planning to offer EO products that share a common underlying nested grid base with scalable cell sizes. Nested grids will allow EO data to become 'sensor agnostic' and readily comparable with spatial data from other sources. This feature is particularly relevant given the growing time series of EO imagery and will significantly simplify the process of monitoring long term changes in land and water resources.

3 For further information on the ACT projection, please visit: *https://www.planning.act.gov.au/tools_resources/survey-data-maps/surveying-data/surveyors_information/coordinate_system/act_standard_grid_parameters*

For example, the ARG25 is the first NNG-compatible product to be released by GA in the DEA (see Section 3.4). This product is structured as a data cube, with time as the third dimension (see Figure 7.14). The surface reflectance data for each imaged date and time are stacked in chronological order, with Landsat TM/ETM+ sources being interleaved as appropriate to their acquisition time. Interrogation of this data cube allows the time series for individual grid cells, as well as their heritage and data quality, to be extracted and compared. The resulting stack of surface reflectance grids is effectively consistent over space and time. This product greatly reduces the processing requirements for numerous studies, and enables the development of a new set of tools for identifying and quantifying environmental change in land and water resources across Australia (see Volume 3).

Figure 7.14 GA data cube for ARG25 product

The data cube organises images from multiple sources and times, such as the images on the left, into stacks of tiles on the right, with the top of the stack being the most recent observations.



Source: Alex Ip, Geoscience Australia.

7.5 Further Information

Intergovernmental Committee on Surveying and Mapping: *http://www.icsm.gov.au/index.html*

Jensen (2016) Chapter 7

7.6 References

Ahuja, N., and Schachter, B. J. (1983). *Pattern Models*. John Wiley and Sons, New York.

ANZLIC (2012). *National Nested Grid (NNG) Specification Guideline – 2012*. ANZLIC NNG Workgroup.

Harrison, B. A., and Jupp, D. L. B. (1990). *Introduction to Image Processing. Part TWO of the microBRIAN Resource Manual* (256 pages). CSIRO Australia, Melbourne.

Harrison, B. A., and Jupp, D. L. B. (1992). *Image Rectification and Registration. Part FOUR of the microBRIAN Resource Manual (draft only)*. MPA, Melbourne.

Jensen, J.R. (2016) *Introductory Digital Image Processing. A Remote Sensing Perspective*. 4th edn. Pearson Education, Inc.

Analysis

Image analysis involves using tools based on statistics and/or numerical analysis to identify features and patterns in the image. The following sections introduce the image analysis topics of:

- statistics (see Section 8);
- classification (see Section 9);
- and segmentation (see Section 10).

These topics are considered in greater detail in Volumes 2C, 2D and 2E.

Contents

Background image on previous page: Landsat-8 image over Weipa, Queensland, acquired on 8 August 2013. This composite image displays bands 6, 5, 3 as RGB and has been sharpened using the panchromatic band. Weipa is situated on the western coast of Cape York Peninsula, facing the Gulf of Carpentaria, and is known for numerous bauxite mines. **Source:** Norman Mueller, Geoscience Australia

8 Statistics

Sections 4 to 7 above introduce the concept of a digital image as a spatial dataset having properties of both colour and geometry. Being in numeric form, digital image data may also be processed using standard statistical techniques to effectively analyse its 'colour' properties. These techniques may be categorised into four groups:

- spectral—techniques that effectively disregard the spatial properties of image data so that an image channel is treated as a one-dimensional array of numbers. These techniques derive statistics that describe the image in terms of the range and variation of data values they contain (see Section 8.1).;
- spatial—describe the data values of an image as they relate to its spatial patterns. These statistics can be used to examine the spatial patterns within the image data and quantify the effects of changes in image spatial resolution (see Section 8.2);
- radiometric—methods to analyse components of image (or ground calibration) values that are not due to surface reflectance, such as atmospheric components, BRDF, terrain-related variations in illumination and scanner inconsistencies (see Section 8.3); and
- temporal—ways to standardise multi-date image data, potentially comprising two to tens of thousands of images, to enable meaningful quantitative comparisons and analysie (see Section 8.4).

8.1 Spectral

A digital image represents spatial attributes as arrays of numbers. Statistics relating to the distribution of values in each channel, or the similarity of values (or correlation) between pairs of channels, can be derived from a digital image. These statistics can then be used to analyse image quality, extract particular spectral features, reduce data volume or transform the data in a variety of ways.

While image data may relate to a wide range of attributes, in EO imagery data channels (or bands) typically represent a measure of spectral radiance or reflectance. Although the following discussion refers to 'spectral' image values, these statistics could be derived for any image attribute data.

8.1.1 Histograms

The distribution of data values in any image channel may be represented as a histogram as introduced in Section 4.1.1. The usual form for histograms in image processing is the 'frequency histogram' shown in Figure 4.2, with the horizontal axis representing the image data range and the vertical axis indicating the number of pixels in the image with each data value.

As the scale of a printed or displayed histogram may not be sufficiently large to enable values to be read exactly, ancillary statistics relating to the absolute minimum, maximum and mode are usually reported separately. In the example in Figure 4.2, no pixels had values below 15 or above 137. The total number of pixels used to generate the histogram is also usually reported: 61440 in this case. If the whole image was used to generate the histogram, the total number of pixels in a channel will equal the number of pixels per line multiplied by the number of lines minus the number of null pixels in that channel. In many image processing systems, histograms may also be generated for sub-regions of an image, on the basis of a sub-range of values in one or more channels (called a spectral theme) or a sub-area (called a training patch or training set). The use of spectral themes and training patches are further discussed in Section 9.1 in the context of image classification.

Percentage points (or percentiles) for the histogram are also generally provided. These values summarise

Background image: Landsat-8 image over Weipa, Queensland, acquired on 8 August 2013. This composite image displays bands 6, 5, 3 as RGB and has been sharpened using the panchromatic band. Weipa is situated on the western coast of Cape York Peninsula, facing the Gulf of Carpentaria, and is known for numerous bauxite mines. **Source:** Norman Mueller, Geoscience Australia

the histogram distribution in terms of cumulative statistics (see Figure 4.2 and Figure 4.3). The percentage points provide a 'normalised' view of histogram values that is useful for operations that require histograms to be compared between channels (see Volume 2D—Section 1.4).

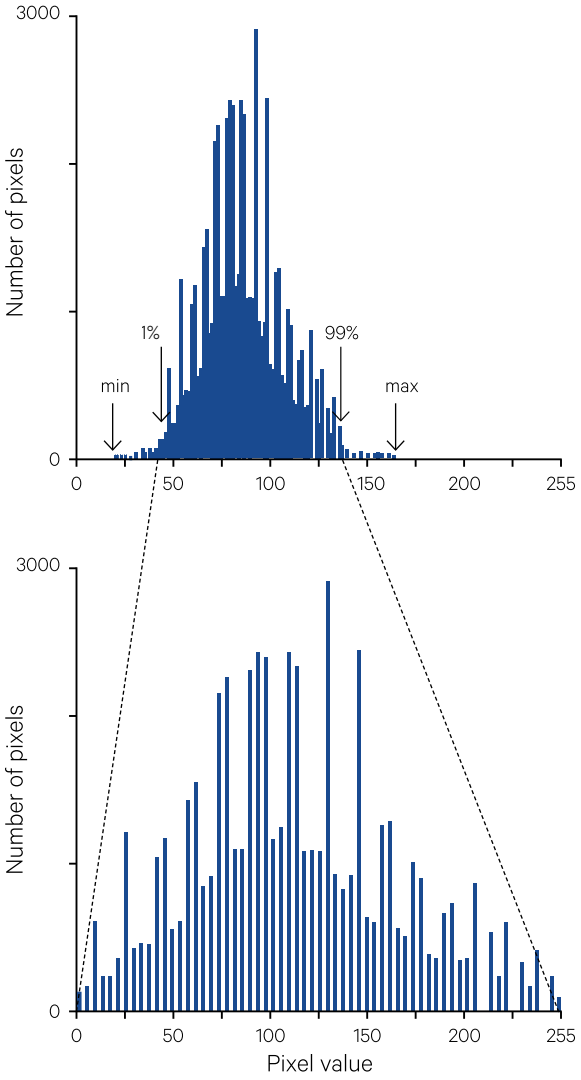
The histogram shape is also useful in interpreting the image data. Skewed data distributions commonly occur in remotely sensed imagery. Image acquisition and calibration operations can sometimes truncate the upper or lower limits of the available data range. For example, during image acquisition the sensor gains may not be set correctly to capture the full range of radiance values from a particular ground scene. When the scanner has been set to measure a much larger range of intensities than occurs in the scene, the range of intensity values in the channel will be relatively narrow. In EO imagery, a narrow range can also be due to poor illumination (such as a low Sun angle). Alternatively, the minimum radiance level detected by a sensor may be higher than the minimum radiance from an image scene and/or the maximum level being detected may be lower than the maximum radiance causing the discrimination between very low and/or very high values to be lost. The image histograms can be used to identify such problems. Histograms should also be checked after all scaling operations, such as transformations, to ensure that the scaled value range of the output image has not been compressed or truncated unnecessarily (see Volume 2C).

Discontinuous histograms are sometimes encountered in EO imagery. Where image data have been recalibrated from 8 bits (values in the range 0–255) to a 9 or 10 bit data range, discontinuities may occur in the image data range so that certain values within the range are not represented in the image. This effect is usually due to pre-processing with an inadequate recalibration algorithm that does not completely account for sensor differences. For similar reasons, image values may be unevenly distributed as shown in Figure 8.1a.

It is the mark of a truly intelligent person to be moved by statistics.
(George Bernard Shaw)

Figure 8.1 Effect of rescaling on image histogram

The original image histogram (top) has a mostly continuous histogram between values 15 (1% minimum) and 137 (99% maximum). When the image values are rescaled to fill the full 8-bit (byte) range (0 to 255), the histogram of the rescaled image (below) is discontinuous.



Source: Harrison and Jupp (1990) Figure 40

Discontinuities may be visible in imagery that has been stretched, especially when the image data span a small range (see Volume 2C). The cause of this effect is shown in Figure 8.1 where a channel, which contains only 60 different pixel values between the 1% and 99% points, is re-written to fill the whole 8-bit (byte) image data range (256 values). Each of the original values maps to a unique value in the expanded range to produce only 60 different output values. Discontinuities in image histograms may also be due to spectral features of the image itself as illustrated in Figure 4.4, or radiometric irregularities, such as sensor striping (see Section 8.3).

8.1.2 Variance

The calculation and interpretation of variance (σ^2) and standard deviation (σ) statistics for an image channel are introduced in Section 4.1.2. The standard equation for variance is given as:

$$\text{channel variance}\left(\sigma^2\right) = \frac{\sum\left(x-\mu\right)^2}{n}$$

where

x is pixel value in channel;
n is total number of pixels in channel (excluding null pixels); and
μ is true mean of the channel data.

This equation can only apply when the real mean of the data values is known exactly, rather than just being estimated from a set of sample values. In image processing, statistics are frequently generated by sampling image values and are usually interpreted to be representative of larger image areas. Thus, the mean value used to calculate variance is only an estimate of the real data mean and the equation is usually modified to use one less degree of freedom as:

$$\frac{\sum\left(x-\overline{x}\right)^2}{n-1}$$

where

$$\text{channel mean}\left(\overline{x}\right) = \frac{\sum\left(x\right)}{n}$$

An alternative formula for variance is:

$$\frac{\sum\left(x\right)^2 - n\left(\overline{x}\right)^2}{n-1}$$

This form can produce very unstable results (such as negative variance values) so should not be used. The following equations are used to update mean and variance statistics in some image processing systems:

$$\overline{x}_n = \overline{x}_{n-1} + e_{n-1}$$

$$\sigma_n^{\,2} = \frac{n-1}{n}\sigma_{n-1}^{\,2} + \left(n-1\right)e_{n-1}^{\,2}$$

where

$$e_{n-1} = \frac{x_n - \overline{x}_{n-1}}{n}$$

and

n is the number of samples.

For large values of n, as typically occur in image analysis, the difference between dividing by n or n–1 is not significant.

Null pixels should be excluded from all statistical calculations in image processing. In most image processing systems, when the variance of a single channel is being computed, a pixel is only treated as masked or excluded from the image if its value equals the null value in that channel. In some multi-channel operations, however, a pixel will be excluded if it has a null value in any channel.

If a channel has a low variance (and hence low standard deviation) we would expect that most pixel values would be close to the mean value (see Section 4.1.2). The standard deviation is related to the probability of a pixel having a particular value in the channel range. A theorem developed by Chebyshev allows us to define, for any data distribution, a lower bound estimate of the probability that a random variable falls within a particular range of the mean; namely, there is greater than or equal to 75% probability that a random pixel value in the range will be within 2σ units of the mean, or at least 89% probability that it will be within 3σ units (Walpole, 1974). Such probabilities can be determined more precisely if the probability distribution of the data is known.

Channel standard deviation is a particularly important statistic for checking training patches when developing an image classification (see Section 9 and Volume 2E). Typically, variance of individual image classes as well as the overall variance for a classification can be computed during the classification process.

8.1.3 Crossplots

Sections 8.1.1 and 8.1.2 above discussed image statistics relating to single channels. These statistics were derived by treating the image channel as a one-dimensional array of numbers, with no account being taken of the location of the pixel values. Most imagery, however, contains multiple channels. Various statistics, which indicate the correlation between channels, may be calculated in the same way, by considering the multi-channel image as a multi-dimensional matrix (see Figure 1.1). For example, the six optical data channels in a Landsat TM image can be considered as a six-dimensional matrix.

The simplest way to view the correlation between channels is to construct crossplots for pairs of channels (see Excursus 8.1). A crossplot is a two-dimensional graph with the horizontal axis showing the values of one channel, the vertical axis showing the values of the second channel and the points on the graph indicating the 'paired values' of pixels in the image. These points can also be colour-coded to indicate the relative number of pixels with each paired value. The crossplots in Figure 8.2 illustrate

the correlation between channels A, B and C in the example image, with the number of pixels having each paired value indicated by the number of dots.

To define the relationship between the two channels, a regression function can be fitted to the crossplotted data. The significance of the function can also be computed and, if relevant, the regression line is displayed on the crossplot as illustrated in Figure 8.2b. It may also be useful to compute the number of pixels at each paired set of values represented on a crossplot by cross-tabulation of the channels.

Correlation between channels indicates a degree of redundancy in image data. If channels are highly correlated the image data can be reduced to a smaller number of dimensions (or channels). This may be advantageous for efficiency in processing, transmission or storage, or to display the three most significant dimensions. Principal Component Analysis (PCA) allows image data to be reduced to fewer dimensions (see Volume 2C). While correlation between channels may simplify image handling, it may also indicate limitations for feature recognition. In the case of Landsat TM for example, the visible channels TM1, TM2 and TM3 are usually highly correlated. The same applies to the near infrared and shortwave infrared channels, TM5 and TM7. This means that in a broad sense there may be only two or three different data dimensions for identifying image features. This aspect of channel correlation is most significant for the design of sensors and choice of airborne scanner channels for particular remote sensing projects.

Excursus 8.1—Band Correlation Statistics

For simplicity, this example is based on a three-channel image containing the pixel values shown in Figure 8.2. Mean and variance statistics (see Section 8.1.2) for these channels are:

$n = 25$

$x_A = 3.3$

$x_B = 4.2$

$x_C = 3.12$

$\sigma^2_A = 2.938$

$\sigma^2_B = 118/25 = 4.72$

$\sigma^2_C = 56.64/25 = 2.2656$

The crossplot graphs in Figure 8.3 were constructed by locating the channel values of each pixel in the image along the relevant axes. For example, in Figure 8.3a, for each pixel in the image, the channel A value was located on the horizontal axis, the value of channel B on the vertical axis, then a dot is marked on the graph at the intersection of these two values. On each crossplot, this process is shown for the values of pixel 1, line 5 in the image.

Figure 8.2 Example image channels

Channel A

	1	2	3	3	4
	3	3	5	6	3
	2	3	3	6	5
	5	4	2	2	1
pixel 1, line 5 →	7	6	3	2	1

Channel B

	3	6	7	8	7
	2	4	7	5	6
	1	1	3	4	5
	1	1	2	5	5
pixel 1, line 5 →	2	3	4	6	7

Channel C

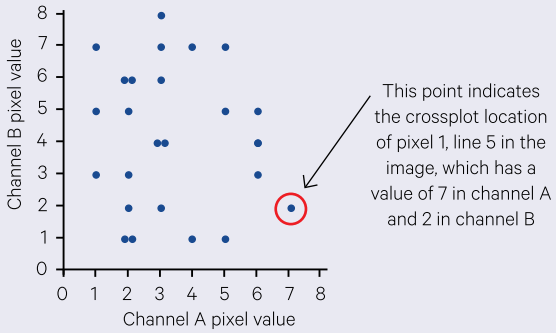
	1	3	4	3	3
	2	3	4	5	4
	2	2	3	6	5
	4	4	2	1	2
pixel 1, line 5 →	6	5	2	1	1

Source: Harrison and Jupp (1990) Figures 41

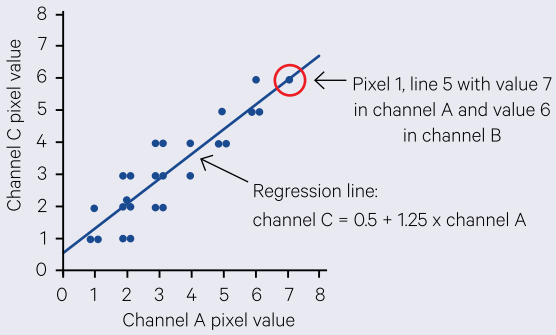
Figure 8.3 Example crossplots

The image channels in Figure 8.2 are crossplotted to show the extent of correlation between channels.

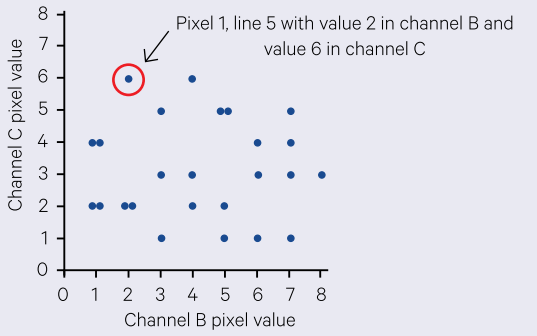
a. Channel B versus Channel A



b. Channel C versus Channel A



c. Channel B versus Channel C



Source: Harrison and Jupp (1990) Figures 42

The distribution of the points on the crossplots immediately indicates the similarity, or correlation, between the two plotted channels. Figure 8.3a shows little correlation between channels A and B, with the points being widely distributed around the graph. The value of a pixel in channel A is unlikely to indicate its value in channel B. However, Figure 8.3b shows a strong correlation between channels A and C. With this type of relationship, it would be possible to calculate the position of the regression line between the two channels and use the values of channel A to predict approximate values in channel C. Figure 8.3c again shows little correlation between channels B and C. This is not surprising considering that channels A and C are very similar and channels A and B are largely uncorrelated. In the case of channels A and C, the correlation is described as 'positive', since an increase in a value in one channel relates to an increase in the other channel. Channels can also be negatively correlated if there is a trend of increasing values in one channel being related to decreasing values in the other.

Table 8.1 details the calculations required to compute the covariance between channels A and B (see Section 8.1.4). When applied to all channel pairs, the covariance statistics for the example image channels are:

$$v_{AB} = -9.6 / 25 = -0.384$$

$$v_{AC} = 57.544 / 25 = 2.3$$

$$v_{BC} = -3.6 / 25 = -0.00144$$

Table 8.1 Covariance matrix calculation

Statistics for channels A and B are based on the example images shown in Figure 8.2.

x_A	$x_A - \overline{x}_A$	x_B	$x_B - \overline{x}_B$	$\left(x_A - \overline{x}_A\right) \times \left(x_B - \overline{x}_B\right)$
1	-2.3	3	-1.2	2.76
2	-1.3	6	1.8	-2.34
3	-0.3	7	2.8	-0.84
3	-0.3	8	3.8	-1.14
4	0.7	7	2.8	1.96
3	-0.3	2	-2.2	0.66
3	-0.3	4	-0.2	0.06
5	1.7	7	2.8	4.76
6	2.7	5	0.8	2.16
3	-0.3	6	1.8	-0.54
2	-1.3	1	-3.2	4.16
2	-1.3	1	-3.2	4.16
2	-1.3	3	-1.2	1.56
6	2.7	4	-0.2	-0.54
5	1.7	5	0.8	1.36
5	1.7	1	-3.2	-5.44
4	0.7	1	-3.2	-2.24
2	-1.3	2	-2.2	2.86
2	-1.3	5	0.8	-1.04
1	-2.3	5	0.8	-1.84
7	3.7	2	-2.2	-8.14
6	2.7	3	-1.2	-3.24
3	-0.3	4	-0.2	0.06
2	-1.3	6	1.8	-2.34
1	-2.3	7	2.8	-6.44
Total				-9.60

In this example, channels A and C have a high covariance value, which is consistent with the strong correlation visible in the crossplot of these two channels. Similarly, both A versus B and B versus C, which had no visible correlation in the crossplots, have low covariance values.

Accordingly, for our example image, the covariance matrix is:

Channel	A	B	C
A	2.938	-0.384	2.300
B	-0.384	4.720	-0.014
C	2.300	-0.014	2.266

and the correlation matrix is:

Channel	A	B	C
A	1.0	-0.10312	0.89148
B	-0.10312	1.0	-0.00404
C	0.89148	-0.00404	1.0

8.1.4 Covariance

The distribution of data values in a single channel could be represented pictorially by a channel histogram or mathematically by the variance statistic. Similarly, the correlation between two channels can be shown pictorially by a crossplot or represented mathematically by a statistic called covariance.

The covariance between two channels i and j (v_{ij}) is calculated as:

$$v_{ij} = \frac{\sum_{k=1}^{n}\left(x_{ik} - \overline{x}_i\right)\times\left(x_{jk} - \overline{x}_j\right)}{n}$$

where

n is the total number of pixels in the image;
x_{ik} is value for pixel k in channel i;
$\overline{x}_i$ is the mean of pixel values in channel i;
x_{jk} is value for pixel k in channel j; and
$\overline{x}_j$ is the mean of pixel values in channel j.

In most statistical analyses in image processing, the mean values are estimated from a sample of the image values so that a degree of freedom is lost in calculating the covariance. This equation is then usually implemented as:

$$v_{ij} = \frac{\sum_{k=1}^{n}\left(x_{ik} - \overline{x}_i\right)\times\left(x_{jk} - \overline{x}_j\right)}{n-1}$$

As mentioned before, any null pixels should be excluded from computation of image statistics. For covariance statistics, a pixel will be excluded if it has the null value in either channel i or channel j. A covariance matrix $\mathbf{C}_{n-1}$ based on n–1 samples, that is:

$$\frac{1}{n-1}\times\sum\left(\boldsymbol{x}_n - \overline{\boldsymbol{x}}_{n-1}\right)\times\left(\boldsymbol{x}_n - \overline{\boldsymbol{x}}_{n-1}\right)^T$$

can be updated using the equations:

$$\boldsymbol{C}_n = \frac{n-1}{n}\boldsymbol{S}_{n-1} = \left(n-1\right)\boldsymbol{e}_{n-1}\boldsymbol{e}_{n-1}{}^T$$

where

$$\boldsymbol{e}_{n-1} = \frac{\boldsymbol{x}_n - \overline{\boldsymbol{x}}_{n-1}}{n}$$

$$\overline{\boldsymbol{x}}_n = \overline{\boldsymbol{x}}_{n-1} + \boldsymbol{e}_{n-1}$$

and

$\overline{\boldsymbol{x}}_{n-1}$ is the vector of channel mean values for n–1 samples.

Negative covariance values indicate negative correlation between the channels; that is, high values in one channel are associated with low values in the other. Variance values, however, can never be negative.

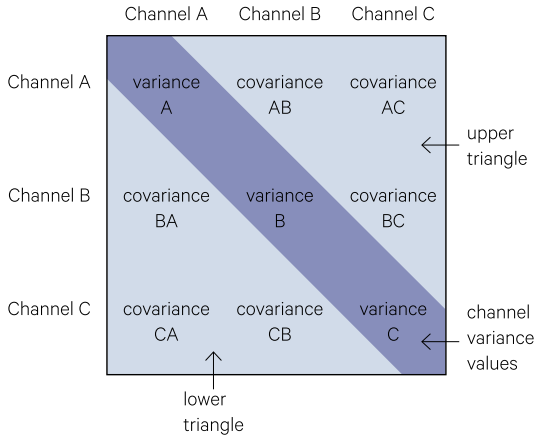
Large positive covariance values indicate a strong positive correlation between the two channels, while a large negative covariance would indicate that a strong negative correlation exists between the channels. The covariance between two identical channels (that is each pixel has the same value in both channels) would be equal to the variance of the channel. In a uniformly random image, that is, when every image value has equal probability of occurring at any pixel and with independent channels, the variance in each channel would approximate:

$$\frac{1}{12}\left(\text{maximum image value-minimum image value}\right)^2$$

and the covariance values should be close to zero. Thus, for a random image with values in the range 0–255, the variance would be approximately 5419.

To evaluate the variance within and between all channels in an image, these variance and covariance values are usually presented in a table of the form shown in Figure 8.4. This table is called a covariance matrix. Since the covariance AB is the same as BA, the matrix is always symmetric about the diagonal of variance values. As such it may also be written as a triangular matrix with either the upper or lower triangle omitted.

Figure 8.4 Format of variance/covariance matrix



Source: Harrison and Jupp (1990) Figure 43

The calculation of all values in the variance/covariance matrix can be summarised in the equation:

$$v_{ij} = \frac{\sum_{k=1}^{n}\left(x_{ik} - \overline{x}_i\right)\times\left(x_{jk} - \overline{x}_j\right)}{n}$$

for

$i = 1$ to m
$j = 1$ to m
$k = 1$ to n

where

m is the number of channels in image; and
n is the number of pixels in the image

or

$$v_{ij} = \frac{\sum_{k=1}^{n}\left(x_{ik} - \overline{x}_i\right)\times\left(x_{jk} - \overline{x}_j\right)}{n-1}$$

where

$\overline{x}_i$ and $\overline{x}_j$ are sample-based estimates of the real mean values.

Another representation of variance between channels is the correlation matrix. This is a normalised form of the covariance matrix that has values of 1 along the diagonal with all off-diagonal values being in the range -1 to 1. Elements of the correlation matrix are calculated from the covariance matrix using the formula:

$$c_{ij} = \frac{v_{ij}}{\sqrt{v_{ii}\times v_{jj}}}$$

where

c_{ij} is the correlation between channels i and j;
v_{ij} is the covariance between channels i and j;
v_{ii} is the variance of channel i; and
v_{jj} is the variance of channel j.

As with the covariance matrix, values close to 0 indicate little correlation between channels. A value of 0 would mean no correlation at all. The advantage of using a correlation matrix is to have the values normalised within a fixed range. In the covariance matrix 'large' positive or negative values indicate strong correlation between channels. In the correlation matrix we have a scale to clearly define 'large', with values close to 1 or -1 showing strong positive or negative correlation respectively. Diagonal values are always equal to 1 since each channel is completely correlated with itself. This would only occur for off-diagonal values if each pixel has the same values in both channels. This normalising allows correlation matrices to be directly compared between imagery, an operation that is not valid using covariance matrices.

Principal Component transformations completely remove the correlation between image channels, that is, create an image that has all zero values for the off-diagonal entries in its covariance matrix. Such computations may be based on either a covariance or correlation matrix (see Volume 2C).

8.2 Spatial

Section 8.1 discusses various statistics that may be derived from the spectral values of digital imagery to assist in its analysis and interpretation. (Here 'spectral' refers to both radiance and other measurable attributes of a pixel.) These spectral statistics are computed without any regard for the spatial arrangement of those values within the image. For example, Figure 8.5 illustrates two images with the same distribution of spectral values, and hence exactly the same spectral statistics, but displaying totally different spatial patterns.

Figure 8.5 Spectrally equivalent images with differing spatial patterns

Both images contain 12 pixels with value 1 and 13 pixels with value 5.

5	5	5	5	5
5	1	1	1	5
5	1	1	1	5
5	1	1	1	5
5	1	1	1	5

5	1	5	1	5
1	5	1	5	1
5	1	5	1	5
1	5	1	5	1
5	1	5	1	5

Source: Harrison and Jupp (1990) Figure 44

Visual interpretation of imagery is largely dependent on the delineation of spatial patterns. Such patterns may not be easily extracted from digital image processing despite being clearly visible to a human interpreter. However, various filtering and texture operators can be used to manipulate the spatial information in imagery (see Volume 2C).

Spatial statistics may also be derived from image values to indicate the image spatial structure. Various models and measures have been proposed to describe the spatial variation within an image (Ahuja and Schachter, 1983; Levine, 1985). For example, spatial statistics have been to forecast large-scale erosion and deposition patterns in arid landscapes (Pickup and Chewings, 1988).

Spatial structure and statistics of digital imagery are significantly influenced by many factors including the size, shape and density of objects in the original scene and the relationship between the size of the objects and the size of the image pixels (Woodcock and Strahler, 1987). Various aspects of image spatial resolution are discussed in Volume 1. Distinct objects tend to be spectrally homogeneous and separable from other objects and background. In this case, the extent to which adjacent pixels in the image are more spectrally similar than expected, that is, the degree to which they are 'spatially autocorrelated', is determined by the size of an image pixel relative to the objects being imaged. The mean distance at which the spatial autocorrelation disappears is related to the size, spacing and shape of the objects being imaged.

The nature and causes of spatial variation in remotely sensed imagery are investigated by Woodcock *et al.* (1988a, 1988b) using the variogram measure with simulated and real data. Jupp (1988b, 1988a) provide a mathematical framework for modelling the effect of image pixel size on spatial image covariance then apply these results to show that the spatial structure of an image depends on the underlying spatial structure of the scene that it represents.

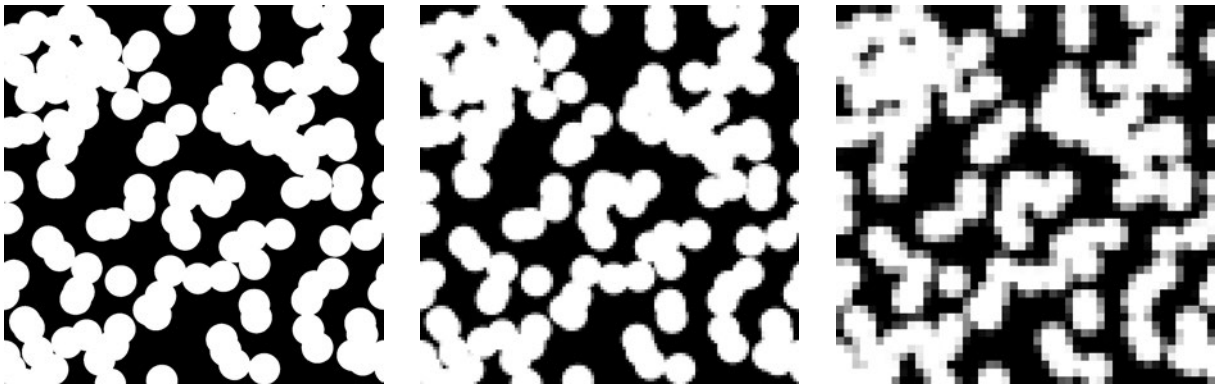
These authors have defined two classes of image spatial resolution that relate image pixel size to object size in the original scene. The H-resolution (high) case occurs when the image pixel size is smaller than the object size so that many pixels will represent the object. For example, a large agricultural field imaged by Landsat OLI would contain many pixels. Similarly, a sugar cane plantation, with fields spanning 200 m by 100 m, imaged by an airborne camera system would contain many pixels per field. When such fields contain a homogeneous land cover, adjacent pixels within the field would be spectrally similar, that is, spatially autocorrelated. The L-resolution (low) case occurs when the image pixel size is larger than the object size so each pixel includes several objects. This case exists in Landsat OLI scale imagery (with 30 m pixels) of urban areas.

Figure 8.6 shows six artificial images of a woodland with decreasing spatial resolution. If a tree crown is the object of interest, Figure 8.6a to Figure 8.6e are examples of the H-resolution case since the pixel size is smaller than the tree crown size, while Figure 8.6f is L-resolution. Despite its 'blockiness', the distribution of individual crowns is still discernible in Figure 8.6e.

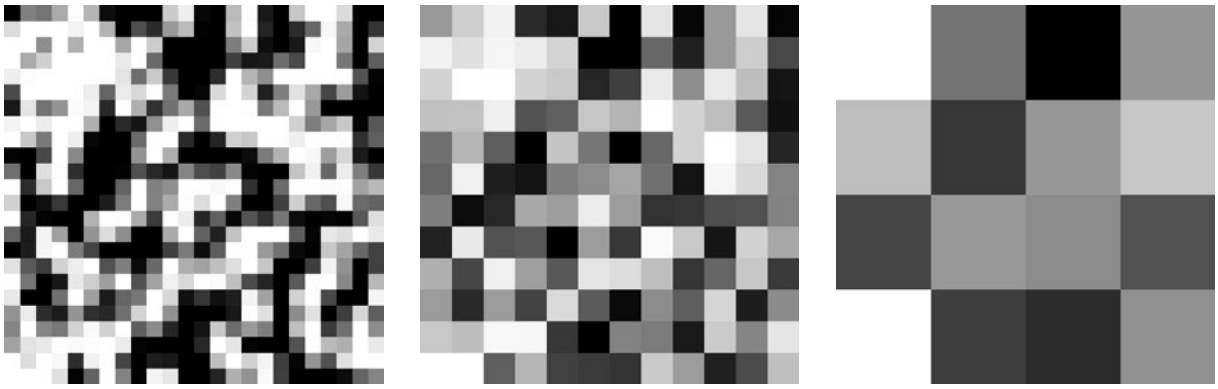
Figure 8.6 Impact of pixel size on feature discernment

In this synthetic image, circles represent tree crowns of 10 m diameter on a contrasting background. The ground area of each image is approximately equal to 120 m x 120 m, that is, the area of 4 x 4 Landsat TM/ETM+/OLI multispectral pixels.

a. 10 cm pixel b. 1 m pixel c. 2.5 m pixel



d. 5 m pixel e. 10 m pixel f. 30 m pixel



Source: David Jupp, CSIRO

The differences between these resolution classes and the properties within them can be determined using spatial statistics. The commonly accepted statistics for analysing image spatial structure and variation are:

- local variance (see Section 8.2.1);
- spatial covariance (see Section 8.2.2); and
- autocorrelation and variograms (see Section 8.2.3).

8.2.1 Local variance

Local variance may be computed for each pixel and averaged over the whole image. If the local neighbourhood of a pixel is defined as those pixels that are adjacent to it, then the values of the pixel and its eight neighbours can be used to calculate a local variance statistic for that pixel. The average of the local variance values over the whole image indicate the extent to which adjacent pixels have similar spectral values in a particular channel.

The effect of changes in pixel resolution on image spatial structure can be determined by computing the local variance values for a set of images with increasing pixel size. A larger pixel size may be simulated for an image using the image blocking technique described in Section 7.2.1.3. The effect of this process is illustrated in Figure 8.6 above.

The average local variance statistics (namely, the mean, standard deviation, minimum, maximum and pixel count) can be computed using a 3×3 pixel neighbourhood centred on every pixel in the image (see Volume 2C for more details of this type of 'filter-based' processing). The local standard deviation (square root of the local variance) is also referred to as the mean 'texture' value and is calculated as:

$$\text{Texture} = \sqrt{\frac{1}{2} \times \frac{\sum \left(x_{ij} - x_{kl} \right)^2}{n}}$$

where

x_{ij} is the central image value at column i and row j;
x_{kl} is the pixel value at column k and row l, where $k = I \pm 1$ and $l = j \pm 1$; and
n is the number of pixels involved in the calculation—usually eight but sometimes less if there are null pixels within the 3×3 window.

Where *n* equals 8, this equation becomes:

$$\sqrt{\frac{\sum\left(x_{ij}-x_{kl}\right)^2}{16}}$$

These spatial statistics may be computed iteratively with image blocking being used in simulate larger pixels between iterations. It is useful to graphically summarise the texture statistics from this iterative sequence by plotting pixel resolution versus local variance as illustrated in Excursus 8.2. The major repeating patterns in the image have been found to occur at a resolution of between two-thirds and three-quarters of the pixel size in the image with maximum variance. This technique has application for determination of the optimal pixel size for studying specific features (Woodcock, 1985).

These statistics may also be used to estimate the values for local pixel-to-pixel uncorrelated noise for each channel. Such noise estimates may then be used as weights in a Principal Components Analysis (see Volume 2C). The local spatial variance can also be used to determine threshold values for image despiking (see Volume 2C) by defining the 'normal' level of local variation in image values. For this purpose, the local standard deviation would be normalised (by dividing by the channel mean) then multiplied by 100 to represent the percentage variation equivalent to one standard deviation. The threshold level would then be some multiple of this to represent an appropriate number of standard deviation units (see Volume 2C).

8.2.2 Variograms

The variogram of an image is the expected squared grey-level difference between two pixels which are a given distance and direction apart (Ahuja and Schachter, 1983). This measure may be estimated using a large number of randomly-selected pixel pairs in the same row or pixel column in the image. The variogram statistic is usually computed for a range of distances, or lags, between the pixel pairs. The variogram is calculated as:

$$\frac{\sum\left(x_i-x_{i+j}\right)^2}{n}$$

where

x_i is the pixel value at column *i*;
x_{i+j} is the pixel value at column *i+j*;
j is the lag distance in pixel columns along an image line; and
n is the number of pixel pairs involved in the calculation.

A similar statistic for quantifying spatial variation within an image is *spatial covariance*. This is calculated as:

$$\frac{\sum\left(\left(x_i-x\right)\times\left(x_{i+j}-\overline{x}\right)\right)}{n}$$

where

$\overline{x}$ is the mean value in the image channel;
x_i is the pixel value at column *i;*
x_{i+j} is the pixel value at column *i+j;*
j is the lag distance in pixel columns along an image line; and
n is the number of pixel pairs involved in the calculation. (Again, n–1 is often used instead of n when the mean value used for channel *x* is only an estimate rather than the real mean value.)

This formula assumes that the mean of the channel is spatially invariant (or 'stationary'). In practical image terms, this would imply that the local mean and variance do not vary systematically over an image. For remotely sensed imagery, variable illumination and reflectance geometry generally invalidate this assumption. The variogram provides a more relaxed measure of spatial variance and has been commonly used in geostatistics for this purpose. When the data are stationary the variogram is related to spatial covariance as:

$$V_j = 2\left(Cov_0 - Cov_j\right)$$

where

j is the distance lag between pixels;
V_j is the variogram for distance lag *j* over the image;
Cov_0 is the spatial covariance for no lag, that is, the variance of the channel; and
Cov_j is the spatial covariance for a distance lag *j* over the image.

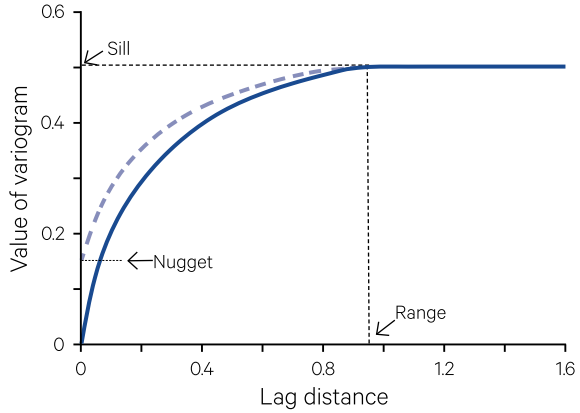
As the name implies, the semi-variogram equals half the variogram:

$$SV = Cov_0 - Cov_j$$

8.2.3 Autocorrelation and variograms

Spatial autocorrelation is another measure of spatial variation. It is computed as the covariance between pixels in a channel divided by the channel variance. These results can be displayed graphically as a 'correlogram' by plotting autocorrelation value versus distance (Cliff and Ord, 1981). This statistic however also assumes a stationary mean value over the channel so it is an inappropriate measure for remotely sensed imagery.

Figure 8.7 One-dimensional variogram



Source: Harrison and Jupp (1990) Figure 46 [Adapted from: Jupp *et al.* (1988a)]

The squared grey-level differences can be computed for pairs of pixels which are a range of distances apart along the same scan line for each line in the image (or set of images). In image processing systems, the distance is often specified as a lag factor, with multiple lags being computed as required. As the length of the lag determines the number of pixel pairs which will be included in the calculation for a given image width, it is advised that the lag distance should be no greater than one tenth of the image line length to retain sufficient confidence in the resulting statistics. These results are best presented graphically by plotting variogram value versus lag distance as shown in Figure 8.7.

Jupp *et al.* (1988a) report there are five especially significant parameters that describe the one-dimensional variogram:

- The 'sill' at which the variogram flattens off to a level equal to twice the regional data variance. (Variogram data are sometimes presented as the 'semi-variogram'—variogram/2—so that the sill height equals the regional variance). The existence and size of the sill are important parameters. The lack of a sill could indicate a trend in the data (such as a scan-angle brightness effect across the image) and/or that the pixel spatial resolution is too small for the scales of pattern in the image (that is, the L-resolution case described above).
- The 'range' value of lag *j* at which the variogram reaches the sill. This parameter is often related to the size or scale of the largest elements (objects) in the scene that produce the correlation structure.
- The 'nugget effect', which, if present, is expressed as a finite limit for the variogram as the lag *j* tends to zero. The variogram value is numerically zero at lag 0, however uncorrelated noise in the data always produces variance. For an image of an uncorrelated noise series, the variogram will have no range and the nugget effect will equal the sill.
- The 'derivative at the origin' is diagnostic of the essential variability of the data at the finest scale. If the limit of the derivative at the origin as lag *j* tends to zero is near zero, the underlying process will be 'smooth'. If the derivative limit is high, the process will be 'rough'. Jupp *et al.* (1988a) have interpreted this parameter as the amount of edge and boundary at the finest scale.
- The 'anisotropy' of the variogram appears if the image is defined in more than one dimension. It reveals a covariance structure such that values tend to be more similar in one or more preferred directions than in others. This will commonly arise in a remotely sensed image when the Sun is not overhead and three-dimensional objects in the ground scene cast shadows (dark shapes) that are elongated in one direction.

Jupp *et al.* (1988b) used an artificial image of a woodland scene to demonstrate how the covariance and variogram statistics are modified. Woodcock *et al.* (1988b) summarise their interpretations of variograms for remotely sensed imagery as:

- The sill is related to the proportion of the area covered by objects, which is a function of their number or density;
- The range is related to the size of objects in the scene. The shape and range of the variogram are more closely related to the area of objects than to their shape;
- The shape of a variogram is related to the variance of the size of objects in the scene. A more rounded or gradually sloping shape is characteristic of higher variance in the size of objects; and
- Increasing the image pixel size has the following effects on variograms:
  - the height of the sill is reduced;
  - the range of influence is increased; and
  - the height of the variogram at the distance equal to one pixel unit increases relative to the sill.

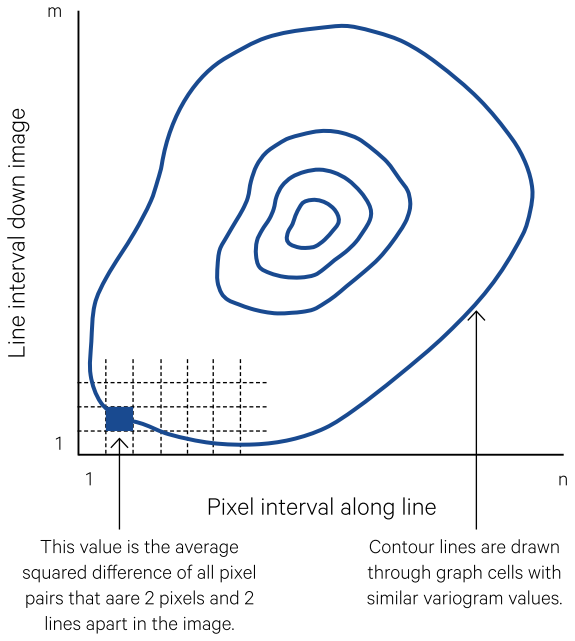
Curran (1988) discusses the use of the semi-variogram (*variogram/2*) in remote sensing analyses and gives examples of computing the spatial variance statistic along transects within specific features. Curran also suggests how spatial statistics can be used to determine an appropriate spatial resolution (for remotely sensed pixels and ground sample plots) for different image features and the number of sample points required to achieve a given confidence level. Spatial statistics of individual image features (such as an agricultural field) can also be determined after isolating the feature using subsetting (see Section 7.2.1.1) and/or segmentation techniques (see Section 10) and possibly re-orienting the resulting image to align image lines with the feature (see Section 7.3 and Volume 2C).

Woodcock *et al.* (1988a) report that one-dimensional variograms are easier to display and interpret, although two-and higher-dimensional displays may also be produced. Variogram statistics computed along an image line are one-dimensional. However, image data could be rotated or transposed to determine these statistics in a different direction (see Section 7.2.2). A two-dimensional variogram is usually displayed as a contour plot as shown in Figure 8.8 and can be useful for revealing anisotropy in data. This plot is constructed by entering the mean squared difference of all pixels pairs that were a particular distance apart in pixels and lines. Contour lines are then drawn between cells with similar values. The interval at which the contour structure breaks down indicates the size of the major repeating patterns in the image (Woodcock, 1985).

Spatial statistics offer an important tool for analysing the scale of an image relative to the objects it represents. The 'optimum' pixel size for image data obviously depends on the cover types being investigated. A finer pixel size is not always preferable since the additional information may simply be increasing the variation within a feature rather than improving its discrimination from other objects or the background. For a woodland area for example, a 2 m pixel would sub-sample individual tree crowns and report changes in density within each crown. If the object of the study is to map woodland as one homogeneous category, this detail will simply complicate the task. A larger pixel size however will integrate multiple crowns into a single pixel size and thereby highlight changes in density of tree cover rather than leaf cover. This integration is most effective for woodland areas when the pixel size is at least 50 m. Thus coarser scale imagery can often resolve broad area composite features such as woodland more effectively than finer scale data.

Figure 8.8 Two-dimensional variogram

The size of the major repeating pattern in the image corresponds to the point at which the contour structure fails.



Source: Harrison and Jupp (1990) Figure 48

The choice of 'optimal' scale for remotely sensed imagery (disregarding the logistical questions of cost and availability) can be summarised by two questions relating to the object of interest:

- Is the feature of interest causing variance between adjacent pixels? If not, is the feature too large or too small relative to the pixel size?
- At what scale does spectral variance in an image relate principally to the object of interest?

The spatial analyses described in this Section attempt to provide answers to these questions. The ability to 'see' features in imagery by visual interpretation often confuses the issue of an appropriate scale for digital data. The type of analyses undertaken in digital image processing requires different criteria for the selection of optimal image resolution for a particular application.

Excursus 8.2—Spatial Statistics

Source: David Jupp, CSIRO

Three example images are illustrated in Figure 8.9. These images were artificially created to represent an area of 80 m square with 50% cover (that is, half the area is covered by a single 'type' of object and the other half is background). The object values are distributed to represent three scales of pattern:

- disks of 8 m diameter;
- disks of 1 m diameter; and
- random distribution.

The centres of the disks are randomly (or Poisson) distributed. The random distribution of disks is essentially objects with diameter a little smaller than a pixel.

In these images, every pixel either has value 0 (background) or 1 (disk) and the proportion of disk is simple the image mean. In general, if the cover of disks is q, then the image mean is q for all cases (random, 1 m or 8 m disks), and the image variance is:

$$\begin{aligned} \text{var} &= \frac{1}{N}\sum_{j=1}^{N}(x_j - q)^2 \\ &= q - q^2 \\ &= q(1-q) \end{aligned}$$

Since the image standard deviation is the square root of the image variance, for 50% cover, q = 0.5, var = 0.25 and standard deviation = 0.5. As occurs with the mean, the standard deviation is the same for all three cases. So the overall mean and standard deviation do not respond at all to the different disk sizes.

In an EO image of a ground scene, the disks could be considered to be representative of tree crowns so that the large and small disks would be like simplified images of woodland vegetation with big and little trees respectively. Each image is 512 pixels by 512 lines, and as it covers an area of 80 m by 80 m, the pixel size is equivalent to 15.625 cm.

The images were blocked four ways to produce images at four new resolutions:

- 2×2: 31.25 cm pixel;
- 4×4: 62.5 cm pixel;
- 8×8: 125 cm pixel; and
- 16×16: 250 cm pixel.

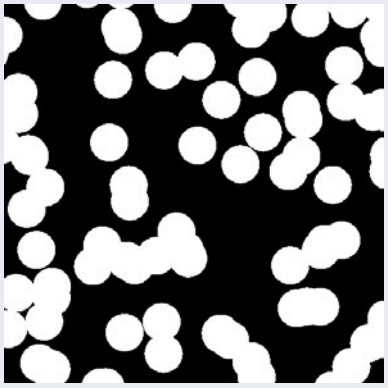
Then, local variance statistics were computed for the five scales (original plus four blockings listed above).

The mean local standard deviation, or texture, can be computed as the average of the square root of local variance (as detailed in Section 8.2.1 above). Figure 8.10 represents the results of the average function graphically. For the large tree image, the maximum pixel size (250 cm) is much smaller than the object size (8 m) so the variance is still increasing at a blocking factor of 16. However, in the small tree image the variance peaks at a pixel size around 62.5 cm, which is between one half and three-quarters of the object size (1 m). The object size in the random image is effectively one pixel, so the variance decreases rapidly with blocking. Unlike the overall mean and variance, the three disk models therefore have very different behaviour with block variance.

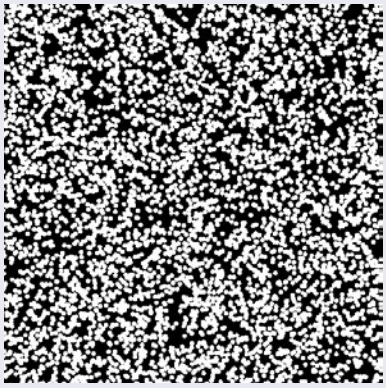
Figure 8.9 Artificial example images

These three artificial images comprise 512 x 512 square pixels and cover an area of 80 m square. Percent cover within each image is 50%.

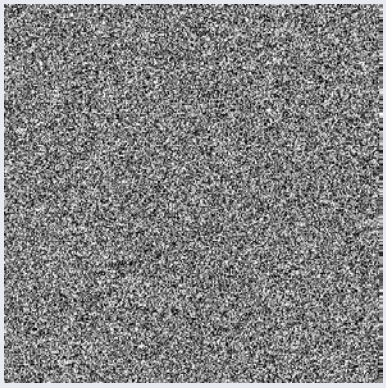
a. Large disks (8 m diameter)



b. Small disks (1 m diameter)



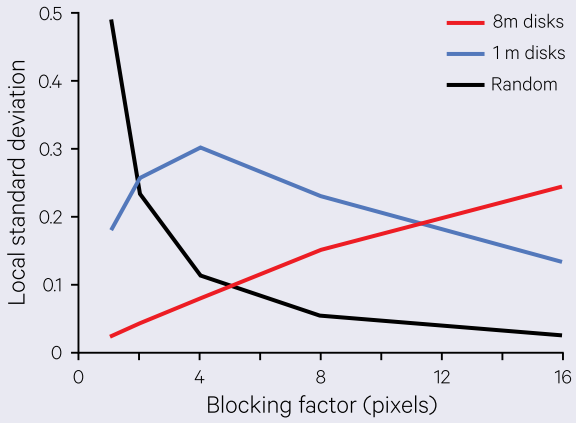
c. Random image



Source: David Jupp, CSIRO

Figure 8.10 Local standard deviation results derived using average function

Object size for large disks (see Figure 8.9) is much larger than the pixel size so variance continues to increase over the distance range used in this example. With the small disk image, variance peaks at a distance approximately equivalent to two-thirds of the object size. The pixel size effectively equals object size in a random image so variance only decreases.



Source: Harrison and Jupp (1990) Figure 45

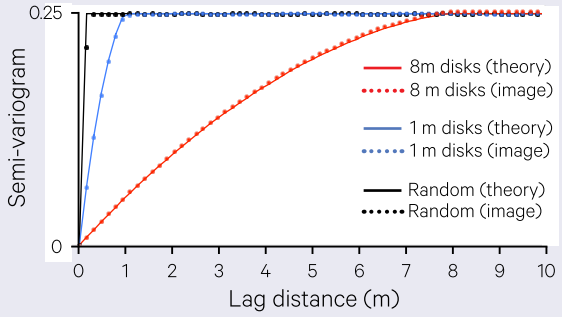
The semi-variogram statistics also show different behaviours for the three disk models. The three artificial images in Figure 8.9 were processed with lags from 1 pixel up to 52 pixels (or 8.125 m). Theoretical model results (Jupp *et al.*, 1989) for the equivalent continuous functions were also computed.

These results are plotted in Figure 8.11 and show good agreement. In this example, given that the cover fraction (q) is 50%, the sill of the variogram occurs at value 0.25 or, as shown above, $q(1-q)$.

Accordingly to theory (Jupp *et al.*, 1989), the range values correspond to disk size, with the curve for the 1 m disk image reaching the sill at a lag distance of 1 m and the curve for the 8 m disk image reaching the sill at a lag distance of 8 m. The random image has a much narrower range as it is effectively a disk of approximately diameter 15 cm (one image pixel).

Figure 8.11 Variogram results for example images

Semi-variogram values are plotted against lag distances to indicate spatial variance of three images with different object sizes and distributions (see Figure 8.9). The curves derived from images with 8 m and 1 m disks match theoretical results.



Source: David Jupp, CSIRO

8.3 Radiometric

In the context of EO imagery, the term 'radiometric statistics' is sometimes used to refer to variations in radiance values resulting from detector inconsistency in both image data values and calibration measurements (see Section 3.2.3). In this Section, we use the term to embrace all variations in EO imagery that distort true surface reflectance measurements (see Section 3.3.2).

The process of EO image acquisition is detailed in Volume 1. This process involves an EO sensor, carried on a travelling platform, observing the undulating surface of the Earth through a changing atmosphere, with a moving illumination source. While the goal of EO is to quantify surface reflectance, the actual process of image acquisition invariably results in measurements that include factors related to atmospheric composition, surface illumination and instrument operation. The major factors that interfere with measurement of surface reflectance during EO image acquisition can be summarised as:

- EMR interactions with target and sensor (see Section 8.3.1);
- EMR interactions with the atmosphere (see Section 8.3.2); and
- sensor operation (see Section 8.3.3).

A wide variety of approaches are used to correct EO imagery for these factors. While most implementations of these methods attempt to correct for all of the major factors in a single, integrated model, we discuss the three major factors individually below. An example of radiometric calibration applied to a continental scale dataset is given in Excursus 8.3.

8.3.1 EMR interactions with target and sensor

As discussed in Volume 1B—Section 3, the radiance measured by EO sensors is impacted by several interactions that can occur between the incident radiance and the imaging target. These interactions also vary with wavelengths of the EMR energy, the texture and type of surface materials, and the three-dimensional relationship between illumination source, target and sensor. Cumulatively, these interactions result in quantifiable variations in 'scene brightness' in EO imagery (see Appendix 1).

In this context, 'correction' of scene brightness variations changes the spectrum of values for each image pixel to what would have been measured by the EO sensor if the geometry of the illumination source and sensor were 'fixed', rather than varying over the image extent. Two corrections that are often applied to EO imagery attempt to compensate for:

- surface orientation (see Volume 1B—Section 3.5); and
- anisotropy (the differences in geometrical reflective properties of targets when they are measured in different directions; see Volume 1B—Section 3.6).

Surface orientation is generally related to surface topography, but can also vary with sensor view angle. Terrain illumination correction uses geometrical parameters derived from accurately registered Digital Surface Models (DSM) to modify topographic shading in imagery (Li *et al.*, 2012). These algorithms are limited both by the resolving power of the EO data with respect to hill slopes and the accuracy of the elevation data (in all dimensions).

Anisotropy can be quantified by the Bidirectional Reflectance Distribution Function (BRDF; Nicodemus *et al.*, 1977). BRDF correction reduces the differences due to varying Sun and sensor positions in EO images, and can be applied to the variation within a single image and across a set of images. Some of the models that have been used for BRDF correction are detailed in Appendix 2. Accuracy of these models depends on the spatial resolution of the available calibration data.

The three main approaches to modelling scene brightness, namely empirical models, measurement models and semi-empirical models, are detailed in Appendix 1.

8.3.2 EMR interactions with the atmosphere

Radiance detected by EO sensors includes effects due to EMR scattering and absorption by atmospheric gases, aerosols and/or water vapour (see Section 3.2.3 above and Volume 1B—Section 4). A range of methods are used to correct for atmospheric path radiance in EO imagery, with variable success. Commonly used methods include:

- subtract 'dark values'—determining the image value corresponding to a very dark object with virtually no discernible radiance, such as a deeply shaded hillside or valley. Any image values below the dark value are assumed to represent atmospheric radiance alone. The image data can then be rescaled to only contain values related to surface reflectance (see Volumes 2C and 2D for details).
- atmospheric normalisation—between two or more images can be based on a relative or standardised reference (Coppin *et al.*, 2004). Some methods that have been proposed for atmospheric normalisation of EO image sequences include:
  - simple regression (SR)—matching reflectance values for the full image scene to a known reference using spectral channel crossplots, then use the regression equation to rescale image values (Jensen, 1983; see Volume 2C);
  - pseudo-invariant feature (PIF)—selects landscape features with near-constant reflectance through time and rescales image ranges so that all features have comparable values in all images (Hall *et al.*, 1991a, Furby and Campbell, 2001). An automated version of this approach called Automatic Scattergram-Controlled Regression (ASCR) uses scattergrams (or channel crossplots; see Section 8.1) to select pixels with near-constant reflectance (Elvidge *et al.*, 1995); and
  - radiative transfer models (RT)—simulate the impact of the atmosphere on reflected or emitted radiance using parameters derived from imagery and field data (Ahern *et al.*, 1988; see Volume 1B—Section 5 and Appendix 1).

Radiative transfer models such as MODTRAN (Berk *et al.*, 2003) account for Mie, Rayleigh and aerosol scattering effects and atmospheric gas absorption using estimates of atmospheric optical depth (visibility), water vapour, and concentrations of major gases (such as ozone and carbon dioxide). Such models estimate a range of optical parameters, such as transmittance (for Sun/sensor directions), path radiance, atmospheric albedo, and the proportion of diffuse to total irradiance (see Appendix 1). Their accuracy is strongly dependent on the quality of available aerosol data to determine the atmospheric profile at the time of image acquisition. Such methods are commonly used to correct hyperspectral imagery where estimates of atmospheric water vapour can be derived from image spectra.

8.3.3 Sensor operation

Imaging sensors can introduce a range of radiometric distortions as reviewed in Section 3.2.3. These sensor-related distortions can be:

- random—such as missing or spurious pixels or lines;
- systematic—such as multiple detectors recording inconsistent measurements; or
- gradual—such as detector sensitivity deteriorating with time.

The impact of such distortions is removed or reduced by the process of radiometric calibration (see Section 3.3.2). EO imagery can be radiometrically calibrated to either an absolute or relative standard.

Absolute calibration relates image values directly to a known reference and enables pixel observations to be converted into radiance measurements. Such methods require precisely calibrated ground reference points which, ideally, are simultaneously measured during image acquisition, and allow removal of calibration differences between detectors and channels. A range of methods have been proposed for absolute calibration of various sensors, such as Teillet *et al.* (1990) for NOAA AVHRR, Slater *et al.* (1996) for EOS sensors, and Teillet *et al.* (2001) using hyperspectral data for multiple sensors.

Relative calibration simply attempts to equalise the variation between multiple detectors. For example, whiskbroom sensors operate by recording multiple lines with different detectors in each imaging scan (see Volume 1A—Section 14). While efforts are made to calibrate these detectors, sensitivity differences still exist and result in a characteristic multi-line, horizontal striping pattern in the imagery (see Section 3.2.3 for details). Similarly, pushbroom scanners can introduce vertical striping patterns into imaged scenes when individual detectors along a scan line deliver different radiance values for the same target reflectance (see Figure 3.6).

One relative method for 'destriping' images with horizontal or vertical striping involves equalising the mean and variance values for the set of detectors. If the mean and variance values of each detectors were accumulated separately, then the image could be rescaled so each detector has a mean and variance equal to either the average values for the whole image or to a reference sensor. A similar method, known as 'histogram equalisation', matches the histograms of each detector to a selected reference histogram (such as the mean). These processes are further discussed in Volume 2C.

Excursus 8.3—Continental Scale Calibration of ASTER imagery

Source: Michael Caccetta, CSIRO
Further Information: Caccetta *et al.* (2013)

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), launched in 1999, acquired global imagery with particular value for geological applications. Over Australia, this imagery enabled continental scale mapping of surface mineral abundance and composition (see Volume 1B—Excursus 6.1).

3,600 scenes of ASTER L1B radiance at sensor data, acquired between May 2000 and March 2008, were converted into a single, calibrated mosaic of apparent surface reflectance by correcting each scene for effects due to both instrument characteristics and atmospheric/illumination characteristics. This extensive dataset included approximately 400 different dates with significant radiometric variation occurring between adjacent scenes (see Figure 8.12a, Figure 8.12c and FFigure 8.12e). To generate a calibrated reflectance dataset, processing required use of:

- new statistical cross-calibration procedures, which compensated for between-scene variations, to create a single mosaic; and
- an extensive set of independent, calibrated reflectance imagery (Hyperion satellite imagery) to convert the mosaic to apparent reflectance (Caccetta *et al.*, 2013).

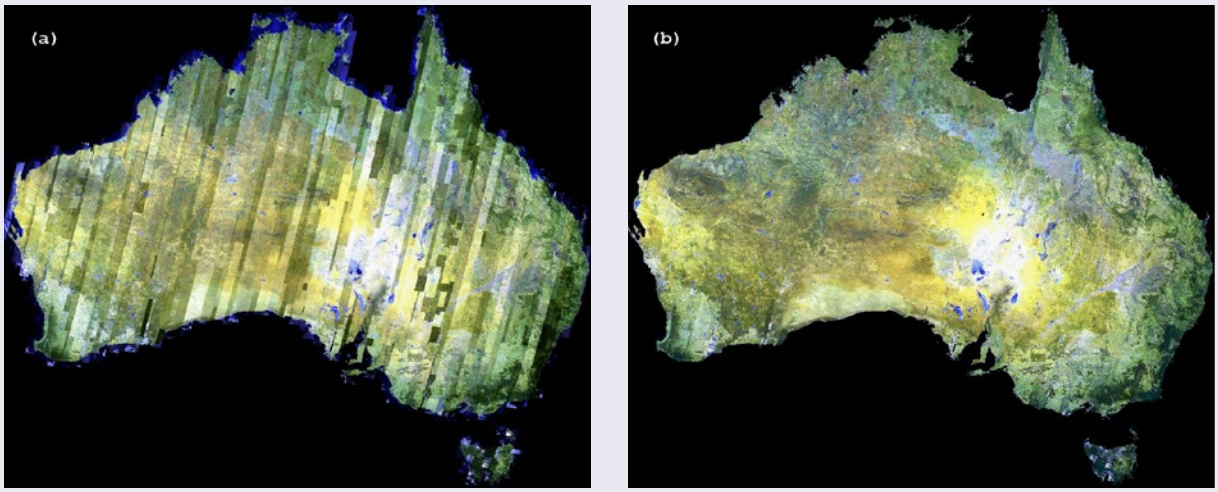
For ASTER optical bands, the seamless, continental mosaic resulting from the cross-calibration process is illustrated in Figure 8.12b, Figure 8.12d and Figure 8.12f. This cross-calibrated mosaic was then converted to apparent reflectance by determining a linear transformation based on robust least squares regression between the ASTER mosaic and a set of 18 co-located Hyperion scenes. A similar approach was used to generate a continental emissivity mosaic from ASTER L2 surface emissivity data.

From these two continental mosaics (reflectance and emissivity), a suite of geoscience information products have been derived, primarily indicating mineral group content and composition (for example, see Volume 1B—Excursus 6.1). These geoscience maps are available from: *http://doi.org/10.4225/08/51400D6F7B335*.

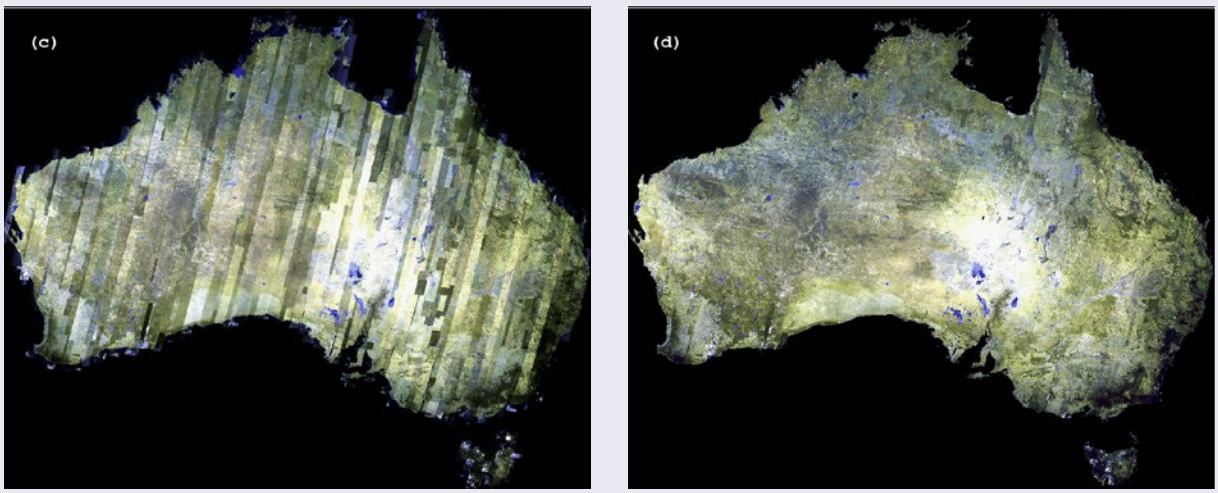
Figure 8.12 Continental mosaic of ASTER imagery

3,600 ASTER L1B scenes (a, c and e) were cross-calibrated to create a seamless continental mosaic of apparent reflectance (b, d and f).

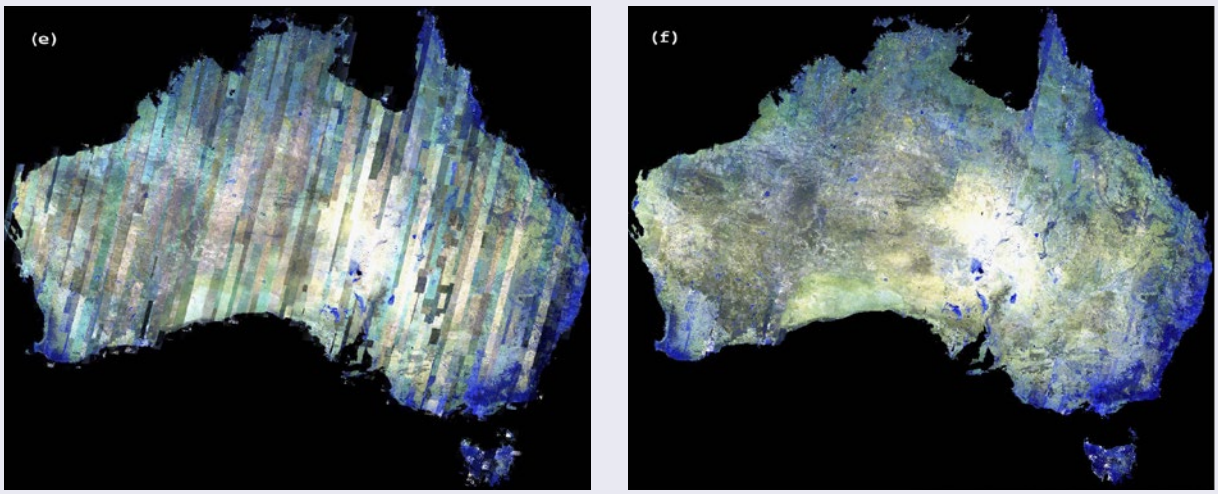
a. and b. ASTER bands 7 (SWIR), 4 (SWIR), 1 (green) shown as RGB



c. and d.ASTER bands 8 (SWIR), 5 (SWIR), 2 (red) shown as RGB



e. and f. ASTER bands 9 (SWIR), 6 (SWIR), 3 (NIR) shown as RGB



Source: Michael Caccetta, CSIRO

8.4 Temporal

Time series datasets differ from many other sets of measurements in that they are assumed to represent a non-random sequence. By definition, a time series dataset comprises consistent measurements taken at equally spaced time intervals. Image time series, then, represent a sequence of two (or more)-dimensional measurement layers, or multivariate time series, whose values are consistent in both space and time. For example, Figure 8.13 shows a temporal profile plot of mean chlorophyll concentration near the mouth of the River Murray, as estimated from MODIS imagery.

Most time series analyses aim to identify those factor(s) that drive changes in the data series measurements and, based on this understanding, predict future values in the series. Each of these goals requires that the underlying pattern(s) in the data be described precisely. The identified patterns can then be related to their drivers and extrapolated to future scenarios.

It is assumed that the measurements in time series datasets include both systematic pattern (often resulting from identifiable components) and random noise (or error). The latter generally obscures the systematic pattern to some extent and a range of filtering or screening algorithms are commonly used to clarify the systematic data content. The underlying systematic pattern results from both seasonal (repeating) trends and longer term trends. Thus, where the time series is the sum of these three components, it can be represented by the equation:

$$X_t = m_t + s_t + Y_t$$

where

X_t is the time series;
m_t is the long term trend;
s_t is the seasonal trend; and
Y_t is random noise.

In some time series, a multiplicative model may be more appropriate to represent the interactions between components:

$$X_t = m_t \times s_t \times Y_t$$

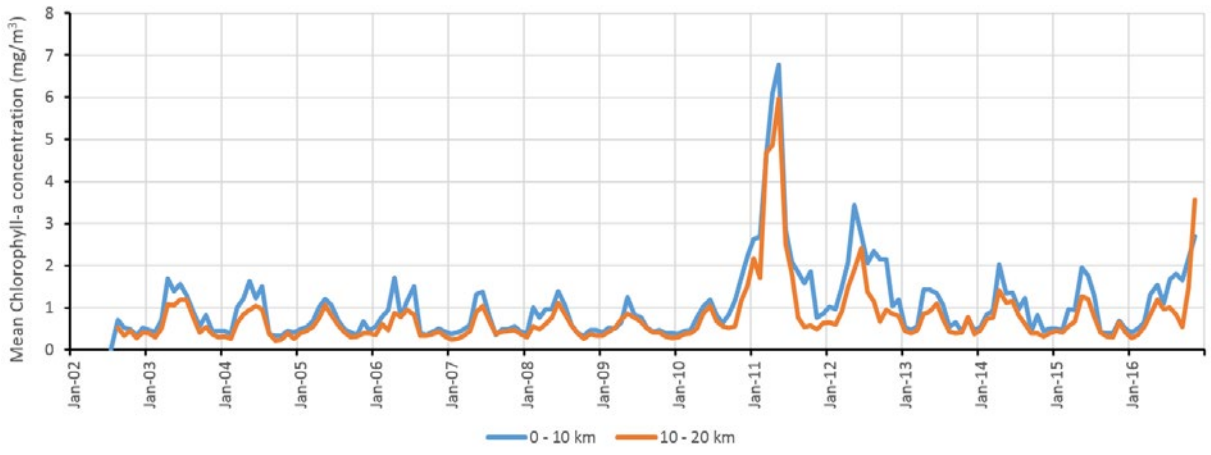
In such cases, logarithmic transformations are used to convert the data into a linear space so that an additive model can be fitted (see Volume 2C).

Where the number of equally spaced observations in a time series is infinite, it can be viewed as a stochastic process for statistical purposes. A time series is considered stationary when its statistical properties are time-invariant, that is the mean, variance and covariance are constant. The goal of time series analyses is to extract the long term (m_t) and seasonal trends (s_t) so that the remaining noise component (Y_t) is stationary.

Many time series datasets demonstrate a seasonal trend based on annual cycles (whether driven by climatic or economic events) and a longer term trend that carries the mean of the seasonal trend up or down over time. There are many examples of time series datasets in which the general shape and duration of the seasonal shape is discernible from year to year, but the amplitude increases (or decreases) with the longer-term trend (see Figure 8.14).

Figure 8.13 Trends in chlorophyll-a concentration

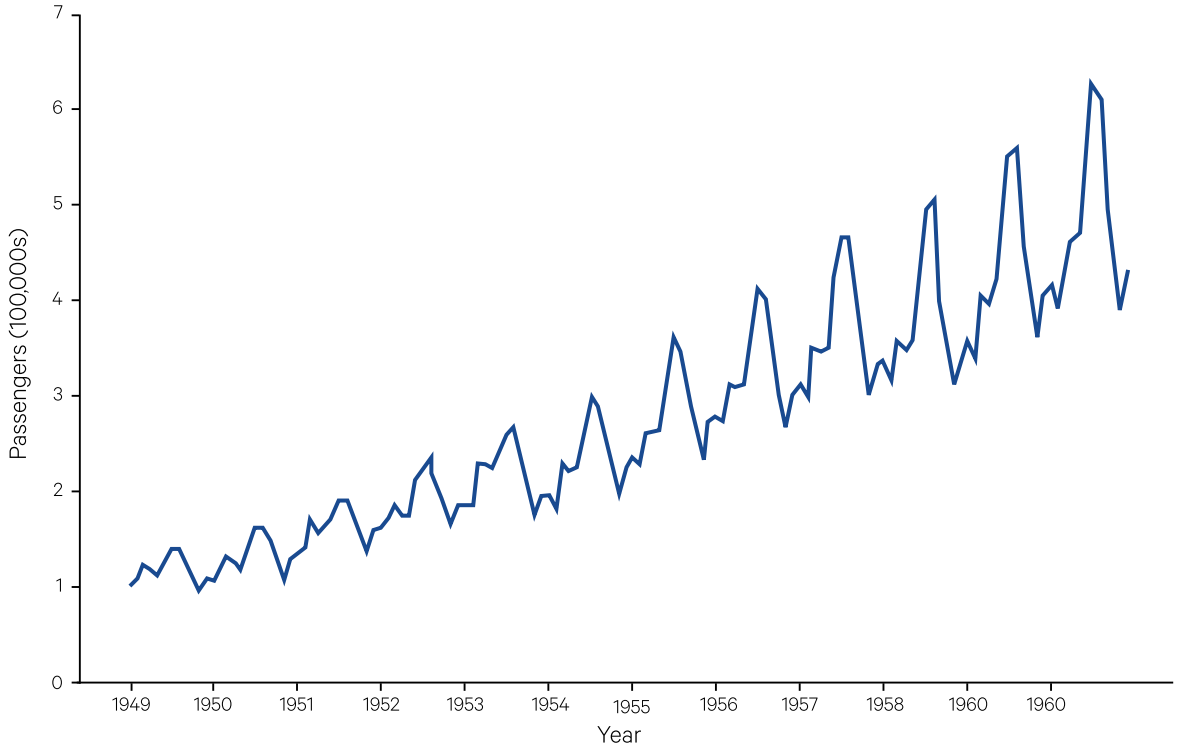
Temporal profile of ocean chlorophyll-a concentration near the mouth of the River Murray, in Australia, as measured from MODIS Aqua Level 3 chlorophyll a data product. The graph shows that high (2011) and normal (2012) river flows result in substantial increases in ocean chlorophyll a concentrations at least 20 km from shore, as compared to the concentrations observed in very low flow years (2002–2010).



Source: Ken Clarke, University of Adelaide (adapted from data underlying Figure 8 in Auricht *et al.*, 2017)

Figure 8.14 Time series example

This example, based on international airline passenger numbers from 1949 to 1960 inclusive, shows both seasonal trends with each year and a longer term over the 12 year period.



Adapted from: *http://www.statsoft.com/Textbook/Time-Series-Analysis#identifying* [Derived from Box and Jenkins (1976)]

Trend analysis requires some form of local averaging of the dataset to eliminate non-systematic components, such as median smoothing (see Volume 2C), before representing the trend as a mathematical function. Standard time series methods determine the trend first, then use that relationship to focus on seasonal perturbations. Time series models can be assessed by the randomness of their error distribution and/or predictive ability.

Seasonality analysis identifies the seasonal dependency, or correlation, of a measurement on preceding measurements. This characteristic is called serial dependence or serial correlation and can be examined using autocorrelation techniques. Once adjusted for the long term trend, standardised seasonal averages should sum to zero. Seasonality can be removed from a time series by block averaging or seasonal differencing (Reinert, 2010).

The traditional approach to statistical analysis of time series datasets can be summarised in terms of three stages:

- identification—pre-processing of data to achieve stationarity. This requires an initial assessment of the data to understand the overall trend, the periodicity and the likely relationship (additive or multiplicative) between components. This assessment should determine whether the data need to be differenced or transformed.
- estimation—of seasonal and longer term trends; and
- verification—by overfitting or residual analysis (Box and Jenkins, 1970).

Autoregressive (AR) processes assume that any measurement in a time series can be estimated from a linear combination of specific, preceding measurements plus an error factor. This requires that the series be stationary, and a stationary series is not serially dependent. Stationarity is generally achieved by differencing each measurement in the series with that prior measurement that is offset in time by a selected lag interval. This process can be considered as a type of linear filter. Log transformation may also be required to stabilise variance.

Moving average (MA) processes assume that each observation involves random error plus the cumulative impact of prior errors. With certain conditions, the moving average equation can be inverted into an autoregressive form. Moving averages can be computed from the linear combination of measurements within a specified time interval, that is:

Moving average for measurement

$$\text{at time } \frac{k}{2} = \sum_{i=0}^{k-1} \frac{\text{value}_{k-i}}{k}$$

Alternatively, an exponentially weighted moving average can be used which weights 'current' time measurements more heavily. An AutoRegressive Moving Average process (ARMA) effectively combines both processes.

Where observations in the time series are incomplete or noisy and result from an unobservable process, Markov models may be more appropriate for analysing trends. Discrete state variables, such as EO time series, can be modelled using hidden Markov Models (Rabiner, 1989; Viovy and Saint, 1994). Examples of studies focussing on EO time series include Scarth *et al.* (2015), Schmidt *et al.* (2016) and Frantz *et al.* (2016). Analysis of EO time series datasets is considered in detail in Volume 2D—Section 3.

8.5 Further Information

Time Series Analysis Tools for Satellite Imagery:
TimeSat—*http://web.nateko.lu.se/timesat/timesat.asp*
Google Earth Engine—*https://earthengine.google.com/*

SEOS: eLearning Tutorials: *http://www.seos-project.eu/modules/*

LEOWorks—Image Processing Tutorials: *http://leoworks.terrasigna.com/files/LEOWorks4_beta_manual.pdf*

8.6 References

Ahern, F. J., Brown, R. J., Chilar, J., Gauther, R., Murphy, J., Neville, R. A., and Teillet, P. M. (1988). Radiometric correction of visible and infrared remote sensing at the Canada Centre for Remote Sensing. '*Remote Sensing Yearbook 1988/89*' (Eds: A. Cracknell, and L. Hayes), pp. 101-127. Taylor and Francis, London, United Kingdom.

Ahuja, N., and Schachter, B. J. (1983). *Pattern Models*. John Wiley and Sons, New York.

Auricht, H. C. C., Clarke, K. D., Lewis, M. M., and Mosley, L.M. (*In press*). Have droughts and increased water extraction from the River Murray (Australia) reduced coastal ocean productivity? *Marine and Freshwater Research,* 68, pp. 1–14.

Berk, A., Anderson, G. P., Acharya, P. K., Hoke, M. L., Chetwynd, J. H., Bernstein, L. S., Shettle, E. P., Matthew, M. W., and Adler-Golden, S. M. (2003). *Modtran 4 Version 3 Revision 1 User's manual*. Airforce Research Laboratory, Hanscom, United States.

Box, G., and Jenkins, G. (1970). *Time series analysis: Forecasting and control*. Holden-Day, San Francisco, United States.

Caccetta, M., Collings, S., and Cudahy, T. (2013). A calibration methodology for continental scale mapping using ASTER imagery. *Remote Sensing of Environment*, 139, pp. 306–317.

Cliff, A. D., and Ord, J. K. (1981). *Spatial Processes: Models and Applications*. Pion, London.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., and Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, 25(9), pp. 1565-1596. doi:*http://dx.doi.org/10.1080/0143116031000101675*.

Curran, P. J. (1988). The semivariogram in remote sensing: An introduction. *Remote Sensing of Environment*, 24(3), pp. 493-507. doi:*http://dx.doi.org/10.1016/0034-4257(88)90021-1*.

Elvidge, C. D., Yuan, D., Weerackoon, R. D., and Lunetta, R. S. (1995). Relative radiometric normalization of Landsat Multispectral Scanner (MSS) data using a automatic scattergram-controlled regression. *Photogrammetric Engineering and Remote Sensing*, 61(10), pp. 1255-1260.

Frantz, D., Röder, A., Udelhoven, T., and Schmidt, M. (2016). Forest Disturbance Mapping Using Dense Synthetic Landsat/MODIS Time-Series and Permutation-Based Disturbance Index Detection. *Remote Sensing*, 8(4), pp. 277. doi:*http://dx.doi.org/10.3390/rs8040277*.

Harrison, B. A., and Jupp, D. L. B. (1990). *Introduction to Image Processing. Part TWO of the microBRIAN Resource Manual* (256 pages). CSIRO Australia, Melbourne.

Jensen, J. R. (1983). Urban/suburban land use analysis. In '*Manual of Remote Sensing*' (Ed: R. N. Colwell). Second Edn. American Society for Photogrammetry and Remote Sensing (ASPRS); John Wiley & Sons, Inc., Falls Church, Virginia.

Jupp, D. L. B. (1988a). *Fastrack Classification*. Paper presented at the First microBRIAN User Group meeting, Brisbane, Australia.

Jupp, D. L. B. (1988b). *The mosaic model for microBRIAN Classification*. Paper presented at the First microBRIAN User Group meeting, Brisbane, Australia.

Jupp, D. L. B., Mayo, K. K., Harrison, B. A., Hutton, P. G., and Ahmad, W. (1988a). Remote sensing and image processing on personal computers : the CSIRO (Australia) microBRIAN system. In '*Desktop Planning : Advanced Microcomputer Applications for Physical and Social Infrastructure Planning*'. Hargreen Pub. Co., Melbourne, Australia.

Jupp, D. L. B., Strahler, A. H., and Woodcock, C. E. (1989). Autocorrelation and regularisation in digital images. II. Simple image models. *IEEE Trans. Geosci. And Remote Sensing*, 27, pp. 247–258.

Jupp, D. L. B., Walker, J., Harrison, B. A., and Reece, P. (1988b). *The Effects of Vegetation Structure in Remotely Sensed Data and its Implications for Operational Monitoring of Large-scale Environmental Changes in Kakadu National Park*. CSIRO Division of Water Resources Research, Canberra, Australia.

Levine, M. D. (1985). *Vision in Man and Machine*. McGraw-Hill Inc., USA.

Li, F., Jupp, D. L. B., Thankappan, M., Lymburner, L., Mueller, N., Lewis, A., and Held, A. (2012). A physics-based atmospheric and BRDF correction for Landsat data over mountainous terrain. *Remote Sensing of Environment*, 124, pp. 756-770. doi:*http://dx.doi.org/10.1016/j.rse.2012.06.018*.

Nicodemus, F. E., Richmond, J. C., and Hsia, J. J. (1977). *Geometrical considerations and nomenclature for reflectance* (NBS Monograph 160). US Department of Commerce, National Bureau of Standards.

Pickup, G., and Chewings, V. H. (1988). Forecasting patterns of soil erosion in arid lands from Landsat MSS data. *International Journal of Remote Sensing*, 9(1), pp. 69-84. doi:*http://dx.doi.org/10.1080/01431168808954837*.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), pp. 257-286. doi:*http://dx.doi.org/10.1109/5.18626*.

Reinert, G. (2010). *Time Series*. Department of Statistics, University of Oxford. Retrieved from *http://www.stats.ox.ac.uk/~reinert/time/notesht10short.pdf*.

Scarth, P., Armston, J., Flood, N., Denham, R., Collet, L., Watson, F., Treyithick, B., Muir, J., Goodwin, N., Tindall, D., and Phinn, S. (2015). Operational Application of the Landsat Timeseries to Address Large Area Landcover Understanding. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci, XL-3/W3,* pp. 571–575. doi:*http://dx.doi.org/10.5194/isprsarchives-XL-3-W3-571-2015*.

Schmidt, M., Pringle, M., Devadas, R., Denham, R., and Tindall, D. (2016). A Framework for Large-Area Mapping of Past and Present Cropping Activity Using Seasonal Landsat Images and Time Series Metrics. *Remote Sensing*, 8, pp. 312. doi:*http://dx.doi.org/10.3390/rs8040312*.

Slater, P. N., Biggar, S. F., Thome, K. J., Gellman, D. I., and Spyak, P. R. (1996). Vicarious Radiometric Calibrations of EOS Sensors. *Journal of Atmospheric and Oceanic Technology*, 13(2), pp. 349-359. doi:*http://dx.doi.org/10.1175/1520-0426(1996)013%3C0349:vrcoes%3E2.0.co;2*.

Teillet, P. M., Fedosejevs, G., Gauthier, R. P., O'Neill, N. T., Thome, K. J., Biggar, S. F., Ripley, H., and Meygret, A. (2001). A generalized approach to the vicarious calibration of multiple Earth observation sensors using hyperspectral data. *Remote Sensing of Environment*, 77(3), pp. 304-327. doi:*http://dx.doi.org/10.1016/S0034-4257(01)00211-5*.

Teillet, P. M., Slater, P. N., Ding, Y., Santer, R. P., Jackson, R. D., and Moran, M. S. (1990). Three methods for the absolute calibration of the NOAA AVHRR sensors in-flight. *Remote Sensing of Environment*, 31(2), pp. 105-120. doi:*http://dx.doi.org/10.1016/0034-4257(90)90060-Y*.

Viovy, N., and Saint, G. (1994). Hidden Markov models applied to vegetation dynamics analysis using satellite remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 32(4), pp. 906-917. doi:*http://dx.doi.org/10.1109/36.298019*.

Walpole, R. E. (1974). *Introduction to Statistics*, Second Edn. Macmillan, New York.

Woodcock, C. E. (1985). *Understanding Spatial Variation in Remotely Sensed Imagery*. (PhD thesis), University of California.

Woodcock, C. E., and Strahler, A. H. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment*, 21(3), pp. 311-332. doi:*http://dx.doi.org/10.1016/0034-4257(87)90015-0*.

Woodcock, C. E., Strahler, A. H., and Jupp, D. L. B. (1988a). The use of variograms in remote sensing: I. Scene models and simulated images. *Remote Sensing of Environment*, 25(3), pp. 323-348. doi:*http://dx.doi.org/10.1016/0034-4257(88)90108-3*.

Woodcock, C. E., Strahler, A. H., and Jupp, D. L. B. (1988b). The use of variograms in remote sensing: II. Real digital images. *Remote Sensing of Environment*, 25(3), pp. 349-379. doi:*http://dx.doi.org/10.1016/0034-4257(88)90109-5*.

9 Classification

Image classification is commonly used for feature or pattern recognition and usually involves converting a continuous tone image into a thematic map. As with most image processing tools, the suitability of image classification for a particular interpretation task will depend on the nature of the application and the resolution of features to be identified in the image. Land cover mapping exercises are frequently based on image classification since all the available channels in an image can help to discriminate between different cover types. Geological analyses, however, often require delineation of linear features so may be based on enhancement techniques rather than classification.

As discussed in Volume 1, different land cover features reflect and emit electromagnetic energy differently and are thus recorded as different spectral values in an EO image. Spectral pattern recognition involves identifying typical spectral values for the various image features and grouping together pixels that have values similar to the feature values. The process of image classification implicitly involves two sets of classes:

- image classes, defined by spectral and/or spatial attributes of one or more image channels; and
- interpreter categories, which are related to ground features in the imaged scene.

Image classification attempts to define a set of image classes that represent interpreter categories as closely as possible. In broad terms, this exercise involves a number of logical stages:

- characterising image classes (see Section 9.1);
- allocating each pixel to an image class (see Section 9.2);
- grouping and labelling image classes to represent interpreter categories (see Section 9.3); and
- assessing the accuracy of labelled interpreter categories (see Section 9.4).

While these logical stages are listed as sequential steps, in practice two or more stages may be integrated into a single image processing algorithm, and such algorithms can differ between image processing systems. Also, 'recipes' for classification can imply that it is a sequential process, however in practice several of the stages involved are iterative—classes are defined, pixels allocated to classes, allocations are assessed then more classes may be defined or some may be merged (see Table 9.1). This iterative nature may be time-consuming, but is essential for a reliable and informative result from classification of EO imagery. The interpreter is encouraged to view the whole process as one of becoming acquainted with the image data and understanding what it has to offer to a particular application. The term 'exploratory data analysis' embraces this spirit of enquiry.

It is once again the vexing problem of identity within variety; without a solution to this disturbing problem there can be no system, no classification.
(Roman Jakobson)

Background image: ISODATA classification of Landsat-8 image over Weipa, Queensland, into 10 land cover classes using all optical bands. The original image is shown in the banner of Section 8. **Source:** Norman Mueller, Geoscience Australia

Table 9.1 The classification process

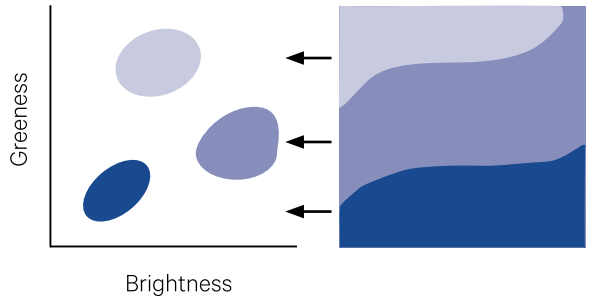
Logical Stage	Potential sub-stages	Example Process	If results of sub-stage not satisfactory
Characterise image classes	Pre-process to highlight feature(s) being classified	Despiking; greenness index; Principal Components Analysis (PCA); image segmentation	Trial other processing options
	Select most informative channels	Feature selection	Remove redundant channels or trial different channel set
	Define initial image classes	Interactive training patches; auto-seeding	Select more classes, possibly using different method(s)
Allocate pixels to image classes	Select and apply allocation algorithm	Minimum Distance classifier; Maximum Likelihood classifer	Trial different allocation algorithm(s)
	Check that final set of classes represents image variation	Create residual image	Generate more classes, possibly using additional (or different) channels, then re-allocate pixels to classes
	Check that class seeds represent all pixels allocated to them	Mean migration; spectral transfer	Recompute class seeds or generate additional classes
Group and label image classes to represent interpreter classes	Analyse class separation using spectral and spatial tools	Canonical Variates Analysis (CVA); co-occurrence matrix analysis	Remove redundant classes and split diverse classes then re-allocate pixels to classes
	Group similar image classes to represent interpreter categories using statistical tools	Spatial and spectral MST; dendograms	Review image classes and interpreter categories
	Select appropriate colour sequences to show grouped interpreter classes	Minimum Spanning Tree; interactive painting	Review available colour systems
Verify interpreter categories with independent information	Select sample pixels in image and reference data set and cross-tabulate	Interactive training; automatic algorithms to select appropriate samples; locate reference sample points in image	Select more samples for poorly represented classes
	If interpreter categories are available as a reference image channel, cross-compare classification channel with reference channel	Channel cross tabulation algorithms	Refine classification channel or reference channel
	Analyse cross tabulation(s) for level of agreement	Contingeny table manipulation; Kappa statistics	Select more samples and/or refine classification

The class development phase is an important preliminary step but is often misunderstood. Its aim is to represent the spectral variation in the image—not necessarily to build ground feature categories as such. Although training patches may be taken with ground features in mind, the classification process simply uses them to represent the structure of the image rather than an interpreter's view of the structure of the imaged scene. Thus, it is important to differentiate between image classes, which can be characterised by their spectral and/or spatial attributes, and interpreter categories, which are quite independently based on Earth surface features.

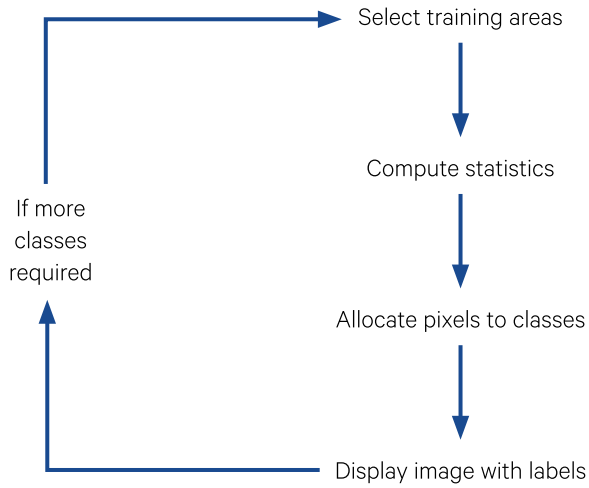
'Supervised' classification procedures have traditionally required sets of sample values, which represent the various ground features in the image (such as a particular vegetation or soil type). The sample values are then used to identify pixels with similar values on the assumption that these pixels are likely to be the same ground feature categories (see Figure 9.1). This approach is valid when the interpreter is sure of the number and range of image classes that exist, but can result in misclassifications when more classes occur than expected, or the attributes of some classes are not consistent with their assumed character. A prudent interpreter should be mindful that EO imagery offers a unique perspective of our planet, and thus can present new information about the ground features being mapped.

Figure 9.1 Supervised classification

a. The spectral values of features that are visible in the image are located on an image crossplot



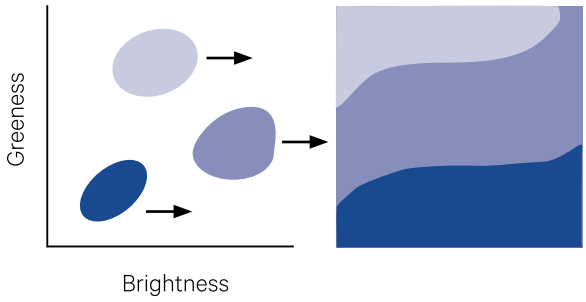
b. Summary of processes



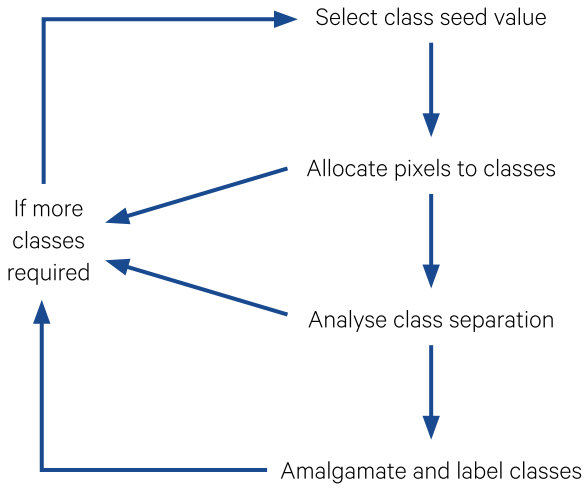
Source: Harrison and Jupp (1995) Figures 2a and 68

Figure 9.2 Unsupervised classification

a. Pixel clusters are identified on the image crossplot then related to features visible in the image



b. Summary of processes

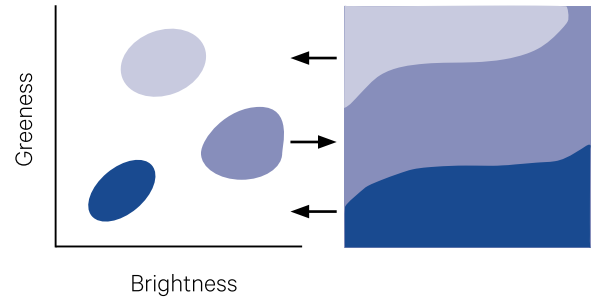


Source: Harrison and Jupp (1995) Figures 2b and 69

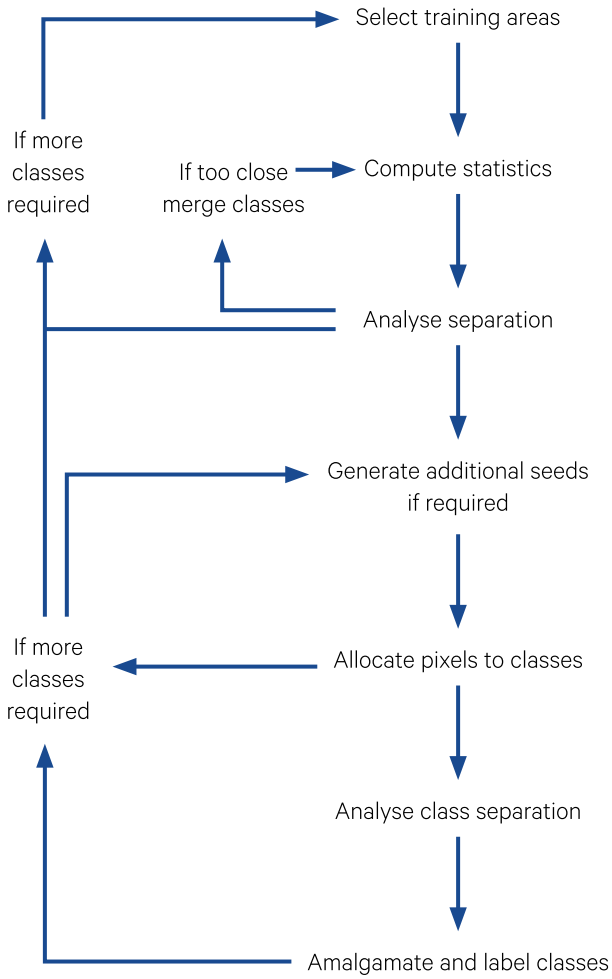
'Unsupervised' methods simply define groups of pixel values based on spectral statistics (using various 'auto-seeding' methods) then subsequently associate category labels with each group (see Figure 9.2). While some supervision is generally required in image classification to ensure that the final classes present a valid mapping of ground features, it is often beneficial to explore the subtle variability of EO imagery using various statistical tools along the way. A number of hybrid classification methods, which combine aspects of both supervised and unsupervised classification, have also been proposed (see Figure 9.3; also Excursus 9.3 and Volume 2E).

Figure 9.3 Hybrid or mixed mode classification

a. A combination of supervised and unsupervised methods are used to create classes



b. Summary of processes



Source: Harrison and Jupp (1995) Figures 2c and 70

The following sub-sections provide an overview of standard image classification methods in terms of:

- characterising classes (see Section 9.1):
  - defining classes—describes how image classes can be defined using interactive training or feature statistics, and also introduces some of the class statistics that are useful to differentiate between classes (see Section 9.1.1);
  - selecting channels—introduces popular distance measures that can be used to assess the most appropriate channels to use for image classification (see Section 9.1.2); and
  - pre-processing—reviews some of the pre-processing options that may be appropriate prior to classifying EO imagery (see Section 9.1.3);
- allocating pixels to classes (see Section 9.2)—introduces four popular approaches:
  - density slicing and painting of single channels (Section 9.2.1);
  - Parallelepiped classification for multi-channel data (Section 9.2.2);
  - Nearest Neighbour classifier (Section 9.2.3); and
  - Maximum Likelihood classifier (Section 9.2.4).
- labelling image classes (see Section 9.3)
  - clustering tools—to assess initial classifications (see Section 9.3.1) and relate image classes to interpreter categories (see Section 9.3.2);
  - aggregating classes—to represent identified clusters (see Section 9.3.2); and
  - post-classification processing—options to improve cohesion of classified image (see Section 9.3.3); and
- verifying the final classification (see Section 9.4).

One implementation of a hybrid classification methodology, the Mosaic Model approach, is introduced in Excursus 9.3. For a more considered overview of the statistics, algorithms, and philosophical issues relevant to Image Classification and Analysis, please refer to Volume 2E. An example of image classification is provided in Excursus 9.1.

Excursus 9.1—Example Classification

Source: Megan Lewis, University of Adelaide
Further information: Lewis (1998)

A vegetation mapping methodology was used to relate Landsat-5 multi-date imagery to ecological classifications. This study was conducted at Fowlers Gap Arid Zone Research Station, a 39,000 ha pastoral property in far western NSW, Australia. The climate in this region is arid. The vegetation is typical of much of the southern Australian arid zone, comprising low open chenopodiaceous shrublands, and some low open *Acacia* and *Casuarina* woodlands and grasslands on the plains (Mabbutt *et al.*, 1973, and Burrell, 1973).

Australian arid landscapes, particularly in their dry state, typically lack spectral variability, with most differentiation being related to soil brightness and vegetation cover. While the use of single-date imagery for classification depends on adequate spectral contrast between different vegetation types at that time, the use of seasonally-contrasting images takes advantage of any differences between vegetation types to improve discrimination. Because this arid vegetation varies considerably between summer and winter and wet and dry conditions, two combined Landsat images were used as a basis for classification of land cover types. The images show contrasting conditions: widespread actively growing grass and ephemeral vegetation after summer rains (see Figure 9.4a) compared with more restricted perennial vegetation growth in mid-winter of the same year (see Figure 9.4b). In total, 12 optical channels—six from each image (3 visible, 1 NIR, 2 SWIR)—were used to classify vegetation.

One hundred 6.25 ha sites were sampled, representative of the diversity of vegetation types and land systems present at Fowlers Gap. The samples were sited within relatively homogeneous areas. At each site a wheelpoint apparatus (Mentis, 1981) was used to estimate percentage cover of all physical (bare ground, stone) and biotic (ephemerals, grass and perennial shrub and tree species) ground cover components. Fifty sites were used as 'training areas' to calibrate relationships between spectral and field data, and the remaining fifty were used as validation sites to test the spectral classification.

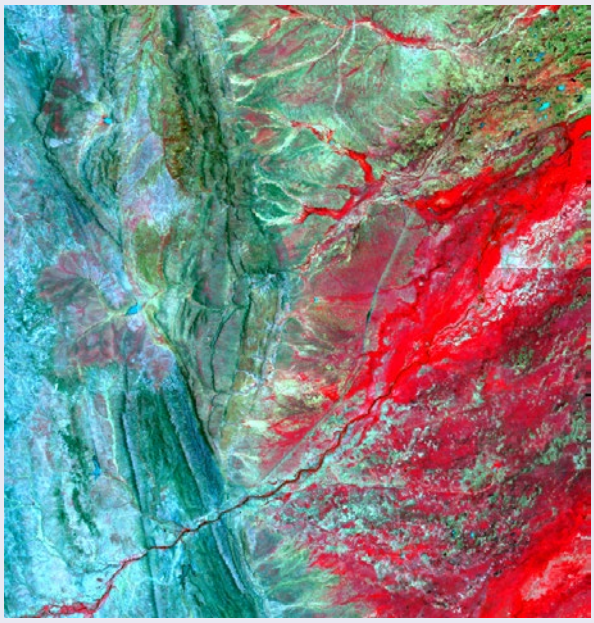
The field site data and multi-date imagery were used to:

- create nine groups of land cover based on agglomerative hierarchical clustering of field data;
- relate cover and spectral classes via a discriminant function; and
- map vegetation using a maximum likelihood classification of multi-date Landsat TM imagery (see Figure 9.5).

Figure 9.4 Study site

False colour composites of Landsat-5 TM images used to classify vegetation of Fowlers Gap Research Station, NSW.

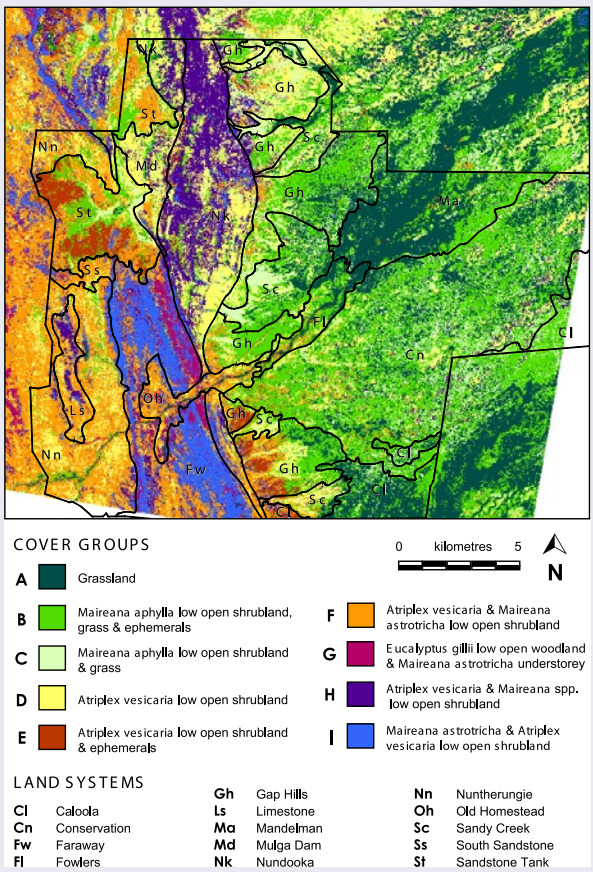
a. Summer image (30 January 1993)



b. Winter (23 June 1993)

Figure 9.5 Vegetation Classification

Maximum likelihood classification of two dates of Landsat-5 TM imagery, overlain with land systems previously mapped from black and white photography by Mabbutt (1973).



A comparison of mapped image class with ground class for the 50 verification sites (see Table 9.2) yielded a Kappa of 0.31 (s.e. 0.08, $p<0.001$). Although these figures suggest overall agreement between predicted and actual vegetation class, the accuracy of mapping for individual vegetation types varied considerably from 100% for the *Maireana astrotricha–Atriplex vesicaria* shrublands and *Eucalyptus gilii* open woodland, to 15% for *Acacia vesicaria* shrublands. These figures need to be treated with caution however, and have not been used to assess the accuracy of individual classes since the number of samples in each class is low.

Table 9.2 Comparison of image and cover classes

This matrix compares image class versus cover class for 50 independent validation sites.

Cover Class	Image Class									Total
	A	B	C	D	E	F	G	H	I	
A	3	3								6
B		5	1	2		2				10
C		4		1						5
D		1		1		4		1		7
E	1		1			1				3
F				2		3				5
G							2			2
H					1	3		4	2	10
I									2	2
Total	4	13	2	6	1	13	2	5	4	50

9.1 Characterising Classes

Remote sensing image interpretation based on spectral classification relies on the observation that different land covers reflect or emit electromagnetic energy in different and often unique ways, while similar land covers have similar spectral signatures. As detailed in Volume 1, the spectral discrimination between different land covers in an EO image will be determined by:

- the spectral and spatial resolutions of the image;
- the radiometric characteristics of the sensing device; and
- the date and time of image acquisition.

To spectrally identify a particular feature in an image requires that its spectral values be different from those of other features in that image. While spatial attributes of image pixels can also be considered during image classification (see Excursus 10.1), for simplicity the following sections primarily focus on spectral attributes.

When contemplating classification of an EO image, an interpreter is encouraged to consider factors relating both to image provenance and project management:

- Does the EO image have appropriate spectral and spatial resolutions and extents to identify the desired output categories? If not, it would be wise to either source imagery with more appropriate pixel size/channels/dates or radiometric detail, or adapt interpreter expectations to what the selected image can deliver (see Volume 2E—Section 1);
- What is the most appropriate method for defining classes? This question is best answered by exploring the image patterns and statistics, but options may be constrained by project budget (see Section 9.1.1 and Volume 2E—Section 4);
- Which image channels will best differentiate between ground features of interest to the project? The most relevant channels should maximise contrast and separation between classes (see Section 9.1.2 and Volume 2E—Section 5); and

- Should the image be pre-processed before classification to remove areas that are not of interest to the project or highlight specific features? Fewer channels and pixels would reduce the 'cost' of processing and, given the potential for misclassification, project focus should be restricted to regions 'known' to the interpreter (or covered by the available reference resources). Options for pre-processing are discussed in Section 9.1.3 and Volume 2E—Section 3.4.

Ultimately, an ill-considered approach to image classification will deliver an inadequate result. Selection of suitable imagery combined with careful management of the classification process, however, can yield valuable information. It is the responsibility of the interpreter to ensure that, subject to project constraints, the final classification reflects the information content of the original EO data set. When constraints are imposed, they should be appropriately acknowledged as the source of shortcomings in the final classification, rather than leaving users to conclude that these are the natural and inevitable consequence of using EO imagery.

9.1.1 Defining image classes

As detailed in Volume 2E, any class can be characterised by either:

- its current membership; or
- a rule that determines its membership.

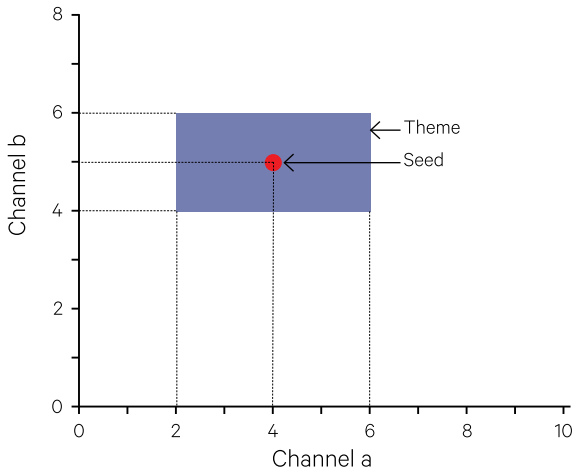
Both of these approaches can be used to define image classes. For example, once a classification has been completed, an image channel can be created in which each pixel has the value of its class 'number'. This 'classification channel' then effectively summarises the membership of each class. Alternatively, when membership is defined by an underlying rule, the relevant spectral attributes of the class need to be enunciated. This requires either:

- a valid range of values to be defined in one or more channels, such that pixels within the range(s) can belong to that class. In this text, such ranges are called 'themes'; or
- a central value is defined in one or more channels, which can be linked to a statistical criterion to determine a spectral boundary for the class. In this text, these central values are called 'seeds'.

Figure 9.6 shows the difference between seed and theme values for a two-channel image using a spectral crossplot (see Section 8.1.3).

Figure 9.6 Theme and seed values

A theme is a range of values in each channel. A seed is the central value in each image channel. In this example, the theme values are 2–6 in Channel a, and 4–6 in Channel b, while the seed values are 4 in Channel a, and 5 in Channel b.



Source: Harrison and Jupp (1990) Figure 49

The range of values for a theme can be defined by two values, the upper and lower bounds, in each image channel, with these bounds being referred to as the minimum and maximum values of the theme in a given channel. A theme can also be used to generate seed values using feature statistics as described in Section 9.1.1.2. Seeds are principally used for image classification and are also required for statistical analyses during the classification procedure (see Section 9.3 below and Volume 2E).

Sample values that represent image classes may be generated either by using:

- training patches (see Section 9.1.1.1); or
- feature statistics (see Sections 8.1 and 9.1.1.2).

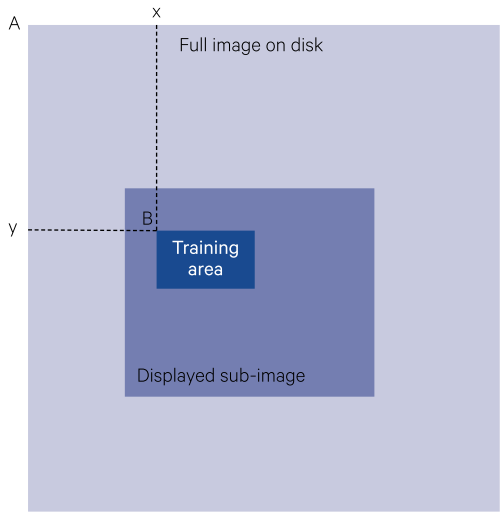
9.1.1.1 Training patches

The values of features that are visible in a displayed image may be determined interactively. This process is often referred to as training, with the resulting sample values sometimes being called a training patch (or more commonly a 'training set'). This terminology relates to the simplest form of image classification where each visually-defined sample is assumed to represent a particular ground feature. Once each pixel in the image is associated with one of these sample values, that is, the computer is 'trained' to identify the ground features based on their spectral values, the image is considered to be classified in terms of these features.

In most image processing systems the screen cursor may be used to identify an individual pixel or an enclosed area. An area is generally represented either as a rectangle, defined by two diagonally opposite corners in the image, or a polygon. As illustrated in Figure 9.7, cursor location is reported in terms of image pixels relative to the whole image, not just the displayed portion (see Sections 6.1 and 7). In many image processing systems, training areas may also be defined by specifying the locations of points along their boundaries. These locations can often be expressed either in terms of image co-ordinates or as registered map co-ordinates (see Section 7 and Volume 2B).

Figure 9.7 Locating a training area

A training area is selected using the screen cursor on the displayed image window with the location of the area being reported relative to the whole image.



Source: Harrison and Jupp (1990) Figure 50

Some image processing systems also allow training patches to be defined as:

- a line—to include those pixels intercepted by a straight line between two defined points;
- a string—ragged lines, which may be useful to sample long, thin features such as roads and rivers; and/or
- a polygon—closed area with irregular outline (see Figure 9.8 and Volume 2E—Section 4.1.1).

Figure 9.8 Five patch types

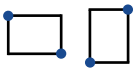
a. Point

single pixel recorded by column and row (referred to as pixel and line) position in the image

b. Line

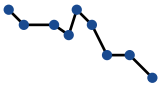
straight line interpolated between any two points

c. Box



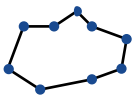
rectangle interpolated between two diagonally opposite corner points

d. String



series of straight lines interpolated between pairs of adjacent, irregularly spaced points to form a ragged line

e. Polygon



closed string which outlines a single region

Source: Harrison and Jupp (1995) Figure 45

In most image processing systems, when the training area comprises a single pixel, the digital values of all specified image channels are determined. For an area or a line, these values are extracted for all pixels in the defined region and results are often summarised in terms of the mean, standard deviation, minimum value, maximum value and/or number of pixels for each channel. If some pixels in a training area contain the null value in selected channels, the reported statistic may vary between channels, since null pixels are not used to generate image statistics.

The reported values can generally be saved as seeds or themes for subsequent use with classification or segmentation processes. Themes are commonly used in a number of image processing operations:

- to generate various statistics, including seeds (see Section 9.1.1.2);
- to implement a simple parallelepiped classification (see Section 9.2.2); or
- to define regions for image segmentation (see Section 10).

Interactive query of image values is a basic process in image analysis. As well as being used in image enhancement, classification and segmentation, sample values of selected features are required for a number of image transformations (see Volume 2C). In many image processing systems, several patches can be grouped to represent a single feature, then cumulative statistics can be generated for all patches associated with each feature type. Interactive training may also be used for identifying appropriate display parameters for particular features (see Section 9.1.1.2). Other options for using training patches with image classification are detailed in Volume 2E—Section 4.

9.1.1.2 Feature statistics

The feature identification process described in Section 9.1.1.1 can provide a basic statistical summary for pixels located within a defined training area. Other statistics that may be generated for training areas include channel histograms, and covariance and correlation matrices.

The histogram percentage points of an image feature may be used to determine display parameters to optimise the contrast of that feature in the image (typically the 1% and 99% values of the feature are selected as the minimum and maximum values for image display; see Section 4 above). Similarly, the covariance matrix of an image feature may be used to tailor a principal components transformation to that feature (see Volume 2C).

In most image processing systems, channel crossplots can be generated for:

- a complete image;
- a sub-image area (specified by screen cursor location or user-defined coordinates); and/or
- a limited spectral range as defined by a theme.

As detailed in Section 8.1.3, crossplots graphically summarise the correlation between pairs of image channels. When based on a training area, the crossplot also clearly indicates the degree of dispersion of image values in a pair of channels for the selected image region. Use of a spectral theme when generating a crossplot will obviously clip or truncate the plotted spectral space to the upper and lower bounds of the theme in the crossplotted channels, but indicates the 'density' of paired values within a particular spectral region.

Crossplots may also be used in conjunction with channel histograms to 'automatically' generate image theme values. This non-interactive mode may be used to create sample values of 'unidentified' image features as an initial step in image classification (see Excursus 9.3 below or Volume 2E). For example, the two most significant channels in an image can be crossplotted (as in Figure 9.9b), then histogram intervals corresponding to 10% or 20% of pixels (that is divisions of the 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% or 20%, 40%, 60%, and 80% points from the channel histograms) are overlaid onto the crossplot to form a grid (see Figure 9.9d). Any cells in the grid containing crossplotted values can be considered as active. The image values corresponding to the percentage points can then be used to form a theme for each active cell as illustrated in Figure 9.9e. Class seeds derived from these themes are shown relative to the image crossplot in Figure 9.10c below in term of original image channels and the first two principal components (see Volume 2C).

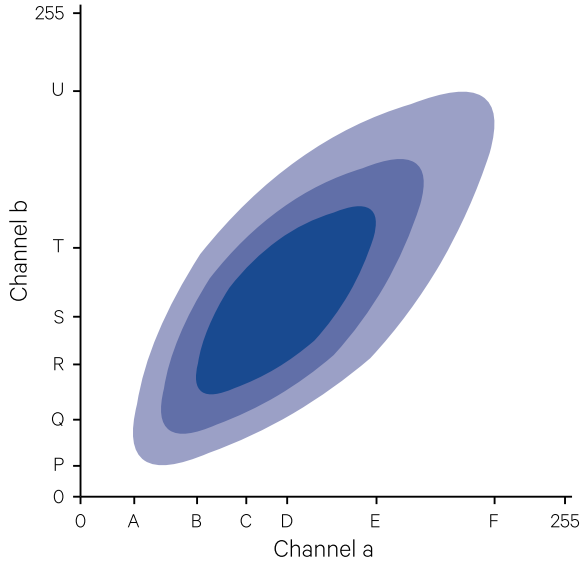
Figure 9.9 Generating themes using image statistics

Themes for auto-seeding can be generated using an image crossplot and histogram percentage points for the two most significant image channels.

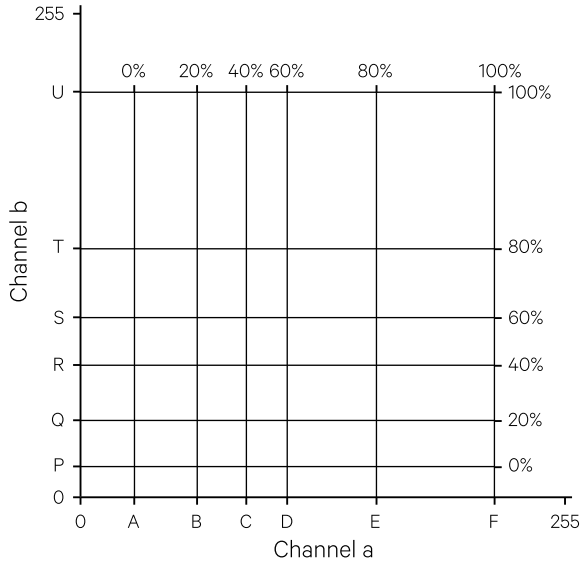
a. Histogram percentage points can be used to stratify each image channel into equal-size categories. For example, in channel a, 20% of the pixels have values in the range A–B.

Histogram % point	0	20	40	60	80	100
Channel a value	A	B	C	D	E	F
Channel b value	P	Q	R	S	T	U

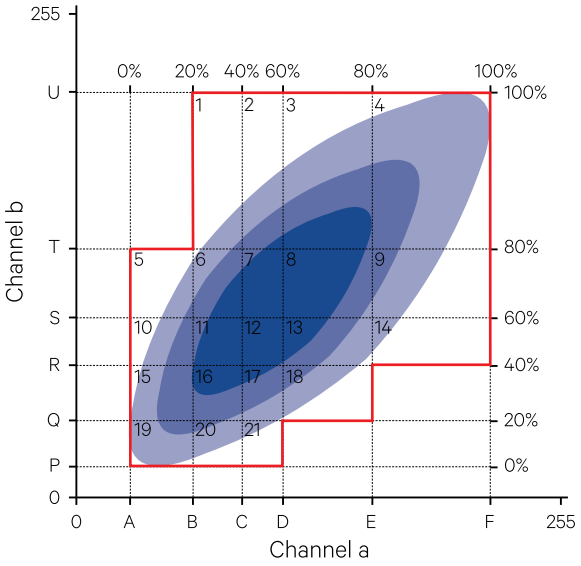
b. Crossplot of image values in channels a and b



c. Histogram percentage points from two channels are paired to form themes.



d. The percentage points can be superimposed on a crossplot to indicate which cells formed by theme ranges are 'active' in the image, that is, they do contain paired values which occur in the image.



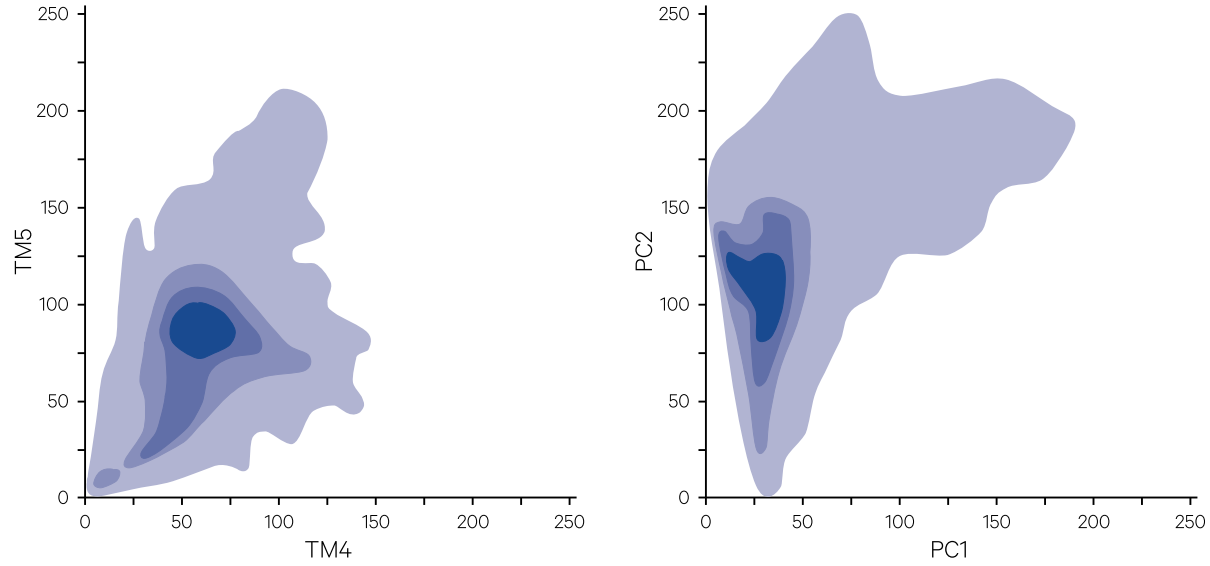
e. The ranges for each cell can then be used to define themes in those channels (see also Figure 9.10).

Theme number	Channel a	Channel b		
	Min	Max	Min	Max
1	B	C-1	T	U
2	C	D-1	T	U
3	D	E-1	T	U
4	E	F	T	U
5	A	B-1	S	T-1
6	B	C-1	S	T-1
7	C	D-1	S	T-1
8	D	E-1	S	T-1
9	E	F	S	T-1
10	A	B-1	R	S-1
11	B	C-1	R	S-1
12	C	D-1	R	S-1
13	D	E-1	R	S-1
14	E	F	R	S-1
15	A	B-1	Q	R-1
16	B	C-1	Q	R-1
17	C	D-1	Q	R-1
18	D	E	Q	R-1
19	A	B-1	P	Q-1
20	B	C-1	P	Q-1
21	C	D	P	Q-1

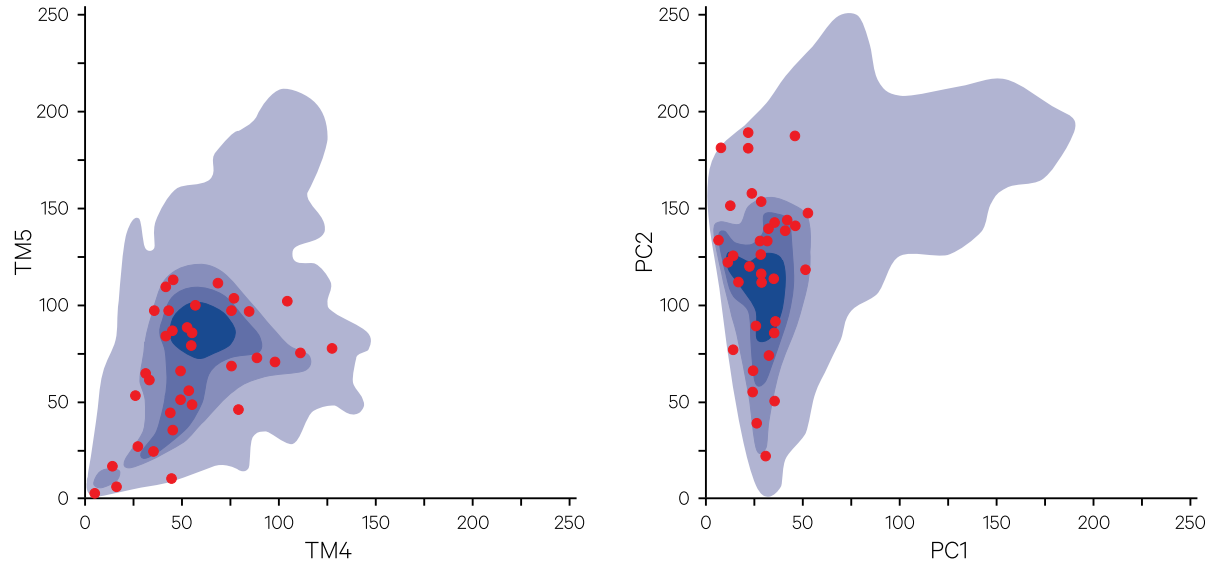
Source: Harrison and Jupp (1990) Figure 51

Figure 9.10 Comparison of class seeds

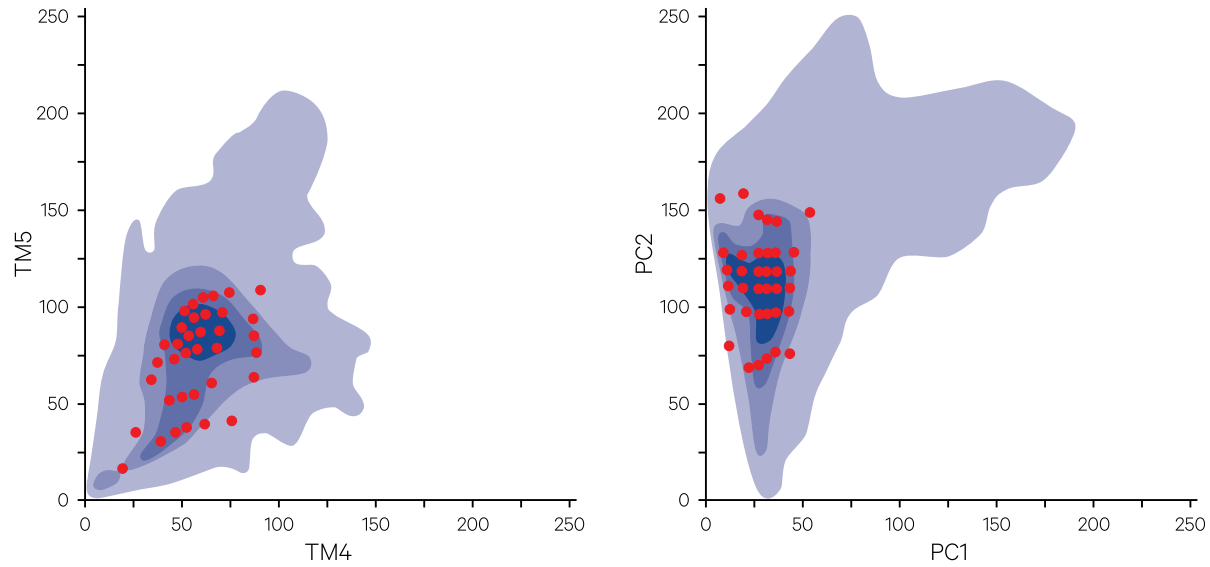
a. Example crossplots for a Landsat TM image in terms of original channels (left) and transformed to principal components (right)



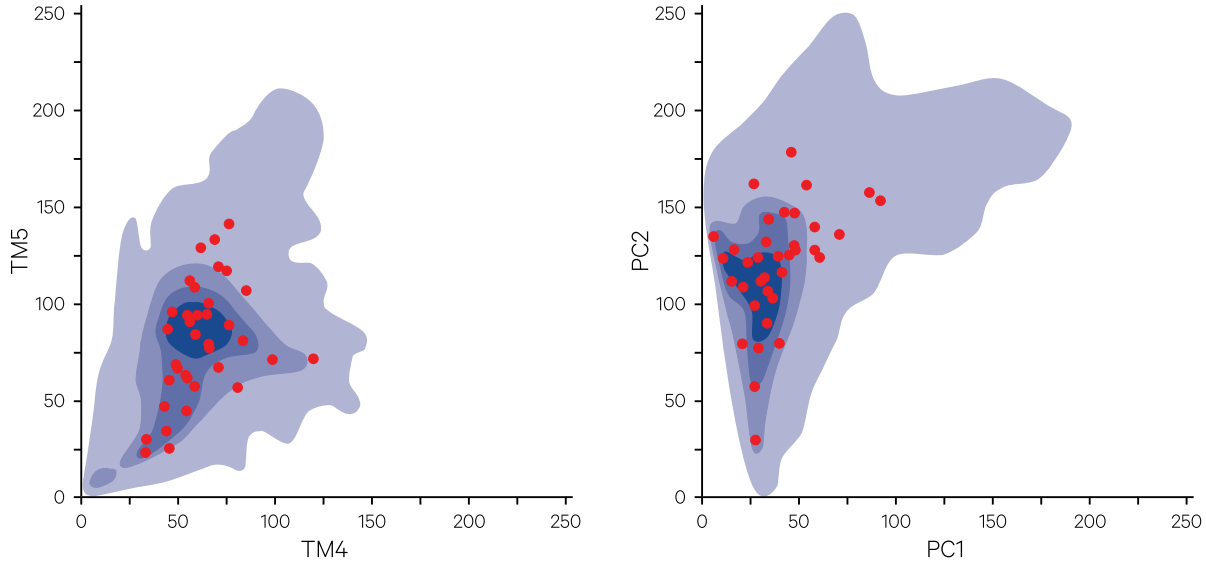
b. Class seeds derived from training patches



c. Class seeds derived from autoseeding themes (see Figure 9.9)



d. Class seeds derived from unsupervised classification



Source: Harrison and Jupp (1995) Figure 55

Various statistics can be computed using image themes. All pixels that satisfy a theme in one or more images would be used to compute the theme statistics of mean, standard deviation, minimum and maximum values, as well as the associated number of pixels. The mean values of a set of themes can often be used as seeds in image classification (see Section 9.2). Class seeds generated using this method effectively represent unidentified image features. The philosophy of the Mosaic Model approach utilises such seeds as sample class values (see Excursus 9.3). After image pixels are grouped or allocated to the classes, the features they represent can be identified interactively and an appropriate label assigned to each class. Pixel counts from themes, that is, the number of pixels in each theme, can also be converted to estimates of ground area using the results of image rectification (see Section 7 and Volume 2B).

A range of other approaches have been proposed to automatically generate themes or seeds for image classes (see Volume 2E—Section 4.2), including:

- crossplotting first principal component (PC1) and second principal component (PC2) (or the first two canonical variates: CV1 and CV2—see Section 9.1.2.1) to locate local maxima points then using 'spectral transfer' to associate new channels values with each class (see Section 9.2.3);
- AMOEBA clustering classifier (Bryant, 1979, 1989, 1990), which uses spatial and spectral characteristics to derive class statistics from homogeneous patches in the image;
- non-parametric distance measures—can use predefined ranges to cluster pixels values around arbitratry seeds, such that each pixel which cannot belong to an existing class becomes a new class seed (eg. Skidmore, 1989);
- semi-automatic approaches based on region-growing away from defined pixels (eg. Bolstad and Lillesand, 1992);
- use clustering algorithms to group pixels into classes defined by gradients shown on channel crossplots (eg. Baker *et al.*, 1991);
- locate values of concentrated pixels as shown on three-dimensional contour plots, that is, a two channel crossplot on which pixel density is highlighted (Goldberg and Shlien, 1976; Schachter *et al.*, 1976; Narendra and Goldberg, 1977; Goldberg and Shlien, 1978). For example, Watershed classification (Watson, 1987; Watson *et al.*, 1992) views the contour plot as a landscape, and uses connectivity principles to define spectral regions as separate 'catchments' (Hashim, 1996); and
- in some implementations of the minimum distance classifier (see Section 9.2.3), unclassified pixels can become new seeds. Subsequently, the mean migration process (see Figure 9.19 below) can be used to redefine class seed values as the mean values of their member pixels.

Three different sets of class seeds that have been generated for a Landsat TM image are shown in Figure 9.10 on crossplots of TM4 vs TM5 and PC1 vs PC2. Seeds derived from training patches tend to be concentrated where pixel values are concentrated whereas the unsupervised class seeds sample a broader cross-section of the image crossplot values.

9.1.2 Comparing classes and classifications

Once a class is defined by its membership, a range of statistics can be computed to further characterise the class, quantify its separation from other classes, and determine its inherent stability and cohesion. Relevant measures of class separation are detailed in Volume 2E—Section 5.2. These metrics can be computed for classifications including and excluding selected channels to determine those channels which are most effective in separating a set of pre-defined classes. Similarly, the 'value' of individual classes can be tested by computing these statistics for a classification that includes them compared with a classification that does not.

9.1.2.1 Classification statistics

A range of statistics related to image classification are detailed in Volume 2E—Sections 2 and 5. Here we provide an overview of two basic methods that can be used to analyse and compare classifications—Canonical Variates Analysis (CVA; see Volume 2E—Sections 2 and 5) and Numerical Taxonomy (see Volume 2E—Section 6). Interested readers are encouraged to refer to Volume 2E for a full explanation of the underlying assumptions and limitations of these tools.

CVA allows the image data space to be rotated and rescaled to maximise the differences between sets of classes relative to the variation within them by defining linear combinations of selected image attributes (see Figure 9.11). The total covariance matrix (***T***) is the sum of the between-class covariance matrix (***B***) and the within-class covariance matrix (***W***), that is:

$$T = B + W$$

where

B is computed from the class mean (or seed) values; and

W is the pooled (or average) within-class variability.

In image terms:

image being classified = mean image + residual image

(minus the offset used to create the residual image; this is 127 for 8-bit (byte) images) and:

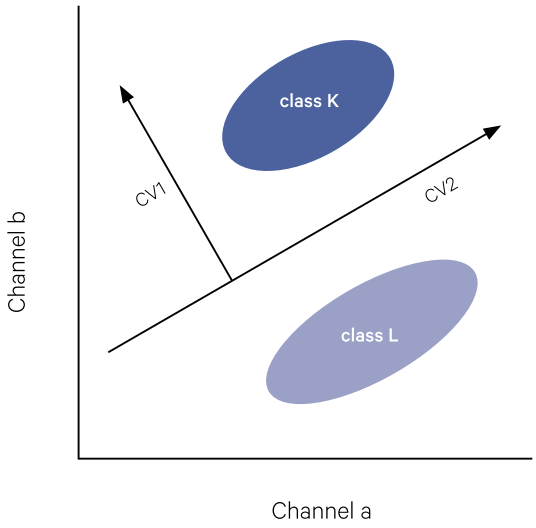
T is the covariance matrix of the image being classified;

B is the covariance matrix the 'mean' image, in which pixel values are replaced by the seed (or mean) values of the class that pixel has been allocated into (see Section 9.2.3); and

W is the covariance matrix the 'residual' image, which is the difference between the original image being classified and the 'mean' image (see Section 9.2.3).

Figure 9.11 Canonical Variates Analysis (CVA)

To maximise the separation between class K and class L, we can rotate and rescale channel axes so that the first canonical variate (CV1) gives greatest separation between class distributions. CV2 is defined as orthogonal (perpendicular) to CV1 (see also Volume 2E—Sections 2 and 5)



Source: Harrison and Jupp (1995) Figure 23

In a 'good' classification, the ratio of the within-class variation to the between-class variation (see Volume 2E—Section 5.1.3) will be small, and the residual image will have the structure of a random sample from a normal multivariate population with a zero mean vector and variability summarised by the within-class covariance matrix. The residual image is never spatially random but it should not contain significantly coherent patterns (see Section 9.2.3).

Various statistical tests based on CVA can be used to compare alternative classifications. These tests and their underlying statistics are presented in greater detail in Volume 2E—Sections 2 and 5. When derived for different classifications of an image (whether differences in classifications stem from using different classes, channels and/or allocation algorithms), such statistical tests can indicate which classification provide the best overall separation between classes. Some commonly used tests for comparing classifications include:

- minimum Trace (sum of diagonal elements) of ***W***:

  Tr ***W***

  This indicates that the within-class sum of squares is small (and is the same as selecting the classification with max Tr ***B***). This statistic implies that the individual classes are spherical and homogeneous, so outliers or small clusters near large ones could render misleading results. Also, it is not invariant to scaling so gives different results for standardised and non-standardised data.

- Wilk's λ or Likelihood ratio statistic:

  $$\text{minimum } \frac{|\boldsymbol{W}|}{|\boldsymbol{T}|}$$

  where

  $|\boldsymbol{W}|$ is the determinant of $\boldsymbol{W}$

  which is the same as testing min $|\boldsymbol{W}|$ or max $|\boldsymbol{T}|/|\boldsymbol{W}|$. This statistic assumes that all classes have the same basic shape but do not have to be spherical. This test involves more computations but is generally preferred over min Tr $|\boldsymbol{W}|$.

- largest root criterion:

  $$\lambda_{max}\left(\boldsymbol{W}^{-1}\boldsymbol{B}\right)$$

  The largest overall ratio of between-to-within variation is used to indicate that one classification is better than another.

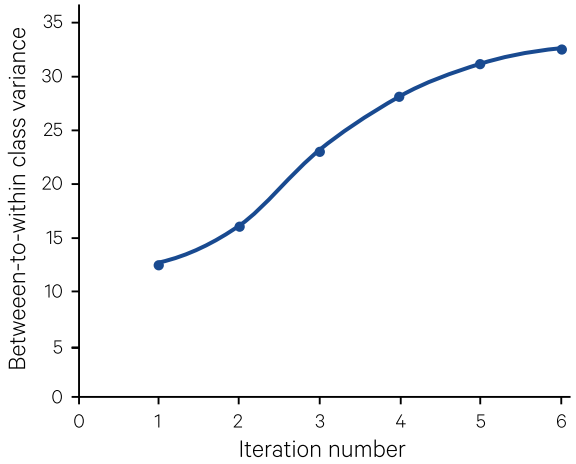
- Hotelling's trace criterion:

  $$\max\left(\text{Tr}\left(\boldsymbol{W}^{-1}\boldsymbol{B}\right) = \sum_{i=1}^{n}\lambda_i\right)$$

  The total between-to-within variance ratio for the individual variables is used to assess class separation. This criterion is more sensitive to image dimensionality.

- As new spectral classes are added to a classification, one indication of the value of additional classes is the trend in between-class to within-class variation, which tends to stabilise for the 'right' number of classes. This can be plotted for each iteration as shown in Figure 9.12.

Figure 9.12 Assessing iterative classifications



Source: Harrison and Jupp (1995) Figure 27 [Adapted from McVicar *et al.* (1989)]

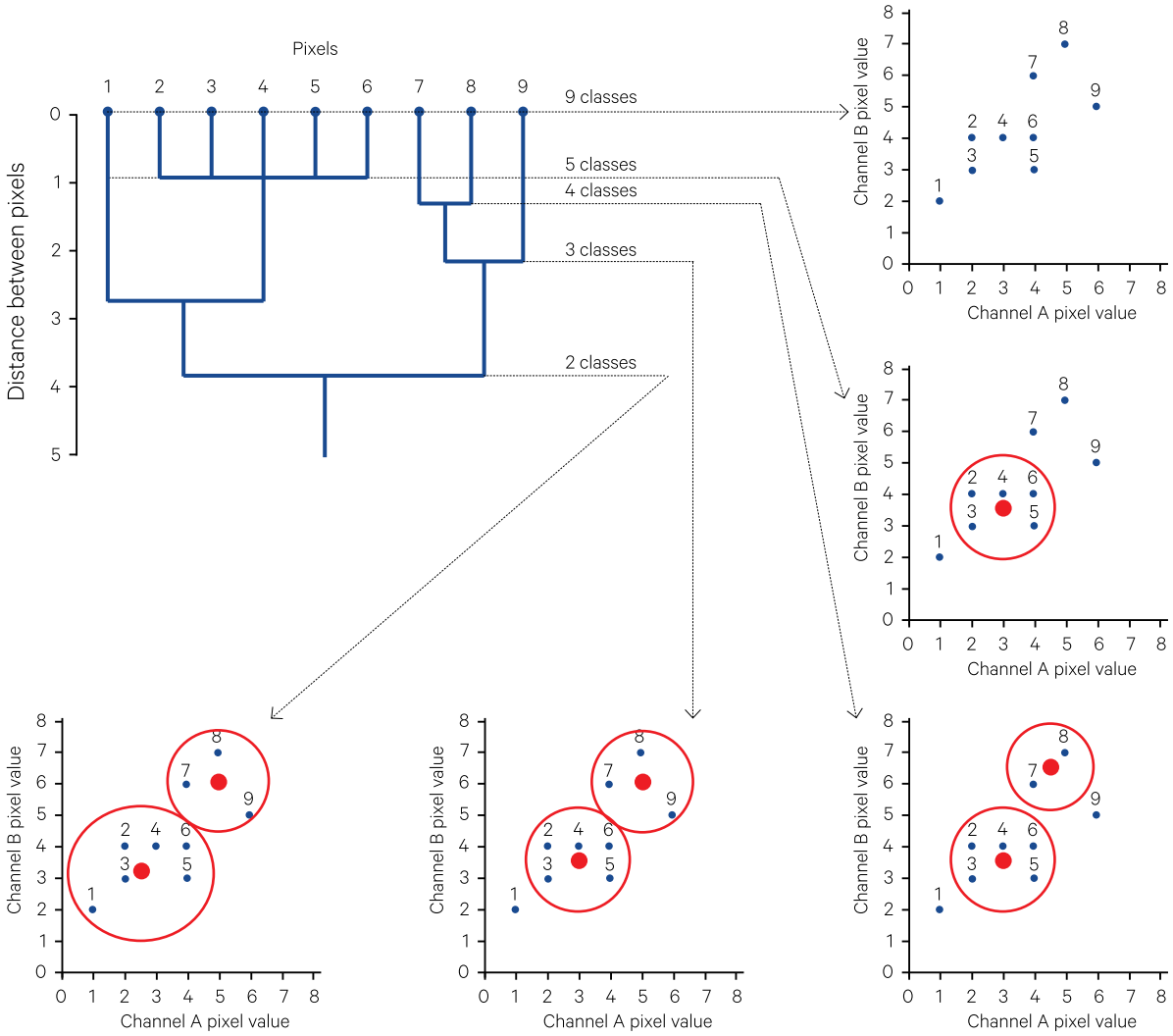
Another approach to analysing class separation uses tools developed for numerical taxonomy. In this context, numerical taxonomy embraces methods that systematically group classes using their numerical characteristics. For image classification, most taxonomic methods use some measure of the distance between classes to determine how to combine them into groups. Methods either start with one group (all classes) and systematically divide this group into multiple sub-groups in a hierarchical fashion, or alternatively, agglomerative methods start with multiple groups (such as a set of image classes) and, on the basis of some 'similarity' measure, cluster the individual classes into related groups. Such methods are detailed in Volume 2E—Section 6.1.

The hierarchical interconnections between classes can be viewed using either:

- dendrograms—a tree diagram, which successively links classes to form larger groups, with linkages being ordered by the distance between groups (see Figure 9.13). Different algorithms use different methods for determining intra-class distance and different strategies for forming groups so can and do create different dendrograms (see Volume 2E—Section 6.1.1). However, the first and last linkages are often similar for dendrograms generated by different algorithms; or
- minimum spanning trees (MST)—a connected, non-cycling graph that links all members of a set. In the context of image classification, MST can be based on both the spectral and spatial similarity of classes. A variety of methods can be used to generate MST, using different axes and different similarity measures (see Volume 2E—Section 6.3).

Figure 9.13 Dendrogram example

This tree diagram progressively groups nine pixels (shown as blue dots) into groups based on some measure of the distance between them. The mean values of the merged groups are shown as red dots.



Source: Adapted from Richards (1986)

For example, the Mahalanobis distances in a selected set of channels between pairs of classes can be computed and ordered in terms of increasing value (see Table 9.3). These ordered distances can be used to construct both a spectral minimum spanning tree (MST) and a dendrogram of taxonomic linkages. To create the MST, the pairs of classes are plotted and linked on a crossplot of selected image channels (or transformed channels such as CV1 versus CV2). The closest pair of classes (eg. classes 9 and 17 in Table 9.3) are joined first, then the second pair (classes 3 and 6 in Table 9.3) and so on until all linked pairs are connected (see Figure 9.14). It is also informative to use different coloured links for different ranges of distances between class pairs when creating the MST.

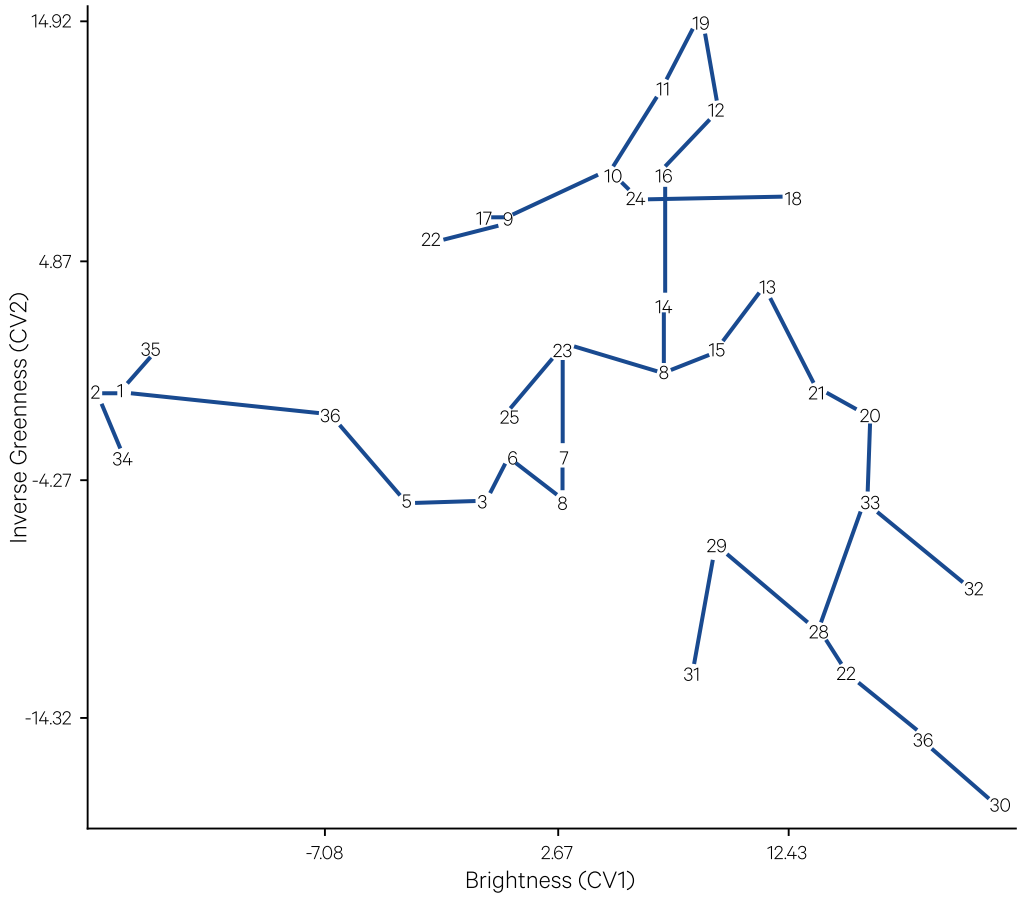
The MST and dendrogram of clusters are useful tools to assess the distance between classes or to evaluate training sets, that is to 'learn' how classes are separated in different attributes, during the initial exploratory phase of image classification. If two class seeds are separated by less than two units in Mahalanobis distance, they could be considered to represent a single statistical class. The MST is valuable during the classification labelling stage to identify classes which form gradients rather than discrete, spectrally distinct clusters (see Section 9.3).

Table 9.3 Minimum spanning tree linkages

Linkage	Class a	Class b	Mahalanobis distance
1	9	17	2.64
2	3	6	2.852
3	21	20	4.078
4	8	35	5.689
5	28	27	5.952
6	4	7	6.225
7	15	13	6765
8	20	33	7.130
9	6	4	7.188
10	5	3	7.206
11	9	22	8.637
12	27	26	10.668
13	16	12	10.680
14	8	14	11.118
15	23	25	11.887
16	11	10	13.276
17	10	9	15.841
18	26	30	16.332
19	7	23	16.476
20	36	5	16.876
21	28	29	18.238
22	33	32	20.799
23	13	21	22.548
24	14	16	22.585
25	33	28	23.997
26	23	8	24.301
27	10	24	29.686
28	24	18	29.725
29	29	31	31.463
30	12	19	38.313
31	19	11	42.595
32	1	36	46.874
33	1	35	77.683
34	1	2	88.883
35	2	34	202.427

Figure 9.14 Spectral MST

In this example, class numbers are graphed by their brightness (CV1) and inverse greenness (CV2) values. Classes are then linked in order of increasing Mahanobis distance between them (see Table 9.3).



Source: Harrison and Jupp (1995) Figure 59

Spatial statistics can also be used to compare classifications. For example, the class co-occurrence matrix, which quantifies the frequency with which pairs of classes occur in adjacent pixels in the image, can be used to construct a spatial MST. This effectively links classes on the basis of their spatial proximity (see Volume 2E—Section 6.4).

9.1.2.2 Measuring class separation

Before mentioning some statistics that are commonly used to measure class separation, we should pause and reflect on what they mean. Given that a class represents a group of pixels, how do we measure the distance between two groups? There are many options for this, including:

- distances between class seeds;
- some 'sensible' average of the set of distances between the values of each pixel in one class and the values of each pixel in another class;
- weighting factors that account for different data ranges in different channels; and/or
- measures that consider the distribution of pixels in each class.

Some of the more commonly encountered distance metrics are based on the Minkowski series of distance functions summarised in Table 9.4. These metrics simply measure the separation between class seeds (which are effectively assumed to be class centres) in the n-dimensional space defined by *n* image channels. Obviously, it is important when using these metrics that the class seeds do represent class centres.
A standardised version of Euclidean distance, the Karl-Pearson distance (K^2), uses weights derived from channel standard deviation (s_k) to account for variability between channels:

$$\sum_{k=1}^{n_c} \frac{\left(x_{ik} - x_{jk}\right)^2}{s_k^2}$$

Table 9.4 Minkowski metrics

Metric	Equation	Description
Generic series of distance functions ($\|.\|.\|_p$)	$d_p\left(X_i,X_j\right)=\left(\sum_{k=1}^{n_c}\left\|x_{ik}-x_{jk}\right\|^p\right)^{1/p}$	p is greater than or equal to 1 (p =1 for city block; p =2 for Euclidean distance; p = ∞ for Chebychev)
City block (taxicab, Manhattan or L_1)	$d_1\left(X_i,X_j\right)=\sum_{k=1}^{n_c}\left\|x_{ik}-x_{jk}\right\|$	X_i is a member of class i, X_j is a member of class j, n_c is the number of attribute (or channel) values at each pixel, x_{ik} is the k'th component of X_i, x_{jk} is the k'th component of X_j, and | | denotes the absolute value (or unsigned difference).
Euclidean distance (L_2)	$d_2\left(X_i,X_j\right)=\left(\sum_{k=1}^{n_c}\left\|x_{ik}-x_{jk}\right\|^2\right)^{1/2}$	$\sum\left\|x_{ik}-x_{jk}\right\|^2$ is commonly written as $\left\|\boldsymbol{x}_i-\boldsymbol{x}_j\right\|^2$
Chebychev (L_∞)	$d_\infty\left(X_i,X_j\right)=\text{Max}_{k=1,n_c}\left\|x_{ik}-x_{jk}\right\|$	
With channel weighting function	$d_p\left(X_i,X_j\right)=\left(\sum_{k=1}^{n_c}w_k\left\|x_{ik}-x_{jk}\right\|^p\right)^{1/p}$	W_k are pre-determined weighting values for the channels

Other class separation statistics consider the distribution of values around class centres. For example, the Mahalanobis distance (D^2; Mahalanobis, 1936; Rao, 1948) can be computed as:

$$d_p\left(X_i,X_j\right)=\sqrt{\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right)^T\boldsymbol{W}^{-1}\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right)}$$

where $\boldsymbol{W}$ is the selected class covariance matrix, $\boldsymbol{W}^{-1}$ is its inverse, and each set of channel values for individual pixels or classes in the image can be considered as a one-dimensional array, or vector, $\boldsymbol{x}_k$. This distance effectively measures the separation between class seeds and assumes all classes can be characterised by equal, normal distributions. Mahalanobis distance is particularly relevant to measuring inter-class distances when the Minimum Distance classifier has been used to allocate pixels to classes and can be shown to be equivalent to the Euclidean distance in the Canonical Variate transformed space (see Section 9.1.2.1 and Volume 2E).

Separation between two overlapping probability distributions can be measured by a statistics called Divergence (see Volume 2E—Section 5.2). This is most relevant to classifications defined by Maximum Likelihood allocation (see Section 9.2.4). For EO analyses, the Transformed Divergence is more commonly used. This is more sensitive than average divergence to classes which are not well-separated (Swain and Davis, 1978; Richards, 1986; Thomas *et al.*, 1987):

$$D'_{ij}=k\left[1-e^{-D_{ij}/8}\right]$$

where k scales the D'_{ij} values between 0 for identical classes and k for maximum separation.

Another particularly robust class separation statistic for classifications based on Maximum Likelihood allocation is the Jeffries-Matusita distance (Swain and Davis, 1978):

$$J_{ij}=\sqrt{\int\sqrt{P\left(\boldsymbol{x}\,|\,class\ i\right)}-\sqrt{P\left(\boldsymbol{x}\,|\,class\ j\right)}dx}$$

For normally-distributed classes this becomes:

$$J_{ij}=\sqrt{k\left(1-e^{-B_{ij}}\right)}$$

where k scales the J_{ij} values between 0 for identical classes and k for maximum separation (that is, no overlap). When the underlying populations are normally distributed ($N\left(\mu,\boldsymbol{W}\right)$):

$$B_{ij}=\frac{1}{2}\log\left(\frac{\left|\frac{1}{2}\left(\boldsymbol{W}_i+\boldsymbol{W}_j\right)\right|}{\sqrt{\left|\boldsymbol{W}_i\right|}\sqrt{\left|\boldsymbol{W}_j\right|}}\right)+\frac{1}{8}\left(\mu_i-\mu_j\right)^T\left(\frac{\boldsymbol{W}_i+\boldsymbol{W}_j}{2}\right)^{-1}\left(\mu_i-\mu_j\right)$$

B_{ij}, the Bhattacharyya distance, is a natural extension of the Mahalanobis distance to cases where classes have different covariance matrices. The general expression for the Bhattacharyya distance for a non-normal distribution is:

$$e^{-B_{ij}}=\int_X\sqrt{P\left(\boldsymbol{x}\,|\,Class\ i\right)P\left(\boldsymbol{x}\,|\,Class\ j\right)}dx$$

9.1.3 Selecting image channels

Ideally, to maximise the efficiency of the image classification process, the image being classified would not contain a significant degree of redundant information. Accordingly, it is useful to assess the intrinsic dimensionality of an image prior to classification using various 'feature selection' methods, which highlight the most informative image channels (see Section 9.1.3.1 and Volume 2E—Section 5.3.1). In some cases, pre-processing of imagery is appropriate prior to classification (see Section 9.1.3.2).

9.1.3.1 Feature selection

While modern computing power happily copes with large datasets, a compromise has traditionally been required between classification accuracy and computational efficiency. Greater classification efficiency is achieved by using as few image channels as possible, that is, identifying the subset of channels that provide maximum discrimination between classes. Two criteria that have been used for this purpose are:

- minimum error rate; and
- maximum separation between classes (McKay and Campbell, 1982b; Young and Fu, 1986).

Relevant measures of class separation are introduced in Section 9.1.2 and detailed in Volume 2E—Section 5.2. The best subset of image channels is commonly assumed to be the one that optimises some selected class separability measure. For example, Mahalanobis distances between classes can be computed with and without selected channels to assess the contribution being made by one or more channels to class separation. Similarly, from a set of potential channel combinations, for example, the combination that maximises the Jeffries-Matista distance (J) could be regarded as the 'best' (see Section 9.1.2.2 and Volume 2E—Section 5.3.1).

McKay and Campbell (1982a) discussed three broad types of method for variable selection, namely:

- methods that use stepwise F-tests;
- methods associated with CVA; and
- methods which evaluate all possible subsets (see Volume 2E—Section 2).

They concluded that a combination of CVA and 'all-subsets' methods provides the most effective basis for channel selection, although for an image with a large number of channels, the 'all-subsets' approach may not be feasible. Some of the tests that are used to compare CVA results are reviewed in Section 9.1.2.1 and Volume 2E—Section 5.1.3.3. When using such statistics, however, it is worth remembering that pooled statistics may overlook the significance of small classes that represent specific ground features.

9.1.3.2 Pre-processing

It may also be appropriate to apply one or more image transformations to imagery prior to classification (see Volume 2E—Section 3.4). Commonly used pre-processing includes:

- removal of 'noise' pixels using despiking routines (see Volume 2C);
- segmentation to remove pixels unrelated to a particular study (see Section 10);
- rescaling selected image channel(s) to emphasise selected portions of the data range (see Volume 2C);
- channel ratios to highlight specific ground feature characteristics, such as vegetation greenness (see Volume 2C and 3A); and
- Principal Component Analysis (PCA) to reduce the dimensionality of the image (see Volume 2C).

9.2 Allocating Pixels

Image classification involves reducing the variation contained within an EO image to a set of classes that can be labelled as scene features, such as land cover. The degree to which different features can be discriminated in an image largely depends on the resolution of the image data itself, but is also affected by the flexibility of the classification method being applied.

Various methods can be used to allocate pixels to classes during image classification. Four of the most commonly encountered approaches include:

- Density slicing and painting of single channels (Section 9.2.1);
- Parallelepiped classification for multi-channel data (Section 9.2.2);
- Nearest Neighbour classifier (Section 9.2.3); and
- Maximum Likelihood classifier (Section 9.2.4).

The first two allocation methods listed above rely on themes to define valid data ranges for classes, while the last two methods determine class membership on the basis of spectral separation from class seeds.

All allocation approaches may be applied in a 'supervised' or 'unsupervised' (or mixed) mode, where the terminology refers to whether the spectral statistics used to define the classes were derived from examples of the ground feature categories to be mapped (the supervised mode) or defined numerically to describe the variation in the data without reference to a pre-determined set of ground feature categories (the unsupervised mode). It is very difficult to 'train' a classifier to recognise categories whose definition is not related to the physics being observed in the EO image. It is also useful to investigate aspects of the data that might not be included in an *a priori* categorisation of the image.

A range of other model-based and knowledge-based classification approaches are described in Volume 2E—Section 8:

- contextual classifiers (see Volume 2E—Section 8.1);
- mixture classifiers (see Volume 2E—Section 8.2);
- decision trees (see Volume 2E—Section 8.3);
- expert systems (see Volume 2E—Section 8.4); and
- neural networks (see Volume 2E—Section 8.5).

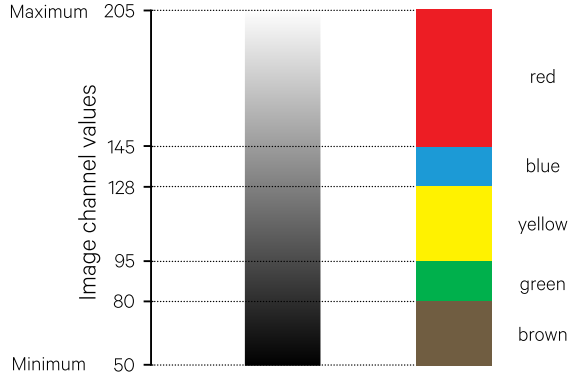
9.2.1 Density slicing

The simplest form of image classification involves grouping ranges of consecutive values in a single image channel to represent different categories. As with most classification results, these categories are typically displayed as different colours so this process also performs an image enhancement function in a a similar fashion to pseudo-colouring (see Figure 9.15 and Figure 5.12).

Most image processing systems utilise the colour LUT of the colour display system to instantaneously redefine colours of the image data being stored in display memory (see Section 6). For example, after a grey-scale image has been displayed, defined ranges of image values could be associated with a selected display colour then reset to their original grey tones.

Figure 9.15 Density slicing

A single image channel can be pseudo-coloured by assigning ranges of image values to specific colours.



Source: Harrison and Jupp (1990) Figure 52

In most image processing systems, the painted colours can be considered as image categories or classes and summarised in a single channel in which each painted pixel in the image is given a class number associated with its painted colour and unpainted pixels are set to the null value. This 'classification channel' can be displayed as a grey-scale image for interactive colouring or 'cleaned' using a modal filter (see Volumes 2C and 2E).

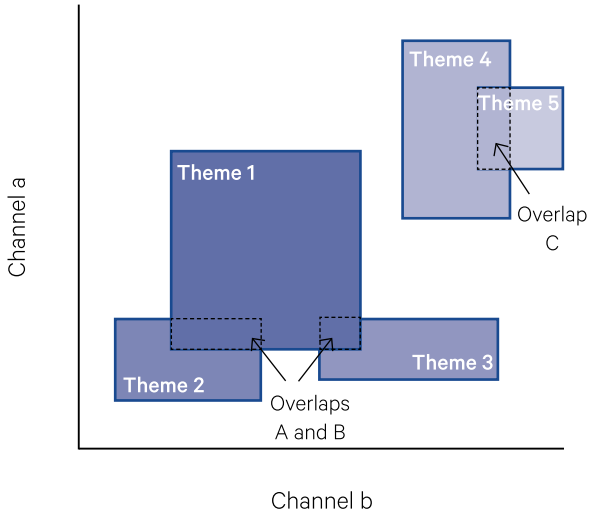
Interactive density slicing is used for labelling and aggregating classes when the classification is represented as a single channel (see Section 9.3 and Volume 2E). The technique is also useful to enhance single channel imagery or transformed data channels (see Volume 2C).

9.2.2 Parallelepiped classification

A simple method for multi-channel image classification involves the use of themes defined by minimum and maximum values (see Section 9.1.1.2). While different implementations of this technique vary, pixels are generally mapped into the first theme that they satisfy. As shown in Figure 9.16, when the categories formed using this process are shown on a crossplot, they are represented as rectangular boxes. This characteristic has resulted in this form of classification being referred to as 'parallelepiped' classification. The process can be viewed as a multi-channel form of density slicing with the requirement that a pixel fall into the specified data range in each image channel.

Figure 9.16 Parallelepiped classification

This simple classification technique uses multi-channel themes to define image classes. If theme values overlap, pixels are allocated to the first theme they satisfy. In this case overlap areas A and B would be assigned to Theme 1, while overlap area C would be assigned to Theme 4.



Source: Harrison and Jupp (1990) Figure 53

Themes may be defined using interactive interrogation (as described in Section 9.1.1.1) or feature statistics (see Section 9.1.1.2). Each image pixel is checked against these theme bounds and associated with a particular theme (generally the first one it satisfies). The operation is performed using the image file values and considers all image channels. Being a disk-based process, this classification method is slower than interactive density slicing (see Section 9.2.1).

The allocation process, that is associating a pixel with one of the themes, may be applied by colour 'overlaying' on a displayed (colour composite) image or by writing an output image to disk using the colour codes associated with each theme. This simple classification technique is also useful for defining appropriate bounds of particular features for image segmentation (see Section 10.2). Since certain transformations are more effective when applied to a pre-segmented image, some implementations of image operations, such as ratioing or destriping, may also allow themes to be specified to implement image segmentation during processing (see Volume 2C).

9.2.3 Minimum distance classifier

A Minimum Distance (or Nearest Neighbour) classifier allocates each pixel in an image to one of the available classes using a minimum distance rule, that is the class whose seed values are deemed to be the 'closest'. As introduced in Section 9.1.2 and detailed in Volume 2E—Sections 5 and 7, various metrics can be used to determine the distance between two classes. Some of the metrics commonly encountered in minimum distance classifiers are illustrated in Figure 9.17.

One of the principal advantages of the minimum distance classifier is its computational efficiency, which generally is linearly proportional to the number of classes. Many implementations of this classifier allow new seeds to be created automatically when a pixel cannot be allocated to an existing class seed.

A common algorithm for this purpose is the 'City Block' metric (see Table 9.1 and Figure 9.17b), which measures the distance between a pixel and a class seed weighted by channel gate values. In mathematical terms, this means that a pixel with attribute values $\boldsymbol{x}$ is allocated to the decision region D_i from which it has the minimum measurement $dist(\boldsymbol{x}, D_i)$:

$$dist\left(\boldsymbol{x}, D_i\right) = \sum_{j=1}^{n_c} \frac{\left|x_j - m_{ij}\right|}{gate_j}$$

where

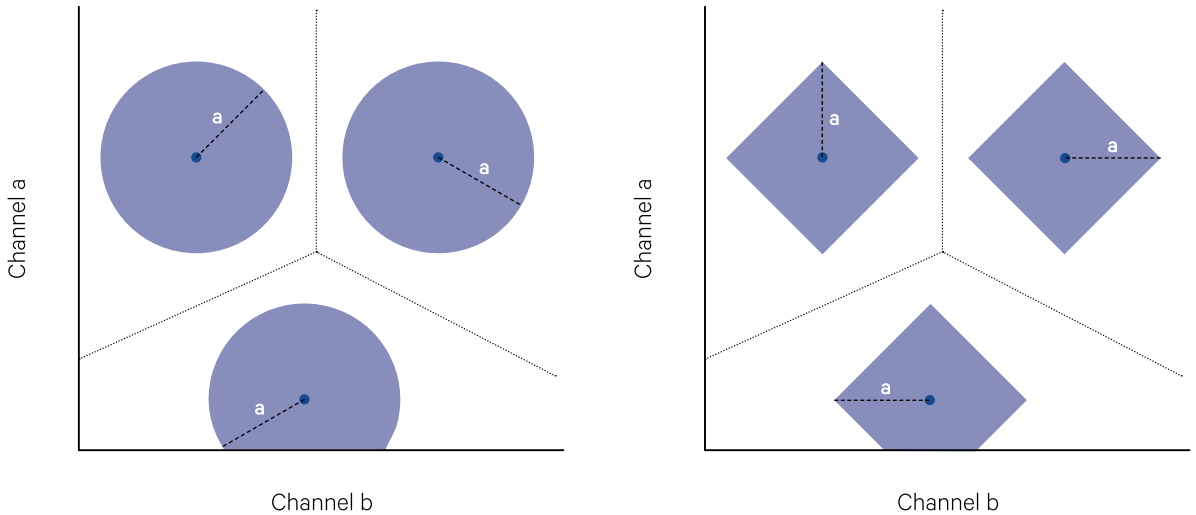
- $\boldsymbol{x}$ is the vector of attributes (or channel values) for the pixel to be allocated,
- D_i is the image attribute decision region for class I,
- x_j is the value of attribute j for the pixel to be allocated
- m_{ij} is the value of attribute j representing class i,
- $gate_j$ is a weight supplied for each attribute or channel, and
- n_c is the number of attributes (or channels) used for the allocation.

Figure 9.17 Distance thresholds for minimum distance classifiers

Different methods can be used to measure distance in image spectral space. Two of the thresholds that are commonly used to determine distance in the minimum distance classifier are:

a. Euclidean distance, where threshold distance = a from the class seed in all directions

b. City block metric, where threshold distance = a from the class seed only in the directions of the channel axes



Source: Harrison and Jupp (1995) Figure 31

Another condition that may be applied during allocation is that the pixel value be within a specified distance of the class seed in each image channel involved in this process (see Figure 9.18a). That is, the attributes must satisfy the condition:

$$L_{ij} < x_j < U_{ij}$$

where

$$L_{ij} = m_{ij} - gate_j \text{ and}$$

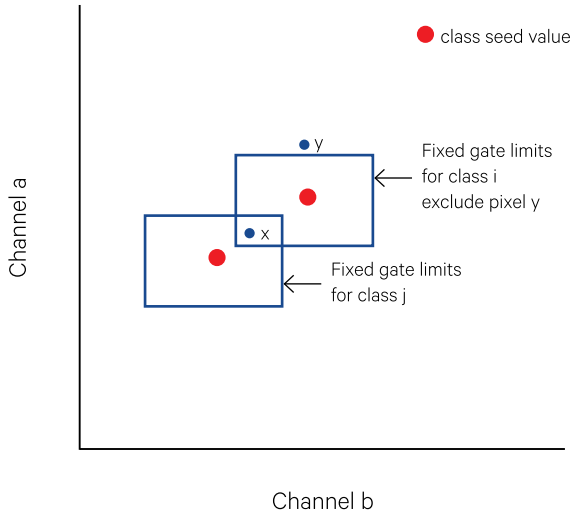
$$U_{ij} = m_{ij} + gate_j$$

For example, the class 'gate' value for a given channel *j* ($gate_j$) may be proportional to the within-class standard deviation in that channel (see Figure 9.18b).

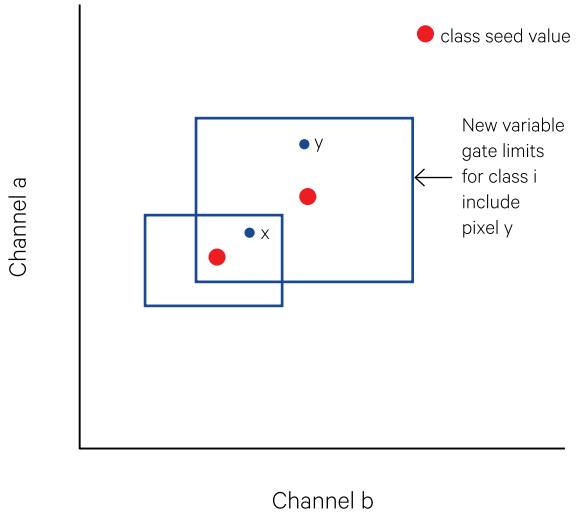
Figure 9.18 Channel gates

Channel gates are shown for two classes, i and j, in relation to two pixels, x and y.

a. Fixed gates use the same gate value for all classes.



b. Variable gates allow different classes to attract outliers at varying distances.



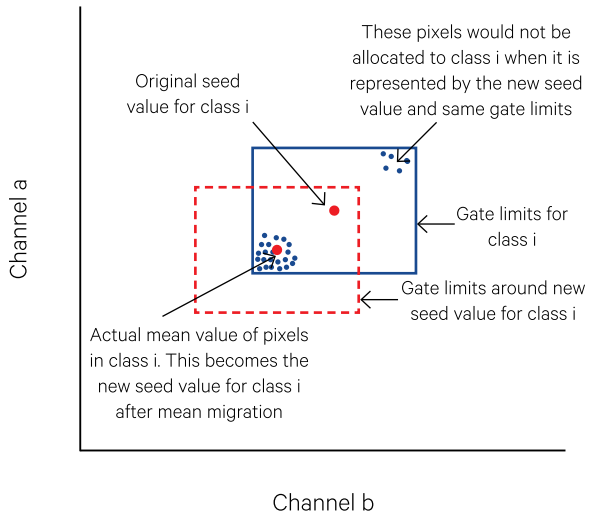
Source: Harrison and Jupp (1995) Figure 29

The results of this classification can be represented as a 'mean' image, in which each classified pixel is given the seed values of its class in each image channel involved in the allocation. For the classification to adequately represent the original image data, the mean image should be similar to the raw data image. Similarity can be measured by a 'residual' image in which the value of each classified pixel is the (shifted) difference between the raw data and mean image values (Jupp and Mayo, 1982). When the mean image has captured the variation of the raw data image, the residual image should appear as random noise. Accordingly, locations where pattern is visible in the residual image may indicate where additional classes are required.

Another image processing operation that is often used in conjunction with the minimum distance classifier is called 'mean migration'. This process is particularly valuable when new seeds have been created automatically, based on a seed value that is not central to its membership. Migrating means, as the name implies, simply involves computing the actual mean for the pixels allocated to a given class, then using this mean as the new seed value. The function of this step is to centre the seed within the current class membership as shown in Figure 9.19. The mean migration process should be followed by another allocation stage so that all pixels are given the 'opportunity' to be associated with the closest class seed, especially when there are several class seeds in close spectral proximity (see Volume 2E—Section 6.2.2).

Figure 9.19 Migrating means process

During mean migration, the actual mean of all pixels allocated to a class is computed. In this example, most pixels allocated to class i are not close to the class seed, so the mean of the pixels in that class 'migrates' the class seed closer to them. In many cases, some pixels originally allocated to that class would no longer be included in it.



Source: Harrison and Jupp (1995) Figure 38

The same process can be used to compute class seed values in terms of a different set of image channels. This 'spectral transfer' operation is particularly useful when a classification developed using a transformed set of channels (such as PCs) is to be transferred back to the original image channels (see Volume 2E—Sections 4.2.4 and 5.3.2).

9.2.4 Maximum Likelihood classifier

Although widely used for classifying EO imagery, the Maximum Likelihood classifier is conceptually more complex and computationally more expensive than most other classification approaches. The foundational statistics for Maximum Likelihood are only briefly introduced below, but the interested reader can discover considerable detail on this topic in Volume 2E—Sections 2 and 7.

While the classifiers described in Sections 9.2.1 to 9.2.3 effectively define decision regions in the image spectral space that correspond to interpreter categories, with multiple decision regions often becoming associated with a single interpreter category during the labelling phase (see Section 9.3), Maximum Likelihood classification relies on probability methods to directly associate interpreter categories with image classes via the following logical steps:

- assume a parametric form for the class distributions;
- estimate the parameters of the distributions from training sets;
- compute the posterior probability that a pixel belongs to each class; then
- allocate that pixel to the class for which it has highest probability of membership.

This method rests on the assumption that pixels known to be in an interpreter category have attributes that are (independent) samples from an underlying probability distribution which is characteristic of that category. Using conditional probabilities (see Volume 2E—Section 2.3.2.1), this distribution may be denoted:

$$P\left(\boldsymbol{x} \mid C_j\right)$$

where $\boldsymbol{x}$ is the set of attributes observed for the pixel. The notation indicates the probability that a pixel from scene class C_j will have the attributes $\boldsymbol{x}$, that is, the probability of a pixel having the values $\boldsymbol{x}$ given the condition that it is a sample from scene class C_j. The overall probability of pixels in the image taking a specific value is assumed to follow the mixture distribution:

$$P\left(\boldsymbol{x}\right) = \sum_{j=1}^{N} q_j P\left(\boldsymbol{x} \mid C_j\right)$$

where the q_j are the relative class densities. The 'reciprocal' probability, which allocates pixels with known attribute values but unknown class membership to one from a set of possible classes, is referred to as the posterior probability or likelihood depending on whether the Bayesian or Likelihood approach is taken (see Volume 2E—Sections 2 and 7). Thus:

$$P\left(C_j \mid \boldsymbol{x}\right)$$

is the probability that a pixel with attribute values $\boldsymbol{x}$ is a member of class j. Bayes' Theorem (see Volume 2E—Section 2.3.2.1) derives this conditional probability as:

$$P\left(C_j \mid \boldsymbol{x}\right) = \frac{q_j P\left(\boldsymbol{x} \mid C_j\right)}{P\left(\boldsymbol{x}\right)}$$

where

q_j is the probability of class j occurring; this can be derived from some additional information about the class, such as the expected frequency of occurrence, and is often referred to as the 'prior' probability;
$P(\boldsymbol{x})$ is the probability of attribute values $\boldsymbol{x}$ occurring in the image and can be estimated from the distribution of all pixels in the image; and
C_j denotes scene class j.

Prior probabilities can be used to introduce non-numerical data into the Maximum Likelihood classifier as a 'cascade' of probabilities. Such implementations of the Maximum Likelihood classifier allow the prior probabilities to be derived from non-image sources, for example a GIS data plane (see Volume 2E—Section 9.5.3).

Since this classifier tends to associate one image class with one interpreter category, the extent to which it is appropriate for a particular project depends heavily on whether the assumptions about class distributions are valid for the image being classified. In general, only the Gaussian distribution is used and the method is often called the 'Gaussian Maximum Likelihood' method. That is (see Volume 2E—Section 2.4):

$$P\left(\boldsymbol{x} \mid C_j\right) = \left(2\pi\right)^{-\frac{n_c}{2}} \left|\boldsymbol{S}_j\right|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\left(\boldsymbol{x}-\mu_j\right)^T \boldsymbol{S}_j^{-1}\left(\boldsymbol{x}-\mu_j\right)}$$

where

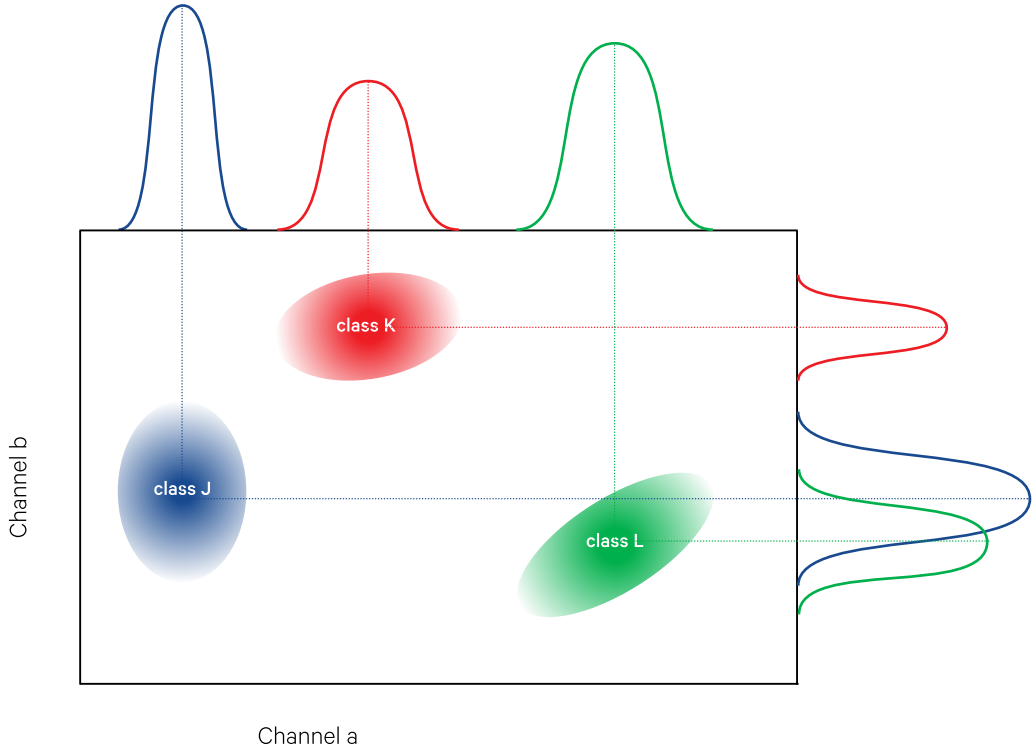
$\boldsymbol{x}$ is the vector of attribute values for the pixel;
C_j denotes scene class j;
n_c is the number of attributes;
$\boldsymbol{S}_j$ is the variance/covariance matrix for class j; and
μ_j is the mean vector for class j.

Other distributions that can be used for Maximum Likelihood classification are discussed in Volume 2E—Section 7.2.3.

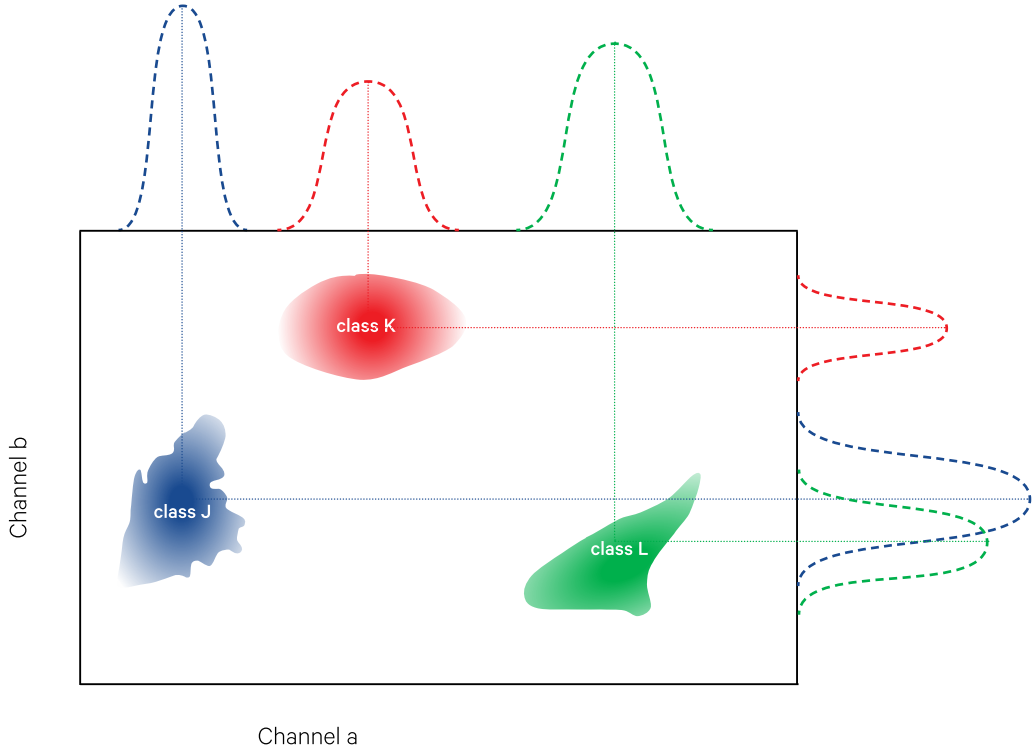
The parameters defining this distribution are generally estimated from training. If some estimates for the class prior probabilities are given, pixels can be allocated to the available classes. With no estimates available for these parameters, they are often taken as equal for each class and normalised to sum to one. The underlying philosophy of this approach assumes that the interpreter categories (such as land cover types) are well represented in the image data as normal distributions around class mean values (see Figure 9.20a). Unfortunately, in most EO images, the actual distributions of image values associated with particular ground features are more complex (see Figure 9.20b) .

Figure 9.20 Maximum Likelihood classification

a. Maximum likelihood classification is based on the premise that image classes can be modelled as statistical distributions.



b. Within a given range of band values, the distribution of class values in EO imagery can be quite irregular.



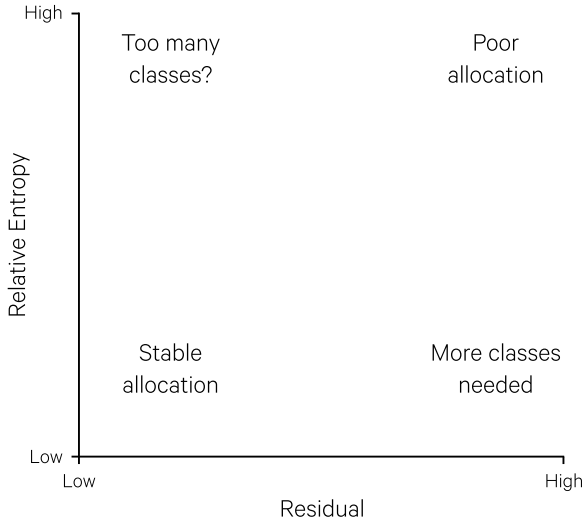
Since Maximum Likelihood classifiers are computationally expensive, especially for large numbers of classes and/or channels, other approaches are often preferred, particularly when combined with post-allocation labelling (see Section 9.3 and Volume 2E—Section 7). Since it is possible that some image pixels do not actually belong in any of the 'available' classes, some versions of the Maximum Likelihood classifier define a rejection probability, below which a pixel will not be allocated to the closest class. Additional statistics may be provided by some implementations of the Maximum Likelihood classifier such as:

- residual—indicates whether a pixel is typical of its allocated class. The residual value is high when a pixel is considered to be a significant 'probability-based distance' from the class mean value, relative to the class distribution;
- relative entropy—indicates whether a pixel has high probability of belonging to several classes; and
- minimum support ratio—equals the difference in probabilities between the allocated class and the next likely class. This highlights those classes whose allocation may change with different prior probabilities.

For example, as illustrated in Figure 9.21, while low 'entropy' and low 'residual' values would indicate stably allocated pixels, those pixels with high 'residual' and high 'entropy' values might be poorly allocated among the available classes. Conversely, pixels with low 'residual' and high 'entropy' values would suggest that there is considerable overlap between classes, and high 'residual' and low 'entropy' values might indicate the need for more classes.

Figure 9.21 Allocation using Maximum Likelihood classifier

Pixels with stable allocations have low residual values and low relative entropy values.



Source: Harrison and Jupp (1995) Figure 34

For more information related to the Maximum Likelihood classifier, please refer to Volume 2E.

9.3 Labelling Classes

The labelling stage involves assigning labels to the allocated spectral classes. The allocation results can commonly be represented as a single classification channel, in which each class is represented by a different numeric value. A classification channel displayed as a grey-scale image can then be interpreted using interactive painting (see Section 9.1). Class labelling is often an interactive process, allowing an interpreter to highlight the spatial extent and continuity of individual spectral classes. It is generally prudent during labelling to associate multiple spectral classes with a single ground-based category.

While the labelling stage implies that the allocation stage has been completed, the iterative nature of the whole classification process necessitates an assessment of each iteration to determine when it has reached completion. Just as statistical tools are available to select the most appropriate image (or transformed) channels (see Section 9.1.3), a range of methods have been developed to highlight where there are too few or too many classes to efficiently capture the intrinsic image variation. These methods are introduced in Section 9.1.2 and discussed in greater detail in Volume 2E. For example, to assess the contribution of a particular class or group of classes to the overall classification, some of the statistics summarised in Section 9.1.2.1 could be computed for classifications with and without those classes (see also Figure 9.12).

9.3.1 Clustering tools

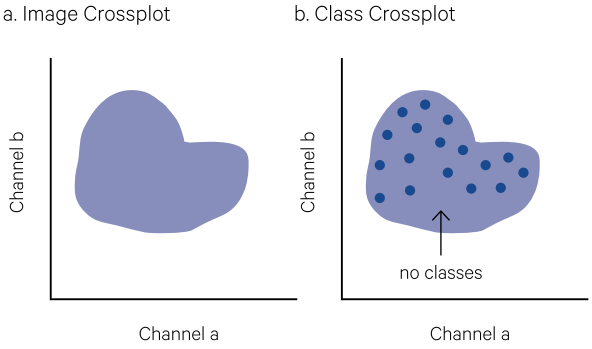
Various spectral and spatial analyses of the classes can also be used at this stage to combine spectral classes into a single interpreter category. It may be appropriate to apply relevant transformations to seed values in order to sort classes by a specific criterion (such as vegetation greenness). Graphically plotting class seed values on crossplots of selected pairs of image (or transformed) channels is a valuable way of viewing the spread of classes in an image classification.

Some of the tools that are generally available for classification analysis include:

- sorting classes by a relevant attribute, such as brightness or greenness. Since most classifiers create class seeds in an arbitrary order, this may require that the classification be spectrally transferred onto channels that are ordered in some recognisable gradient, such as PC1 or a vegetation index.
- crossplotting the values of class seeds using pairs of significant channels (see Section 8.1.3)—comparing crossplots of image channels with crossplots of class seeds can be a useful way of highlighting spectral regions in the image that are not represented by the current set of classes (see Figure 9.22).
- canonical variates analysis (CVA) highlights class separation by maximising the differences between sets of classes relative to the variation within them. As with PCA, the first two canonical variates (CV1 and CV2) are the most informative with respect to class separation. Accordingly, to visualise class distribution and separation it is most useful to plot class seeds on a crossplot of CV1 versus CV2 (see Section 9.1.2.1 and Volume 2E—Section 5.1.3 and 5.3).
- taxonomic clustering of the basic image classes to group them into a given number of interpreter classes (see Section 9.1.2.1). Given the variations that can occur between distance metrics and clustering algorithms (see Volume 2E—Section 6), the first and last (or earliest and latest) aggregations are likely to be the most consistent and reliable.
- minimum spanning tree (MST) orders class pairs by their spectral or spatial proximity (see Section 9.1.2.1):
  - the Mahalanobis distance between class pairs can be used to create a spectral MST (see Section 9.1.2.2); and
  - the co-occurrence matrix, which reports the frequency with which each pair of classes occur as adjacent pixels, shows the spatial similarity of classes. Co-occurrence results can be used to generate a spatial MST in which the linkages indicate spatial association between pairs of classes (see Section 9.1.2.2 and Volume 2E—Section 6.4).

Figure 9.22 Highlighting class deficiencies

Comparing crossplots of image channels with crossplots of class seeds can be a useful way of highlighting spectral regions in the image not represented by the current set of classes.



Source: Harrison and Jupp (1995) Figure 57

9.3.2 Aggregating Classes

The process of aggregating classes, and associating each group of clases with an interpreter label, usually involves interactive image painting of a displayed classification channel to highlight selected classes and potential clusters (see Sections 5.2.1 and 6.1.2.2). Given that most image classifications involve large numbers of classes, it is useful to approach the interactive labelling stage in a structured way using the various clustering tools mentioned above and in Section 9.3.1. At this stage any ancillary data related to the imaged scene and/or the desired classes (such as relevant maps) would be consulted.

Some suggestions to prepare for labelling include:

- order classes so that visual interpretation of the classification channel is intuitive, for example, lighter grey shades would relate to higher reflectance or vegetation greenness;
- plot classes on crossplots of CV1 vs CV2. The relationship between the original channels and each canonical variate is summarised in the CVA transformation matrix (see Volume 2E—Sections 2 and 5). For most EO image classification involving land cover, CV1 is associated with brightness and CV2 is associated with vegetation greenness. Awareness of the gradients along the axes of the CV crossplot will help to label the image classes;
- compute spectral and spatial MST and plot each onto crossplots of CV1 vs CV2 using different colours to represent different spectral/spatial distance ranges. When distance information is included in this way it is obvious to the interpreter which classes are closest;
- compute dendrogram of classes and note the largest and smallest clusters; and/or
- highlight on the MST and/or dendrogram those classes that contain the largest numbers of pixels.

Interactive painting involves display of a classification channel as a grey scale image, then changing selected classes to specific colours. By highlighting a class this way, both its spatial extent and cohesion can be visualised. At this stage, any ancillary data related to the imaged scene and/or the desired interpreter classes (such as relevant maps) would also be consulted. Some guidelines for interactive painting include:

- large classes are best coloured first and labelled using ancillary information so that their region of the MST can be identified;
- using a different colour, the 'closest' class can then be highlighted. If these two classes form coherent patterns, they are likely to belong to the same group. Reference to ancillary information can help to decide when classes should be combined. This process is repeated for all 'adjacent' classes;
- given that classes tend to be linked along environmental gradients in the MST, it is valuable to identify the extremes of the MST early in this process, unless those classes are too small or indistinct; and
- the danger in moving along a particular gradient is that there is a tendency to keep adding classes to one group. It is beneficial to move to a different part of the MST once several classes have been combined.

The result of this process should be that all image classes become associated with several class groups, which are assumed to represent the interpreter classes. The interpreter class groups are often presented as a thematic image and methods for selecting appropriate colours for this are discussed in Excursus 9.2. Those groups can also be recorded as a classification channel, so that the spectral transfer process could be used to compute a mean and residual image for the groups. These images could then be used in the same way as described in Section 9.3.1. to determine whether the aggregated groups are spectrally homogeneous (see Volume 2E—Section 9).

Excursus 9.2—Colour Selection

One of the principal uses of image processing is to identify image features and patterns and highlight their differences and similarities. Colour is frequently used to differentiate classes (or identified numeric relationships) in visual assessment and interpretation. An inappropriate selection of colours for this task can mean that the perceived colour differences do not consistently reflect differences in the numeric values.

The selection of effective colour sequences requires consideration of:

- our perceptual discrimination of colours (see Excursus 5.1);
- their representation on the display/hardcopy device;
- the size and shape of the features being represented;
- our perception of appropriate colours for the ground features being represented;
- the relationship between the different features that we are trying to portray; and
- the interactions between different colours in adjacent features.

In image painting or presentation of classification results, colour selection can be associated with the progression of image values to form natural sequences. The selection of colours using uniform colour spaces can attempt to compensate for the differences between measured and perceived properties of colours. Sequences are best selected using some logic such as following a continuous path in a colour space model and/or restricting colours to planes of constant value or intensity. For example, all colours may have the same hue, or be selected from the sequence of mixtures between two complementary hues. Colours are best spaced at equal perceptual distances in any colour model (as discussed in Appendix 3, linear interpolation between colours in different models produces different results in colour spacing due to the non-linear scaling of the models).

Progressions in the RGB colour space can be easily defined in most image processing systems. The expression of composite colours in terms of their component proportions of the red, green and blue primaries allows a sequence to be developed by varying only one component primary. A number of sequences can also be linked to cover a larger range of colours. For example, starting with a reddish brown, we can increment the amount of red up to the maximum level then add increasing amounts of green to stop at a partially saturated yellow. The colours in this sequence will display a natural progression due to their relationship in the colour space.

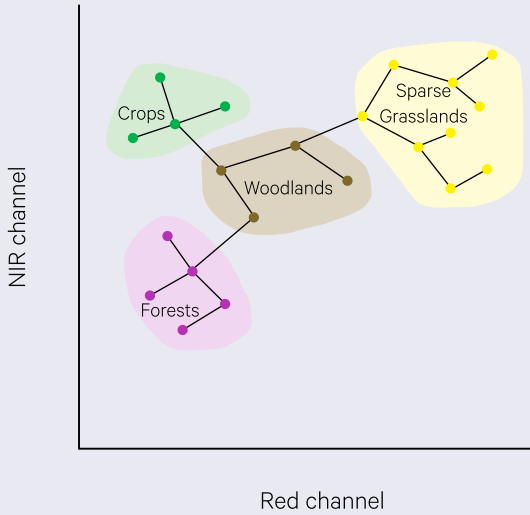
The gradual changes between interpreter categories, such as forest to woodland to open woodland to grassland could be effectively represented as a sequence of related colours such as bright green through darker greens to browns. These sequences may associate specific colour groups with different land covers, for example:

- purples for forest classes;
- browns for woodland classes;
- greens for crop classes; and
- yellows for sparse grasslands and bare ground classes.

The key 'node' points in the MST can be used as the point of changing colour sequences as shown in Figure 9.23. Thus, the MST not only provides a useful base for defining the progression but it can also be used as an effective colour key and implicitly incorporate some spectral (or spatial) information in the descriptive labels. Some methods for using different colour spaces to determine equal perceptual intervals between colours are given in Appendix 3.

Figure 9.23 MST colour sequences.

The minimum spanning tree, which is used for classification labelling in the Mosaic Model approach (see Excursus 9.3 and Volume 2E), can be used to coordinate colour selection for the final classes. In this example, a sequence of yellows are used to map sparse grassland classes, a sequence of browns map woodland classes, a sequence of purples map forest classes and a sequence of greens map crop classes.



Source: Harrison and Jupp (1990) Figure 56

9.3.3 Post-classification processing

To improve the cohesion of a labelled classification, it may be appropriate to apply a 3x3 modal filter to the interpreter group classification channel before final presentation. This process replaces the central value of the filter with the mode (most popular value) of its neighbouring pixels (see Volume 2C), often with constraints such as a minimum number of pixels having the same interpreter class. Modal filtering as a post-classification process is a very effective method of removing small, spurious patches, which enables greater compression of classification results and reduces computational needs for any subsequent GIS processing. Given that for most EO imagery the optical pixel size is larger than the geometric pixel size (see Volume 1), this process should not be viewed as undermining data fidelity, but simply removing imaging artefacts in the classification context. Other options for post-classification processing are discussed in Volume 2E—Section 9.

9.4 Verifying Results

Class labelling based purely on statistical data is rarely satisfactory for a detailed interpretation of EO imagery, so verification with ground data (or 'ground-truthing') within a multi-stage statistical sampling regime is usually required. Most image processing systems contain statistics routines for the final or verification stage of the classification process.

Image processing systems typically provide a number of options for generating sample pixels and comparing the purity of spectral class groupings against the labels supplied by an interpreter. Where an ancillary data channel represents an alternative 'classification', possibly based on GIS data, the classification and an ancillary data channels may also be cross tabulated. This process allows every pixel in the image to be included in the tabulation.

The verification stage effectively checks the degree to which the labelled classification represents ground features. Many analyses cross tabulate the spectral classes versus the labelled classes to determine the degree of agreement. This cross tabulation is commonly called a contingency table. A range of statistics can be derived from the contingency table to summarise the extent to which the spectral classes correspond to the labels groups (see Excursus 9.1). This topic is detailed in Volume 2E—Section 10.

Excursus 9.3—Mosaic Model Classification

Source: Harrison and Jupp (1990)
Further information: Volume 2E; Jupp *et al.* (1985a, 1985b)

Section 9.1 above introduced the concepts of supervised, unsupervised and hybrid or 'mixed mode' classification. The processes involved with each of these approaches are summarised in Figure 9.1, Figure 9.2 and Figure 9.3 respectively. In one version of hybrid classification, the underlying model used for both an EO image and the land cover it represents is the 'mosaic' model (Jupp, 1988). This can be considered as a mixed mode approach to classification (see Figure 9.3). The mosaic model approach provides a sophisticated methodology for flexible class generation, labelling and verification. This model views the image or imaged scene as a composite of coloured patches or 'components' that do not overlap. In this context, for EO imagery the term 'colour' can be considered as a generalised one for spectral signature, but in other data sets it may simply be a set of features or attributes attached to each patch.

Mosaic modelling

Both the scene and an image of the scene can be modelled as a mosaic, although the nature and purpose of the two models may be different. The extent to which the two models can be reconciled defines the value of EO and image processing as a means for inferring the scene 'mosaic' from the image mosaic model.

A patch model of the underlying scene is usually based on a 'logical' assignment of labels or descriptions to the patches. These may be geographical labels or descriptions of contents in terms of land cover categories. The scene model describes the land surface from a particular viewpoint and often for a particular purpose or application. Patches composing the models for real scenes obviously need to be hierarchical in that a 'patch' at one scale, such as an area of open forest, is composed of smaller components at a larger scale, such as tree crowns and areas of bare soil. Similarly, patches at one scale may be part of larger units at a smaller scale.

The patch model of an image considers the pattern and form of the image. Its purpose is to accurately represent the image structure and form rather than incorporate a viewpoint or particular application. The pixel size of an image effectively defines the finest scale for the patch size in the image model. Patches larger than individual pixels are usually required in EO analyses. Geographic mapping typically imposes a logical or descriptive patch structure on the landscape. This logical description of the image may be based on different considerations from the spectral patch structure. One objective of exploratory image analysis is to establish the degree to which they might be brought into coincidence. If possible, this is accomplished by the way logical class labels are assigned to the image-based patches.

Methodology

The mosaic model can be realised by first classifying the image into classes, which effectively creates patches of connected groups of pixels with similar image properties. Given the variability of the real world, these patches are usually rather ragged and spotty, however (if necessary) such problems can be overcome with subsequent processing.

Class seed values may be generated in three ways:

- interactive training;
- theme statistics; or
- unsupervised classification.

The result of the allocation stage is not assumed to be a one-to-one classification of ground features, but should be a mosaic model that represents the spectral (colour) variation in an image precisely. In this case, the 'residual' image (the difference between the raw data image and the 'mean' image—Section 9.2.3) has no significant unmapped features, that is, there would be no coherent features visible in the residual image. Volume 2E describes various statistical analyses that may be used during the generation of the image mosaic model to achieve these requirements.

The Mosaic Model approach effectively combines the processes known as sub-class modelling and post-classification labelling (Haralick, 1976; Strahler *et al.*, 1981). Sub-class modelling, as the name implies, models the image as a set of spectral classes. These classes are defined statistically with the goal of representing the spectral variation of the image. Once pixels have been allocated to these defined sub-classes, the interpreter assigns an appropriate ground feature label to each class in the post-classification labelling stage (see Section 9.3). A major advantage of the mosaic model approach is the flexible way in which the image mosaic can be labelled for different interpretation purposes. Once our image model represents the essential spectral variation of the image, the classes may be aggregated in different ways to indicate different scene features or different levels of detail within the features.

Spectral ambiguities, which cause different ground features to be represented by the same image values, become a problem in the labelling stage. This is effectively the point at which the scene model cannot be reconciled with the image mosaic model by simply using spectral data. In such cases, additional image channels, patch shape or ancillary data may need to be used to resolve ambiguities (see Volume 2E).

For example, two image patches that have the same colour can occur in different contexts and represent different land covers. This commonly occurs in EO imagery for the tropical land covers of mangroves and rainforest. In images that contain both land covers, spectral classification results are typically confused since both cover types represent healthy vegetation that absorbs blue and red wavelengths and strongly reflects near infrared radiation. In this case however, ancillary data such as a 5 m elevation contour could be used to separate the two land covers, which typically occur at different elevations.

The Mosaic Model classification approach can be viewed as a hierarchical model in terms of the level of detail contained within both the image mosaic and the labelled mosaic (or scene interpretation). This method can also be applied iteratively, using progressive aggregation of classes. For example, for an initial classification comprising 100 spectral classes, the first aggregation may reduce the number of classes to 60, then the second to 40, followed by the third and final aggregation resulting in 25 labelled groups.

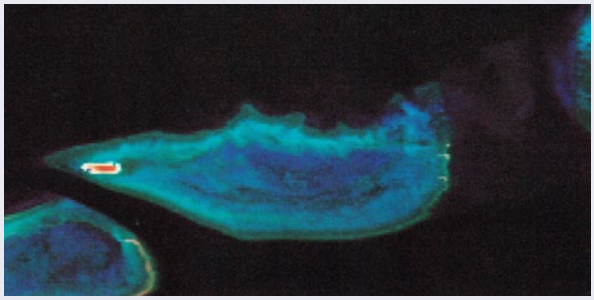
Obviously, a greater level of detail in the image mosaic (indicated by class separability not simply the number of classes) will require more effort during both the classification and labelling stages, while a larger number of reliably labelled groups in the labelled mosaic will require more effort during the labelling stage. The value of this detail however depends on the purpose(s) of the analyses. Volume 2E describes the sequence of operations required to produce and verify a detailed labelled mosaic.

Example classification

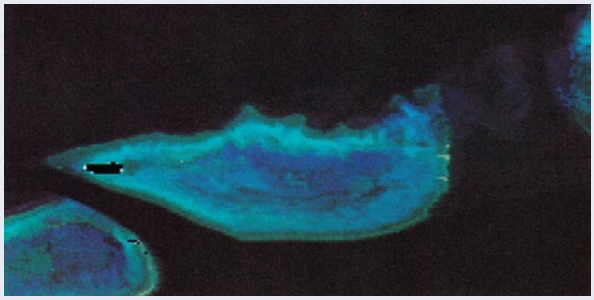
A Landsat MSS image of Heron Island Reef, Queensland, was classified in terms of reef cover classes (Jupp *et al.*, 1985a; Jupp *et al.*, 1985b). The original, mean and residual images for this classification are shown in Figure 9.24.

Figure 9.24 Heron Island Reef example

a. Original image—Landsat MSS image of Heron Island Reef.



b. Mean image—classification into 85 classes with each pixel being represented by the mean class value.



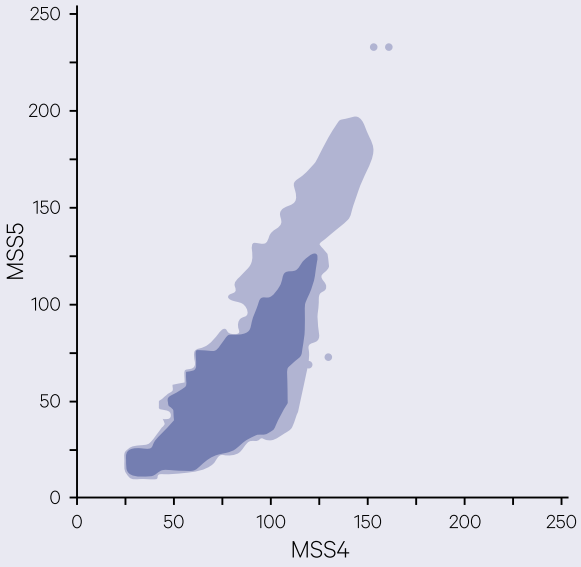
c. Residual image represents the difference between the mean and original images.

Figure 9.25 illustrates the results of image classification using image crossplots which correspond to the original, mean and classified images. (Pixels are actually allocated to classes on the basis of their values in all four image channels but the data variation is shown in only two dimensions here for simplicity of presentation.) The crossplot of the mean image represents the same data range as the original image but with only 85 different image values. The effectiveness of this data reduction stage can be easily seen in the colour similarities between the original and mean images in Figure 9.24a and Figure 9.24b respectively. This data simplification has not significantly reduced the variation in the image data; it simply represents it in a form that can be manipulated more easily.

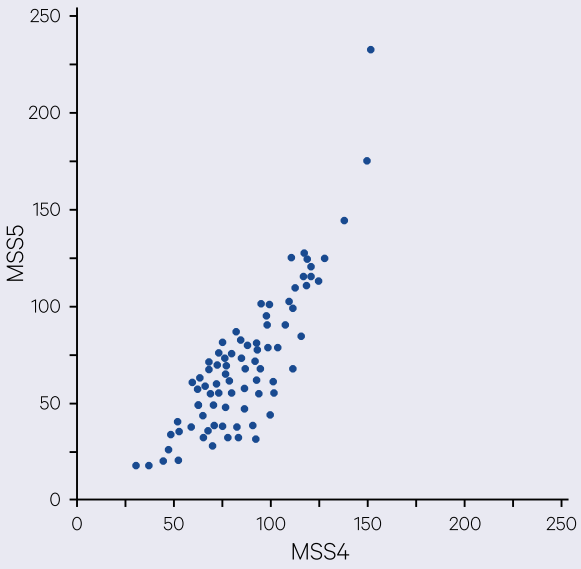
Figure 9.25 Image classification sequence

An example of Mosaic model classification based on a Landsat MSS image of Heron Island Reef.

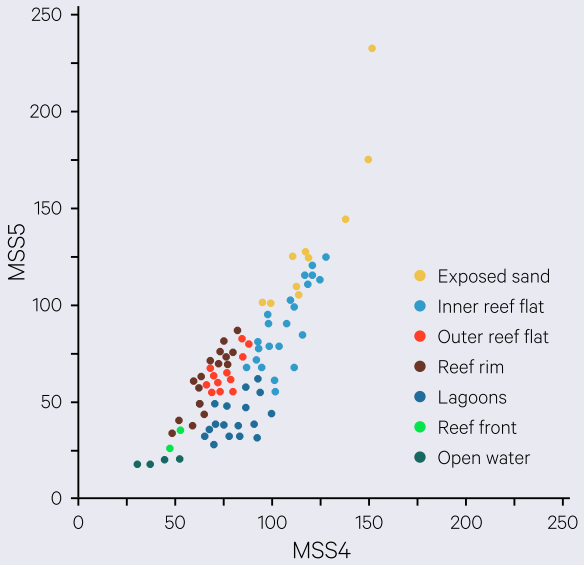
a. A crossplot of bands 4 and 5 in the original image shows the distribution and variation of data values.



b. A crossplot of the mean image (in which each pixel assumes the seed values of the class to which it has been allocated) shows that the classification summarises the distribution of the original image data range but contains less variation. In this case only 85 classes are used.



c. The mosaic model classification methodology allows the 85 classes to be flexibly grouped and labelled. This crossplot shows a labelled mosaic of seven broad categories.

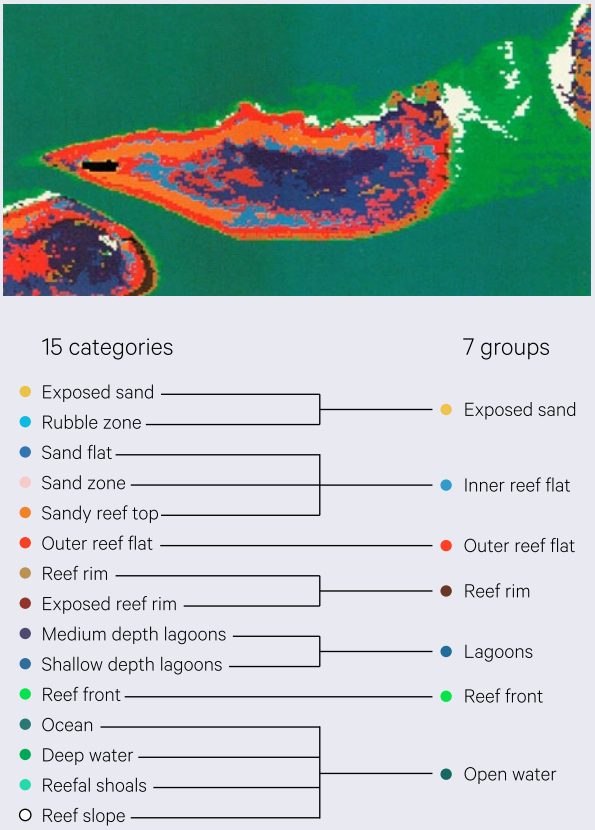


Source: Harrison and Jupp (1990) Figure 54

The labelling stage has reduced the previous 85 spectral classes into 7 major (and 15 minor) categories that represent differences in reef cover type (see Figure 9.26). As introduced in Section 9.3.1 and detailed in Volume 2E, the procedure for aggregating classes can involve statistical analyses of both spectral and spatial similarity between classes in conjunction with interactive labelling. One advantage of the Mosaic Model approach is that these categories could be redefined at a later stage to provide a more, or less, detailed analysis of reef cover, or to represent a different ground feature such as water depth.

Figure 9.26 Classified image

The classification has been labelled into 15 categories which can be further reduced to 7 broad groups.



Source: Harrison and Jupp (1990) Plate 12

9.5 Further Information

Image Classification

SEOS: eLearning Tutorials: *http://www.seos-project.eu/modules/classification/classification-c00-p01.html*

GIS Geography: *http://gisgeography.com/image-classification-techniques-remote-sensing/*

9.6 References

Baker, J. R., Briggs, S. A., Gordon, V., Jones, A. R., Settle, J. J., Townshend, J. R. G., and Wyatt, B. K. (1991). Advances in classification for land cover mapping using SPOT HRV imagery. *International Journal of Remote Sensing*, 12(5), pp. 1071-1085. doi:*http://dx.doi.org/10.1080/01431169108929711*.

Bolstad, P. V., and Lillesand, T. M. (1992). Semi-automated training approaches for spectral class definition. *International Journal of Remote Sensing*, 13(16), pp. 3157-3166. doi:*http://dx.doi.org/10.1080/01431169208904108*.

Bryant, J. (1979). On the clustering of multidimensional pictorial data. *Pattern Recognition*, 11(2), pp. 115-125. doi:*http://dx.doi.org/10.1016/0031-3203(79)90057-8*.

Bryant, J. (1989). A fast classifier for image data. *Pattern Recognition*, 22(1), pp. 45-48. doi:*http://dx.doi.org/10.1016/0031-3203(89)90037-X*.

Bryant, J. (1990). Characterization of the Landsat-7 ETM+ Automated Cloud-Cover Assessment (ACCA) Algorithm. *AMOEBA clustering revisited*, 56, pp. 41-47.

Burrell, J. P. (1973). *Vegetation of Fowlers Gap Station. Lands of Fowlers Gap Station*, New South Wales. Fowlers Gap Arid Zone Research Station Series No. 3 (Ed: J. A. Mabbutt), University of New South Wales, Sydney.

Goldberg, M., and Shlien, S. (1976). A Four-Dimensional Histogram Approach to the Clustering of Landsat Data. *Canadian Journal of Remote Sensing*, 2(1), pp. 1-11. doi:*http://dx.doi.org/10.1080/07038992.1976.10854944*.

Goldberg, M., and Shlien, S. (1978). A Clustering Scheme for Multispectral Images. *IEEE Transactions on Systems, Man, and Cybernetics*, 8(2), pp. 86-92. doi:*http://dx.doi.org/10.1109/TSMC.1978.4309905*.

Haralick, R. M. (1976). Automatic remote sensor image processing. 'Digital picture analysis', Vol. 11, pp. 5-63. Springer-Verlag, Berlin and New York.

Harrison, B. A., and Jupp, D. L. B. (1990). *Introduction to Image Processing. Part TWO of the microBRIAN Resource Manual* (256 pages). CSIRO Australia, Melbourne.

Hashim, M. (1996). Image Classification with Higher-Dimension Watershed Classifier: A Test to Malaysian National Land Use Classification Scheme. *Bulletin Ukur, Penerbitan Akademik Fakulti Kejuruteraan dan Sains Geoinformasi*, 7(2), pp. 141-148.

Jupp, D. L. B. (1988). *The mosaic model for microBRIAN Classification*. Paper presented at the First microBRIAN User Group meeting, Brisbane, Australia.

Jupp, D. L. B., Heggen, S. J., Mayo, K. K., Kendall, S. W., Bolton, J. R., and Harrison, B. A. (1985a). *The BRIAN Handbook. An Introduction to Landsat and the BRIAN (Barrier Reef Image Analysis) System for Users, CSIRO Division of Water and Land Resources, Natural Resources Series No.3*. CSIRO, Canberra, Australia.

Jupp, D. L. B., and Mayo, K. K. (1982). The use of residual images in Landsat image analysis. *Photogrammetric Engineering and Remote Sensing*, 48(4), pp. 595-604.

Jupp, D. L. B., Mayo, K. K., Kuchler, D. A., Heggen, S. J., Kendall, S. W., Haywood, M. J., and Ayling, T. (1985b). *A Landsat-based interpretation of the Cairns Section of the Great Barrier Reef Marine Park. Interpretation of Landsat data by Computer-based Classification and Labelling, CSIRO Division of Water and Land Resources, Natural Resources Series No.4*. CSIRO, Melbourne, Australia.

Lewis, M. M. (1998). Numeric classification as an aid to spectral mapping of vegetation communities. *Plant Ecology*, 136 (2), pp. 133–149.

Mabbutt, J. A., Burrell, J. P., Corbett, J. R., and Sullivan, M. E. (1973). *Land Systems of Fowlers Gap Station*. Lands of Fowlers Gap Station, New South Wales. Fowlers Gap Arid Zone Research Station Series No. 3 (Ed: J. A. Mabbutt), Sydney.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings National Institute of Science, India*, 2(1), pp. 49-55.

McKay, R. J., and Campbell, N. A. (1982a). Variable selection techniques in discriminant analysis: I. Description. *British Journal of Mathematical and Statistical Psychology*, 35(1), pp. 1-29. doi:*http://dx.doi.org/10.1111/j.2044-8317.1982.tb00638.x*.

McKay, R. J., and Campbell, N. A. (1982b). Variable selection techniques in discriminant analysis: II. Allocation. *British Journal of Mathematical and Statistical Psychology*, 35(1), pp. 30-41. doi:*http://dx.doi.org/10.1111/j.2044-8317.1982.tb00639.x*.

McVicar, T. R., Ahmad, W., and Hill, G. (1989). *Landsat Thematic Mapper data for management of Conondale State Forest, Queensland*. Paper presented at the QRSA 1989 Remote Sensing Seminar, Brisbane, Australia.

Mentis, M. T. (1981). Evaluation of the wheel-point and step-point methods of veldt condition assessment. *Proc. Grassland Soc. Southern Africa*, 16, pp. 89–94.

Narendra, P. M., and Goldberg, M. (1977). A non-parametric clustering scheme for landsat. *Pattern Recognition*, 9(4), pp. 207-215. doi:*http://dx.doi.org/10.1016/0031-3203(77)90005-X*.

Rao, C. R. (1948). Tests of Significance in Multivariate Analysis. *Biometrika*, 35(1/2), pp. 58-79. doi:*http://dx.doi.org/10.2307/2332629*.

Richards, J. A. (1986). *Remote Sensing Digital Image Analysis: An Introduction*. Springer-Verlag, Berlin, Germany.

Schachter, B. J., Davis, L. S., and Rosenfeld, A. (1976). Scene segmentation by cluster detection in color spaces. *SIGART Bulletin*(58), pp. 16-17. doi:*http://dx.doi.org/10.1145/1045264.1045267*.

Skidmore, A. K. (1989). Unsupervised training area selection in forests using a nonparametric distance measure and spatial information. *International Journal of Remote Sensing*, 10(1), pp. 133-146. doi:*http://dx.doi.org/10.1080/01431168908903852*.

Strahler, A. H., Franklin, J., Woodcock, C. E., and Logan, T. (1981). *FOCIS: A forest classification and inventory system using Landsat and digital terrain data* (NFAP 255). Nationwide Forestry Applications Program, USDA, Houston, United States.

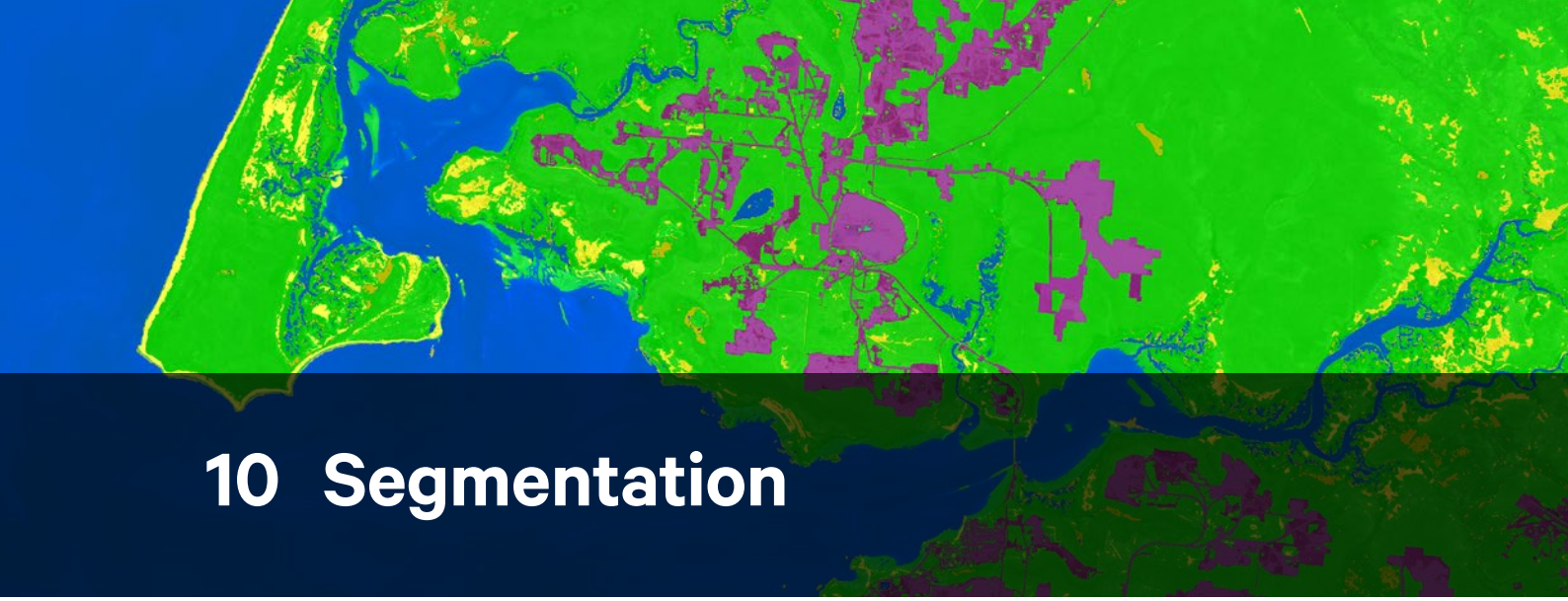
Swain, P. H., and Davis, S. M. (1978). *Remote Sensing: The Quantitative Approach*. McGraw Hill, New York, United States.

Thomas, I. L., Ching, N. P., Benning, V. M., and D'Aguanno, J. A. (1987). A review of multi-channel indices of class separability. *International Journal of Remote Sensing*, 8(3), pp. 331-350. doi:*http://dx.doi.org/10.1080/01431168708948645*.

Watson, A. I. (1987). A new method of classification for Landsat data using the 'watershed' algorithm. *Pattern Recognition Letters*, 6(1), pp. 15-19. doi:*http://dx.doi.org/10.1016/0167-8655(87)90044-4*.

Watson, A. I., Vaughan, R. A., and Powell, M. (1992). Classification using the watershed method. *International Journal of Remote Sensing*, 13(10), pp. 1881-1890. doi:*http://dx.doi.org/10.1080/01431169208904237*.

Young, M. T., and Fu, K.-S. (1986). *Handbook of Pattern Recognition and Image Processing*. Academic Press, Berkeley, United States.



10 Segmentation

Segmenting describes the process of dividing an image into two or more regions on the basis of:

- its spatial patterns (see Section 10.1);
- its pixel values (see Section 10.2); or
- ancillary data (see Section 10.3).

The segmentation may be required before various forms of image analysis, for example to remove cloud or other features that are not of interest prior to classifying an image. Alternatively, stratifying an image into regions, possibly using ancillary data such as cadastral boundaries, allows statistics to be derived for each sub-region. Recent advances in image processing systems allow sophisticated forms of image segmentation to be used in conjunction with standard classification approaches to analyse EO imagery (see Excursus 10.1).

If the regions can be clearly defined on the basis of pixel values, such as cloud or water features (see Figure 4.10), they may be separated using spectral themes. For regions that are visible from spatial patterns in the image (for example, irrigation channels, roads or rivers), the boundaries may also be defined by manually tracing with the screen cursor on a displayed image. When the regions are independent of, or indistinct within, the image, these boundaries may be defined from some ancillary data source, such as a map, using a line-digitising process. To incorporate digitised map boundaries into the image requires that a transformation exists between the coordinate systems of the relevant map and image data. The digitised strings would be converted into image coordinates and then be used to segment the imagery.

Segmentation algorithms analyse the spatial pattern relationships in an image to define boundaries for regions then allocate pixels to those regions based on their position in the image. This process attempts to mimic the role of an aerial photography interpreter and aims for spatially-coherent segments, starting with either:

- a single pixel, then 'growing' a region until boundaries are detected; or
- defined boundary edges, then grouping those pixels they enclose.

In most image processing systems, segmented regions may be identified in two ways in an output image. Pixels in the segments may be set to:

- a 'mask' or null value in all channels (in which case they can be 'ignored' in subsequent processing); or
- a user-specified value in a selected image channel.

Segmenting images using the null value allows regions of the image to be removed from the image and excluded from further processing, as may be required when only a portion of an image is being studied. Alternatively, segmenting with user-defined values allows the creation of ancillary data channels.

10.1 Image Spatial Patterns

Where a clearly visible spatial pattern in the image forms a boundary for image segmentation, cursor tracing of the boundary on the displayed image can be used to define the regions for segmentation. Linear features such as roads, rivers or irrigation channels, which frequently indicate cadastral or management zones, are usually selected for this type of segmentation. Due to 'jitter' problems with human operators, this method is generally less satisfactory than spectral segmentation for features such as coastlines where spectral differences (such as a near infrared channel) could be used to clearly differentiate between land and water.

Background image: Segmentation of Landsat-8 image over Weipa, Queensland, into four major land cover groups: water, vegetation, bare and infrastructure. The original image is shown in the banner of Section 8. **Source:** Norman Mueller, Geoscience Australia

Implementations of spatial segmentation can vary between image processing systems. Some of these options may include:

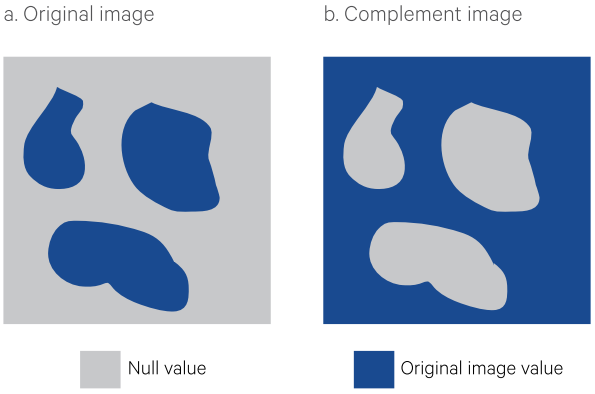
- excluding pixels inside a defined region by setting them to the null value;
- excluding pixels outside a defined region by setting them to the null value; or
- identifying pixels inside a defined region by setting them to a defined value in one or more channels.

Pixels that fall onto the region boundary are typically processed in the same way as those inside the region.

The segmented image can be considered as a mask, in which one or more regions are defined by those pixels set to either a null value or a user-specified value. A segmentation mask can often be complemented as illustrated in Figure 10.1. This process transfers null 'status' from the excluded pixels to the included pixels and sets the excluded pixels to the values in a specified image channel.

Figure 10.1 Complementing an image mask

Pixels that have been excluded from an image (shown in Figure 10.1a) can be complemented to their original value (Figure 10.1b).



Source: Harrison and Jupp (1990) Figure 57

10.2 Spectral Themes

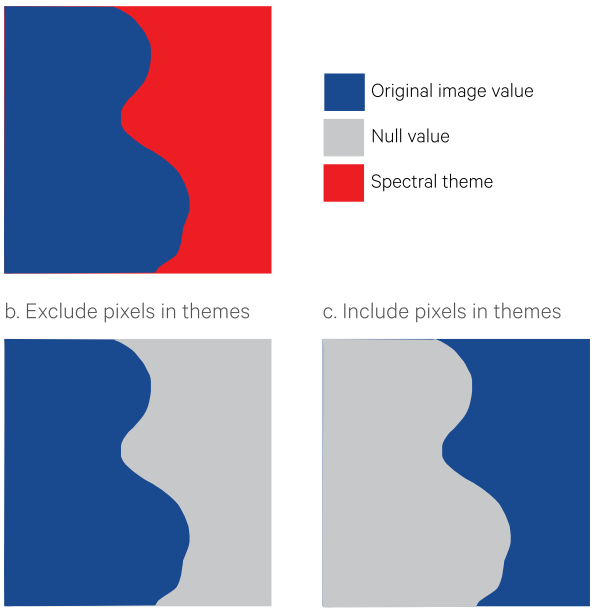
The development of spectral themes to represent particular image features is discussed in Sections 9.1 and 9.2. Spectral segmentation represents a simple form of feature classification where one or more themes can be used to define a feature for the purposes of including it in or excluding it from an image. The themes may define those features that are to be included or excluded from the image, depending on which are easier to develop.

Image segmentation is often an intermediate step in image processing to reduce data volume so it is usually preferable to leave some unwanted pixels in the image rather than remove some required ones. Spectral segmentation may be best effected using particular image channels to define a feature. A mask defined using one set of channels may be transferred to another (registered) set of channels using image pasting as defined in Volume 2D—Section 1.2.

Spatial and spectral segmentation themes can also be combined. For example, spectral classes defined for a deep water feature may also detect land pixels of shadows in steep terrain. In this case a spatial mask can be used to remove the land pixels, often without needing to be accurately located.

Most image processing systems allow images to be segmented spectrally by setting selected pixel values to the null value in all channels of the output image (see Figure 10.2). Typically, pixels that are either included in, or excluded from, the defined themes can be replaced with the values of unmasked pixels. In addition, the complement of the digitised image can often be computed.

Figure 10.2 Including and excluding regions in spectral segmentation



Source: Harrison and Jupp (1990) Figure 58

10.3 Ancillary Data Boundaries

When the boundaries required for image segmentation are defined by some ancillary data source, such as a map, the polygonal region boundaries may be digitised and registered with the image for segmentation purposes. The process required for registration of map and image data is introduced in Section 7 and detailed in Volume 2B. Once a registration model exists, the digitised boundary can be converted to image coordinates and incorporated as an ancillary data channel in the image.

Typically, either one selected channel, or all channels, of an image are processed. Many image processing systems offer the choice of:

- including, excluding or separating from each polygonal region those pixels that constitute its boundary. If the boundary string pixels are to be separated, a unique value can often be specified for them in the output image;
- treating areas of overlap between regions in different ways, such as retaining either the first or last region value, or setting them to ('flagging as') a particular value; and/or
- 'cleaning' boundaries prior to segmentation by removing any closed regions smaller than a specified size.

Pixels in each region may be assigned values in a number of ways as illustrated in Figure 10.3. This assumes that each region has a value associated with it in the vector file, which can simply be assigned to the region pixels to create an ancillary data channel (see Figure 10.3a). If the pixels in the regions were copied from the input images into a new null image, the resulting channels would only include non-null values inside the polygonal regions (see Figure 10.3b). Alternatively, to exclude the polygonal regions from the output image, only those pixels would be set to the null value (see Figure 10.3c).

Figure 10.3 Image segmentation with ancillary data boundaries

a.Set region pixels to another value

b. Copy region pixels from image

c. Set region pixels to null value

Source: Harrison and Jupp (1990) Figure 59

Regions can only be defined by closed boundary vector data. Where ancillary data are in the form of (non-closed) lines or (isolated) points, image pixels on the lines or at the points can often be assigned to a particular value. Where point or line data are sufficiently dense, filter interpolation may be used to fill in any missing pixel values and create a continuous image channel (see Section 10.5 or Volume 2C). Typically, pixels that coincide with multiple points and/or line locations in a vector file can only be set to the first or last point/line value or given a separate value. If this overlapping occurs frequently it may be worth initially registering the ancillary data to a finer image grid to separate the locations as different pixels. In the case of non-nominal data, image blocking could be used to average the values into a larger pixel size (see Section 7.2.1.4).

10.4 Recombining Segments

The complementary operation to image segmentation allows 'null' regions to be transferred between images for the purposes of segmentation or recomposition ('cutting and pasting'). This process naturally assumes that both input images are registered and have the same size. The null mask may be defined by one or more channels in the input image(s) and will only be transferred to the same channel(s) in the output image. Such processes are further discussed in Volume 2D—Section 1.2.

10.5 Cleaning Boundaries

The boundaries of segmented regions may be 'cleaned' using various filtering operations (see Volume 2C). In particular, Section 9 in Volume 2C describes the process of interpolating selected 'missing' values. In conjunction with thickly spaced, but discontinuous, ancillary data points or lines, this operation can create a continuous data channel.

Alternatively, isolated pixels of a selected value (usually equal to the null value by default) may be iteratively removed using a modal filter option. As mentioned in Section 10.3, image processing systems often allow ancillary data boundaries to be 'cleaned' before merging with image data.

Excursus 10.1—Geographic Object-Based Image Analysis

Source: Guy Byrne, Geoscience Australia

Geographic Object-Based Image Analysis (GEOBIA) represents a synthesis of the approaches to land cover mapping used by traditional visual interpretation of imagery and statistically-driven, pixel-based image analysis. A class or object can be described based on a hierarchy of relationships to other classes within the scene, and this hierarchy can be multidimensional. Hay and Castilla (2008) and Blaschke *et al.* (2014) suggest that GEOBIA's ability to define an image model based on the explicit specification of its member elements and their interrelationships (or 'ontology') constitutes a paradigm shift in digital image analysis.

In addition to standard spectral classification descriptors, some discriminates that can be used to describe the relationship between objects and classes within the Definiens GEOBIA Software include:

- geometry (extent, shape);
- position (location, within specified region or not);
- texture; and
- relations to neighbour objects, sub-objects and super-objects (existence of, number of, border to, relative area of, distance from, overlap, aggregration indices).

Using visual interpretation, a landscape is mapped into meaningful units based on the analyst's *a priori* understanding of the scene. That understanding is fundamentally based on the topological relationships between the visible elements within the scene. More recently, pixel-based image analysis uses a variety of statistical and machine learning engines to partition an image into a number of spectrally distinct classes, which are then labelled according to their ontological characteristics.

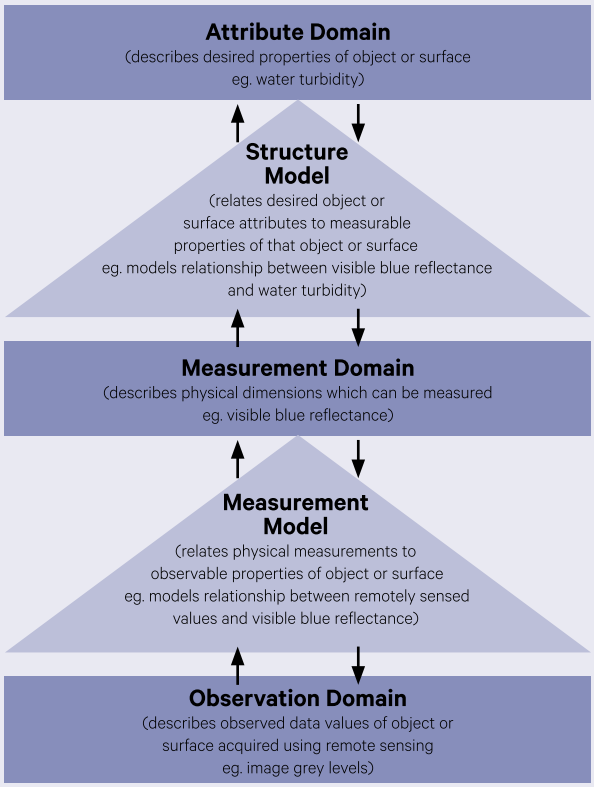
Figure 10.4 shows a conceptual model of the relationship between the observation and attribute domains relevant to EO image interpretation (for a full discussion of this model, please refer to Volume 1A—Section 1). The interplay between the scales of objects, or features, in the landscape and the resolution of the EO image fundamentally defines the sorts of question an interpreter can ask of the data. GEOBIA is most effective when the image pixel size is smaller than the image features to be identified. This scaling relationship between pixels and features has been described as the H-resolution sensor scene model (Strahler *et al.*, 1986) and is depicted in Figure 2.12c (see also Volume 1B—Section 2.2).

To be useful, segments must be measurable, substantial, accessible, differentiable and accountable.
(Philip Kotler)

Using GEOBIA, a classification model is iteratively built through a hierarchy of binary classifications. A feature of interest is 'assigned to a class' by identifying single, or combinations of, attributes based on its spectral, spatial and geometrical properties and ontological relationships. A simplified conceptual model of a typical GEOBIA workflow is shown in Figure 10.6.

Figure 10.4 Conceptual model of domains relevant to EO image interpretation

Interpreting imagery from EO sensors is an indirect process, whereby a measurement model is used to transform EO observations to measurements of some measurable property, and a structure model is used to relate those measurements to application-specific attributes.



Adapted from: Harrison and Jupp (1989) Figure 33

Geographic Object-Based Image Analysis (GEOBIA) is a sub-discipline of Geographic Information Science (GIScience) devoted to developing automated methods to partition remote sensing imagery into meaningful image-objects, and assessing their characteristics through spatial, spectral and temporal scales, so as to generate new geographic information in GIS-ready format. (Hay and Castilla, 2008)

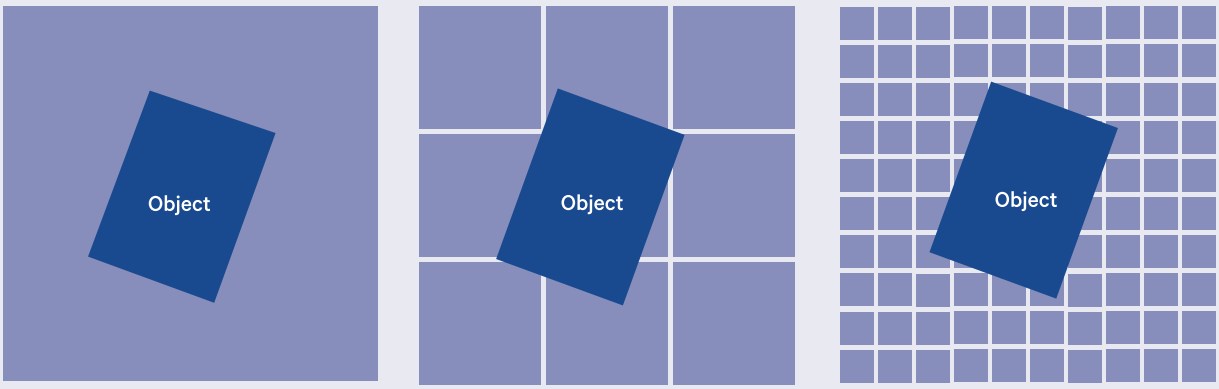
Figure 10.5 Image sampling resolution relative to imaged objects

Depending on image spatial resolution, a single object can be imaged by a single pixel, several pixels or many pixels:

a. When the object being imaged is smaller than an image pixel, then sub-pixel techniques are most appropriate;

b. When the size of the object being imaged is similar to the pixel size, then pixel-by-pixel methods are relevant; or

c. When the object being imaged is much larger than image pixels, object-based techniques are most appropriate.



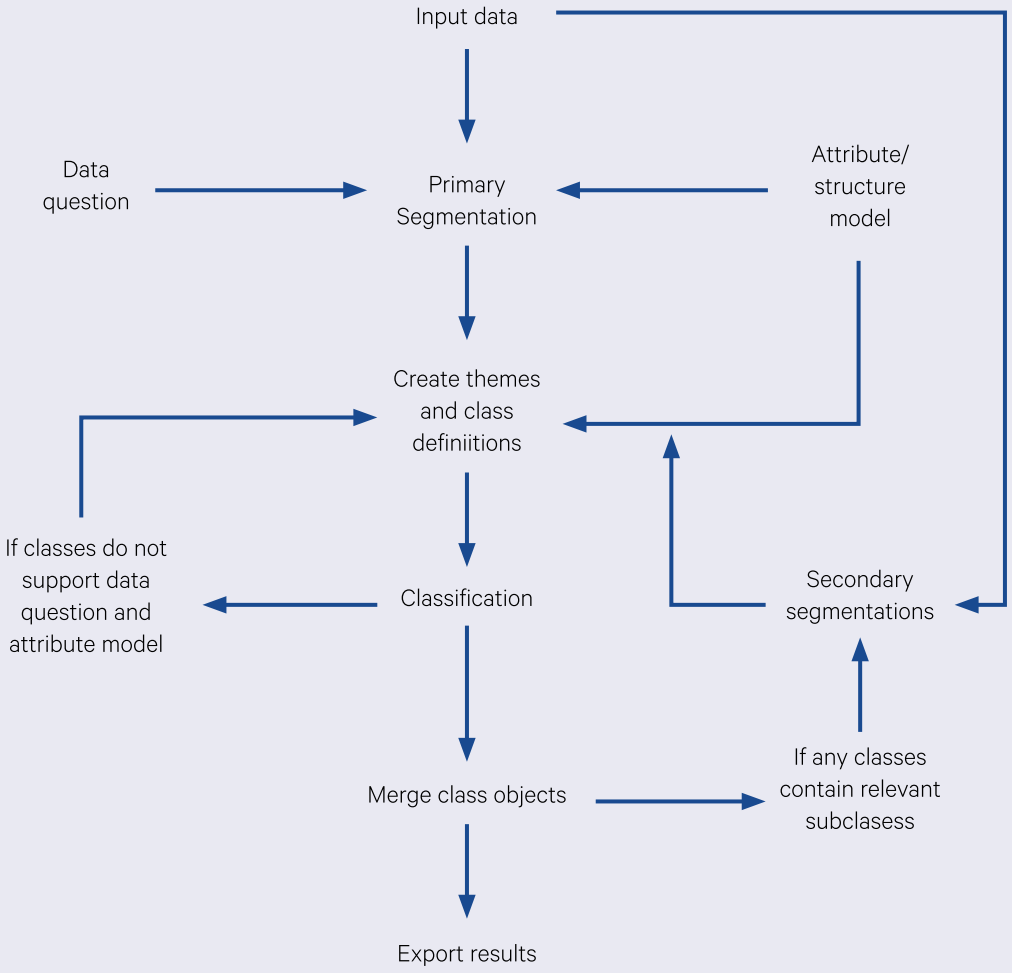
Single 30 m pixel

Nine 10 m pixel

One hundred 1 m pixels

Figure 10.6 Typical GEOBIA classification workflow

This classfiication workflow is similar to the hybrid classification model introduced in Section 9 (see Figure 9.3). In this case, the input data can include EO image bands, GIS shape files and raster surfaces.



The GEOBIA processing model is particularly useful for mapping features that are spectrally similar. For example, vegetation growing at the margins of forests is often spectrally confused with riparian vegetation. However, by including consideration of the shape, size and location of these two types of vegetation, they can be more easily distinguished.

Using a hypothetical example image of Jervis Bay, NSW (see Figure 10.7), the GEOBIA processing sequence to separate spectrally similar features would be:

- create an initial set of objects using a primary segmentation (see Figure 10.8);
- use a NIR spectral theme to partition the objects into groups that represent land and water then merge contiguous objects within these groups (see Figure 10.9);
- to differentiate specific vegetation types, define a secondary segmentation using criteria most appropriate to the land group (see Figure 10.10); then
- progressively classify the land group into a series of sub-classes until all the spectrally-distinct and relevant scene model covers are assigned.

Figure 10.7 Example image

180-day geomedian product over Jervis Bay, derived from Landsat-8 imagery acquired between 1 January 2016 and 28 June 2016, displayed using bands 5,4,3 as RGB.

Figure 10.8 Primary segmentation

This initial segmentation created 110,850 objects for land and 17,469 objects for water within the image shown in Figure 10.7. A subset of this image is shown here for clarity.

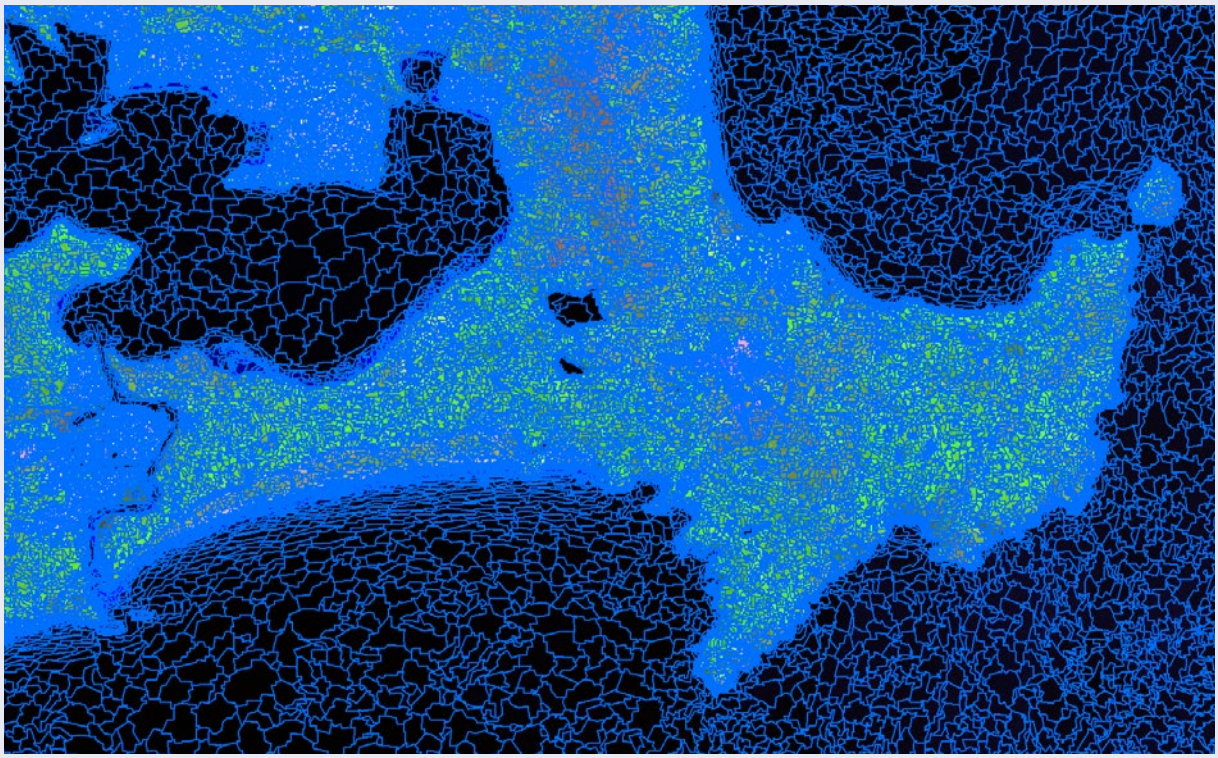


Figure 10.9 Classification of land and water

The segments defined by the primary segmentation stage are merged into two distinct groups that separate land from water. At this stage, the land group (grey) comprises 27 contiguous objects and the water group (aqua) comprises 260 contiguous objects.

Figure 10.10 Secondary segmentation

A secondary segmentation process uses a set of attributes (such as appropriate spectral bands and scaling parameters) that are relevant to the land group to define specific classes. Similarly, to classify the water group, the most appropriate attributes would be used. In this case, 49,563 objects were created in the land group and 1,127 objects were created in the water group. Individual classes can then be refined using new criteria until their mappings are deemed satisfactory. Again a subset is shown for clarity.

In the case of the spectrally-similar forest margin and riparian vegetation classes, the topological landscape relationships unique to each vegetation group need to be understood and then described using sets of rules that 'pull' the mixed spectral class apart. For example, since forests often grade into pasture, an adjacency metric may be a useful class descriptor. In other locations proximity to water or streamlines could be used. As an intermediate stage, a number of smaller sub-classes may be required to fully describe the geospatial relationships of the riparian and forest boundary classes. Such classes could then be merged to define the final set of classes prior to validation.

GEOBIA has been used successfully for many applications including mapping mangroves (Kamal *et al.*, 2015), seagrass (Roelfsema *et al.*, 2014), and coral reefs (Phinn *et al.*, 2012), and monitoring mine site rehabilitation (Bao *et al.*, 2014).

10.6 Further Information

Geographic Object-Based Image Analysis

Blaschke *et al.* (2014)

Johansen *et al.* (2010)

10.7 References

Bao, N., Lechner, A. M., Johansen, K., and Ye, B. (2014). Object-based classification of semi-arid vegetation to support mine rehabilitation and monitoring. *Journal of Applied Remote Sensing*, 8(1), pp. 083564. doi:*http://dx.doi.org/10.1117/1.JRS.8.083564*.

Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Feitosa, R. Q., van der Meer, F., van der Werff, H. M. A., van Coillie, F., and Tiede, D. (2014). Geographic Object-Based Image Analysis – Towards a new paradigm. ISPRS *Journal of Photogrammetry and Remote Sensing*, 87, pp. 180–191. doi:*http://dx.doi.org/10.1016/j.isprsjprs.2013.09.014*.

Harrison, B. A., and Jupp, D. L. B. (1990). *Introduction to Image Processing. Part TWO of the microBRIAN Resource Manual* (256 pages). CSIRO Australia, Melbourne.

Hay, G. J., and Castilla, G. (2008). Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. Chapter 1.4 in '*Object-Based Image Analysis*. Lecture Notes in Geoinformation and Cartography.' (Eds: Blaschke T., Lang S., and H. G.J.). Springer, Berlin, Heidelberg.

Johansen, K., Bartolo, R., and Phinn, S. (2010). Special Feature—Geographic Object-Based Image Analysis. J. *Spatial Science*, 55(1), pp. 3–7. doi:*http://dx.doi.org/10.1080/14498596.2010.494653*.

Kamal, M., Phinn, S., and Johansen, K. (2015). Object-based approach for multi-scale mangrove composition mapping using multi-resolution image datasets. *Remote Sensing*, 7(4), pp. 4753–4783.

Phinn, S., Roelfsema, C., and Mumby, P. J. (2012). Multi-scale Object Based Image Analysis for Mapping Geomorphic and Ecological Zones on Coral Reefs. J. *Remote Sensing*, 33(12), pp. 3768–3797. doi:*http://dx.doi.org/10.1080/01431161.2011.633122*.

Roelfsema, C., Lyons, M., Kovacs, E. M., Maxwell, P., Saunders, M. I., Samper-Villarreal, J., and Phinn, S. (2014). Multi-Temporal Mapping of Seagrass Cover, Species and Biomass: A Semi-Automated Object Based Image Analysis Approach. *Remote Sensing of Environment*, 150, pp. 172–187. doi:*http://dx.doi.org/10.1016/j.rse.2014.05.001*.

Strahler, A. H., Woodcock, C. E., and Smith, J. A. (1986). On the nature of models in remote sensing. *Remote Sensing of Environment*, 20(2), pp. 121–139. doi:*http://dx.doi.org/10.1016/0034-4257(86)90018-0*.


crc•si

CSIRO

Australian Government
Geoscience Australia

bushfire&natural
HAZARDSCRC