

# **Earth Observation: Data, Processing and Applications**

Volume 2X: Processing—Appendices



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](mailto:info@crcsi.com.au).

**Publisher:**

Australia and New Zealand CRC for Spatial Information

**ISBN [ONLINE]:**

978-0-6482278-4-7

**Copyright:**

All material in this publication is licensed under a Creative Commons Attribution 4.0 Australia Licence, save for content supplied by third parties, and logos. Creative Commons Attribution 4.0 Australia Licence is a standard form licence agreement that allows you to copy, distribute, transmit and adapt this publication provided you attribute the work. The full licence terms are available from <https://creativecommons.org/licenses/by/4.0/legalcode>. A summary of the licence terms is available from <https://creativecommons.org/licenses/by/4.0/>.



**Disclaimer:**

While every effort has been made to ensure its accuracy, the CRCSI does not offer any express or implied warranties or representations as to the accuracy or completeness of the information contained herein. The CRCSI and its employees and agents accept no liability in negligence for the information (or the use of such information) provided in this report.

**Recommended Citation for Volume 2X:**

CRCSI (2018). *Earth Observation: Data, Processing and Applications. Volume 2X: Processing—Appendices*. (Eds. Jupp, D.L.B., and Harrison, B.A.) CRCSI, Melbourne.

# Acknowledgements

Production of this series of texts would not have been possible without the financial support of CSIRO, CRC SI, GA and Digital Concepts, 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)<sup>1</sup>. Many illustrations and some text from these publications have been reproduced with permission from CSIRO.

Other contributors are gratefully acknowledged:

- reviewers: Megan Lewis, Stuart Finn and Tony Sparks provided comments on early drafts of selected appendices;
- illustrations: Norman Mueller, Andreas Hüni, Linda Skoog (EagleView);
- 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.

All volumes in this series are covered by the copyright provisions of CC BY 4.0 AU.

---

<sup>1</sup> 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.

# Table of Contents

Acknowledgements

i

## Appendix 1 Brightness Correction Models

1

### A1.1 Empirical Methods

1

### A1.2 Measurement Model Approach

3

#### A1.2.1 Atmospheric correction and reflectance normalisation

3

#### A1.2.2 Analytic BRDF models

4

### A1.3 Semi-empirical Methods

7

#### A1.3.1 Theoretical description

7

#### A1.3.2 Kernels

7

#### A1.3.3 Kernel-driven models

9

#### A1.3.4 Modified Walthall model

10

#### A1.3.5 Advantages of linear models

10

### A1.4 The Issue of Scale

10

### A1.5 References

11

## Appendix 2 BRDF Models

15

### A2.1 Kernel Models

15

### A2.2 Primary AMBRALS Kernel Models

16

#### A2.2.1 Ross-thin (vol) model

16

#### A2.2.2 Ross-thick (vol) model

16

#### A2.2.3 Roujean (geo) model

16

#### A2.2.4 Li-sparse (geo) model

16

#### A2.2.5 Li-dense (geo) model

16

#### A2.2.6 Cox-Munk (geo) model

17

#### A2.2.7 Walthall (vol) model

17

#### A2.2.8 MODIS products

17

### A2.3 Extensions and Alternatives

17

#### A2.3.1 Reciprocal Li models

17

#### A2.3.2 RPV model

18

#### A2.3.3 MRPV model

18

#### A2.3.4 Staylor and Suttles model

18

#### A2.3.5 Shibayama and Weigand model

18

#### A2.3.6 Dymond and Qi model

18

#### A2.3.7 Chen modification to Roujean model

18

#### A2.3.8 Liang modification to Walthall model

19

#### A2.3.9 Pickup and Chewings model

19

### A2.4 Slightly More Complex Models: Not in 'Kernel' Form

19

#### A2.4.1 Otterman model

19

#### A2.4.2 Verstraete, Pinty and Dickinson (VPD) model

20

#### A2.4.3 Strahler and Jupp simple model

20



<b>A2.4.4</b>	Li's 'top easy seen' modification	21
<b>A2.4.5</b>	Hapke BRDF model for soils	21
<b>A2.4.6</b>	Ross' simplified vegetation canopy formula	22
<b>A2.4.7</b>	Minnaert's original model	22
<b>A2.5</b>	References	23
<b>Appendix 3</b>	<b>Colour Systems and Sequences</b>	<b>25</b>
<b>A3.1</b>	Colour Measurement	25
<b>A3.2</b>	Colour Systems	26
<b>A3.2.1</b>	RGB and HSI	26
<b>A3.2.2</b>	XYZ and UVW	27
<b>A3.3</b>	Colour Coordinates	28
<b>A3.4</b>	References	28
<b>Appendix 4</b>	<b>Dithering</b>	<b>29</b>
<b>A4.1</b>	Dither Matrices	29
<b>A4.2</b>	Designed Order	30
<b>A4.3</b>	References	32
<b>Appendix 5</b>	<b>Filters</b>	<b>33</b>
<b>A5.1</b>	Averaging Filters	33
<b>A5.2</b>	Directional Filters	35
<b>A5.2.1</b>	Exposure	36
<b>A5.2.2</b>	Relief shading	36
<b>A5.2.3</b>	Curvature	37
<b>A5.3</b>	Geophysical Filters	38
<b>A5.4</b>	References	38
<b>Appendix 6</b>	<b>Matrix Algebra</b>	<b>39</b>
<b>A6.1</b>	Geometric Effect of Matrix Operations	39
<b>A6.2</b>	Matrix Arithmetic and Properties	43
<b>A6.2.1</b>	Trace	43
<b>A6.2.2</b>	Matrix equality	43
<b>A6.2.3</b>	Matrix addition	43
<b>A6.2.4</b>	Scalar multiplication	43
<b>A6.2.5</b>	Matrix multiplication	43
<b>A6.2.6</b>	Transpose of a matrix	43
<b>A6.2.7</b>	Additive inverse of a matrix	43
<b>A6.2.8</b>	Multiplicative inverse of a matrix	43
<b>A6.2.9</b>	Determinant of a matrix	44
<b>A6.2.10</b>	Characteristic equation	45
<b>A6.2.11</b>	Eigenvalues of a matrix	45
<b>A6.2.12</b>	Eigenvectors of a Matrix	46
<b>A6.3</b>	Derivation of the Principal Components Transformation	46
<b>A6.4</b>	References	48

<b>Appendix 7 Predictive Error and Variance</b>	<b>49</b>
<b>A7.1</b> Predictive Error for Least Squares	49
<b>A7.2</b> Generalised Cross-Validation (GCV)	51
<b>A7.3</b> Singular Value Decomposition	51
<b>A7.4</b> Predictive Variance	52
<b>A7.5</b> References	52
<b>Appendix 8 ‘Sieving’ Outliers</b>	<b>53</b>
<b>A8.1</b> Basic Assumptions and Framework	53
<b>A8.2</b> Fitting Transformation Models	53
<b>A8.3</b> The SIEVER Model	54
<b>A8.4</b> Solving for Underlying Model and Errors	55
<b>A8.5</b> Analysing the SIEVER Solution	56
<b>A8.5.1</b> Basic ‘PCA’ summary	56
<b>A8.5.2</b> Row or feature-based error analysis	57
<b>A8.6</b> Using the Q-Q Plot	57
<b>A8.6.1</b> Quantile-Quantile plot	57
<b>A8.6.2</b> Gamma probability plotting	58
<b>A8.6.3</b> Interpreting the Q-Q plot	58
<b>A8.7</b> Data Plots to Indicate Outliers	59
<b>A8.7.1</b> Listing and plotting the Matrix <i>US</i>	59
<b>A8.7.2</b> Listing and plotting the Error Matrix	59
<b>A8.8</b> Using SIEVER to Locate Outliers	59
<b>A8.9</b> References	60

# List of Figures

<b>Figure A2.1</b>	Geometry of remote observer	15
<b>Figure A3.1</b>	RGB colour coordinates	26
<b>Figure A3.2</b>	HSI coordinates	27
<b>Figure A3.3</b>	CIE rg chromaticity space	27
<b>Figure A3.4</b>	Perceptual colour differences	28
<b>Figure A4.1</b>	Dithering example	29
<b>Figure A4.2</b>	Dithering sequence	29
<b>Figure A4.3</b>	Example dither matrices	30
<b>Figure A4.4</b>	Formation of a balanced dither matrix	31
<b>Figure A4.5</b>	Colour representation using dithering techniques with a 2×2 dither.	31
<b>Figure A5.1</b>	Averaging filters	34
<b>Figure A5.2</b>	Derivative filters	35
<b>Figure A5.3</b>	Directional, edge-detecting filters	35
<b>Figure A5.4</b>	Directional filters	36
<b>Figure A5.5</b>	Kirsch directional filters	36
<b>Figure A6.1</b>	Effect of matrix operation on data geometry	40
<b>Figure A6.2</b>	Reflecting axes	40
<b>Figure A6.3</b>	Reflecting axes about the line $Y=X$	41
<b>Figure A6.4</b>	Rescaling axes	41
<b>Figure A6.5</b>	Rotating axes	42
<b>Figure A6.6</b>	Skewing axes	42
<b>Figure A6.7</b>	Effect of matrix operation on unit area	44
<b>Figure A6.8</b>	Effect of singular matrix operation	45
<b>Figure A6.9</b>	Example image channels	46
<b>Figure A8.1</b>	Q-Q plot	58

# List of Tables

<b>Table A1.1</b>	Semi-empirical weight formulas	9
<b>Table A6.1</b>	Principal component percentage variation	47



# List of Excurses

**Excursus A1.1**—Geometric Optical Models

5





# Appendix 1

## Brightness Correction Models

The need for image calibration is introduced in Volume 2A—Section 3 and is particularly important for multi-date or multi-sensor image datasets (see Volume 2D). The general approach to image brightness variation is to ‘correct’ it in some way. Correction means changing a pixel spectrum so that it is what would have been measured if the Sun and sensor geometry were a fixed standard pair rather than a varying pair of angles over the image extent. As discussed later, this action may be valid as an average over a spatial area but can be a difficult concept at the scale of a single tree crown or soil clump.

We will outline three main approaches to scene brightness modelling and ‘correction’, namely:

- empirical methods (see Appendix A1.1);
- measurement modelling methods (see Appendix A1.2); and
- semi-empirical methods (see Appendix A1.3)<sup>2</sup>.

### A1.1 Empirical Methods

Empirical methods involve fitting the variation in general brightness ascribed to the atmosphere and Bidirectional Reflectance Distribution Function (BRDF) effects by a parametric function (using least squares or some other criterion) and then normalising by some combination of additive and multiplicative image transformation. The simplest methods involve fitting polynomials to average variations over scan lines or frames and subtracting or dividing out the general level to ‘flatten’ the data. These methods are well known and have been used with single frame photographs or scanner runs prior to other image processing operations.

For mosaics, brightness normalisation has a strict analogy with geometric mosaicking where images are geometrically matched to the ground by Ground Control Points (GCP) and to other images by tie points (see Volume 2B). At tie points, in geometric mosaicking, the constraint is that the same feature must be located at the same point in the geometric frame model. In the same way, in brightness normalisation, there may be targets in some frames of known reflectance. These are sometimes called ‘invariant targets’. When their reflectance is only assumed from previous work they are sometimes called ‘pseudo-invariant’ targets. These are analogous

to GCP. In the overlap between frames, the analogy with tie points is patches or features that are assumed to have the same corrected spectra. For a given functional form of brightness model, the estimation reduces to a linear ‘bundle adjustment’.

The greatest problems with the simple empirical approach are image variance and heterogeneity. It is often extremely difficult to determine which components of the variation are atmospheric or BRDF brightening and which are differences in surface type. The high image spatial variance that is most obvious in high resolution images also makes fitting functions very difficult. Another problem is to decide which effects are additive and which are multiplicative. As long as the variation is not too great, the additive approximation is not too bad but when raw, uncalibrated data are used, the interaction between additive and multiplicative effects can be a problem.

<sup>2</sup> Most text and equations in this appendix have been sourced from Jupp and Strahler (2000).

**Background image:** This 180° panorama near William Creek, SA, shows the hotspot effect around the focal point of the camera, caused by the predominant backscatter of vegetation. This image is shown here with vertical exaggeration (see also Volume 1B—Figure 3.22). **Source:** Andreas Hüni, University of Zurich (taken on iPhone6s)

For example, calibration and atmospheric correction are linear effects but the atmospheric terms are angle dependent and vary with incident ( $\mu_i$ ) and view ( $\mu_v$ ) directions. That is, even if the surface were Lambertian (with a flat BRDF) the digital data recorded by a scanner in waveband  $j$  ( $dn_j$ ) will be related to the reflectance ( $\rho_j$ ) by:

$$L_j = \alpha_j dn_j + \beta_j$$

$$L_j = \frac{1}{\pi} E_T^*(\mu_i) T(\mu_v) \rho_j + L_p(\mu_i, \mu_v)$$

hence

$$dn_j = \gamma_j(\mu_i, \mu_v) \rho_j + \delta_j(\mu_i, \mu_v)$$

where

$L_j$  is the radiance recorded in waveband  $j$ ;  
 $\mu_i$  is the Sun or 'incident' radiation direction;  
 $E_T^*$  is the total irradiance at the target; and  
 $L_p$  is the radiation scattered into the sensor from the atmosphere.

The brightness variation described here (which has no surface BRDF effect in that  $\rho_j$  is not dependent on Sun or view angles) obviously has both additive and multiplicative terms, which are complex functions of the incident and view angles. When the surface target itself has a strong BRDF effect the variability will compound.

In the face of this, the empirical methods are generally constrained to be of low order functional forms that still have the structure of the variation being fitted. In this way, it is hoped that variance and image heterogeneity are 'orthogonal' to the function being fitted. In addition, the fitting is best done in the physical framework of the brightness variation, that is, in terms of the phase angles between Sun and sensor view angle and between sensor view angle and the specular vector. There have been many functions of this type developed such as Walthall, Hapke, Roujean and others. These will be addressed below in Appendix A1.3.

For the purely empirical approach and for small frames such as occur with aerial photography and video data (or even single scanner runs) the main problem is scene heterogeneity. There will be scene variations that occur at spatial scales near to or greater than the image and it is very difficult to separate these effects from atmospheric and broad BRDF effects. To illustrate this you can consider an image to be roughly approximated by:

$$I = I_H + I_L + I_A$$

where

$I_H$  is the high spatial frequency component;  
 $I_L$  is the low frequency component; and  
 $I_A$  is the angular brightness variation.

In many images, over the extent of a single frame, the low frequency and angular components are not in any way orthogonal. Hence it is near impossible to estimate the angular function from the single frame data. Moving to models based on view geometry and semi-empirical functions can help (Pickup *et al.*, 1995) but essentially some way to bring in the wider spatial context is needed.

Among empirical methods are also 'referencing' methods. In the case of image digitising, a reference approach is to scan an image and scan a reference (white Lambertian) standard sheet with the same illumination. The image is then 'normalised' by subtracting or dividing the reference from the image to remove lighting based view angle effects. In reconnaissance image data this is not possible unless a large reference target were to exist and scene BRDF effects would still remain. Hick and Ong (Ong *et al.*, 1995) have used a referencing method to extract brightness variations from video frames by referencing against the background of a Landsat TM image. It may be possible to fit empirical functions to the residual between the individual frame and the reference image. In this way, the problem of heterogeneity and correlated spatial frequencies described above could be reduced significantly.

It is worth mentioning here that as soon as the empirical methods extend from simple functional forms to ones involving image and Sun geometry, the processing methods start to become more costly. The cost is in extra work registering images and/or in extra costs associated with collecting attitude and position information (such as INS and GPS) for the platform. If the end result is worth this extra cost then it will be worth some extra modelling as discussed below.

## A1.2 Measurement Model Approach

### A1.2.1 Atmospheric correction and reflectance normalisation

It would clearly be an advantage to model the atmospheric and BRDF effects by physical models and estimate the corrected image in terms of a normalised reflectance factor. The added value would be the link this gives to physical models of the relationship between Earth parameters of interest and the measured reflectance. Obviously, a reflectance is closer to a material property and therefore resolves the problem of monitoring.

If reference targets in the image have measured reflectance (that is, the invariant or pseudo-invariant targets) it is possible to get close to reflectance from the empirical approach. However, it will never be clear how much residual atmosphere and BRDF effect remains. Obviously, most natural targets that are chosen as 'invariant' will have some BRDF that may not be known and that effect will be folded into the final image data.

Assuming the data are accurately calibrated to radiances, there are different ways to describe the process of correction. One is as follows:

An equation relating the recorded radiance sensed at altitude  $h$  above the target to the target reflectance factor is:

$$L_t(\mu_v, \mu_s, h, \lambda) = \frac{1}{\pi} E_T^*(\lambda) t(\mu_v, h, \lambda) (\rho_t + \rho_{env}) + L_p(\mu_v, \mu_s, h, \lambda) + L_g(\mu_v, \mu_s, h, \lambda)$$

where

$L_t(\mu_v, \mu_s, h, \lambda)$  is the radiance observed by the instrument from altitude  $h$ , with look (or view) direction  $\mu_v$  and Sun direction  $\mu_s$  at wavelength  $\lambda$ ;

$t(\mu_v, h, \lambda)$  is the beam transmittance through the layer between the surface and altitude  $h$  in direction  $\mu_v$ ;

$T(\mu_v, h, \lambda)$  is the diffuse transmittance for a layer of thickness  $h$  and for initial beam direction  $\mu_v$ ;

$L_p(\mu_v, \mu_s, h, \lambda)$  is the path radiance of light which did not interact with the surface; and

$L_g(\mu_v, \mu_s, h, \lambda)$  is the glint term that is most significantly present over water-covered targets and is sometimes present over land targets.

$E_T^*$  is the effective irradiance at the target, or

$$E_T^*(\lambda) = \frac{E_T(\lambda)}{1 - s \rho^*}$$

$E_T(\lambda)$  is the irradiance at the target for a 'black' earth;

$s$  is the sky hemispherical albedo;

$\rho^*$  is the background earth albedo;

$\rho_t$  is the target directional reflectance factor; and

$\rho_{env}$  is the environmental reflectance due to the background albedo  $\rho^*$  or

$$\rho_{env} = \rho^* \left( \frac{T(\mu_v, h, \lambda)}{t(\mu_v, h, \lambda)} - 1 \right)$$

If the atmosphere is characterised then it is possible to retrieve the directional reflectance factor ( $\rho_t$ ) for each pixel. This term needs to be defined carefully as there are many different types of 'reflectance' used.

The directional reflectance factor ( $\rho_t$ ) as used here is defined as:

$$\rho_t(\mu_s, \mu_v, f_d, \lambda) = \frac{\pi L(\mu_v, \lambda)}{E_T^*(\mu_s, f_d, \lambda)}$$

in which the irradiance ( $E_T^*$ ) is the sum of diffuse and direct terms and the fraction of diffuse ( $f_d$ ) is included as a parameter. The assumption that the irradiance can be characterised in this context by the Sun position and the fraction of diffuse radiation is one that needs evaluation. The value of using this form of reflectance is that it corresponds to what is measured in the field using an irradiance radiometer or a reference standard.

The physical approach depends on two steps. The first is to determine this reflectance factor for a surface by atmospheric correction. In atmospheric correction, the atmospheric terms are modelled and measured from image and ancillary data. The reflectance factor in an image may be obtained iteratively if the atmospheric turbidity makes the adjacency and other atmosphere/surface interactions significant. The second step is to normalise the reflectance factor in some way to account for its BRDF variation.

In order to go from the reflectance factor to such a corrected value, however, we effectively need to assume that:

$$\rho_t(\mu_s, \mu_v, f_d, \lambda) = \rho_t^0(\lambda) i(\mu_s, \mu_v, f_d, \lambda)$$

where  $i(\mu_s, \mu_v, f_d, \lambda)$  is the angular variation function that is assumed to characterise the land surface type and be normalised to 1.0 at a reference pair of Sun and view angles and standard atmosphere. Then 'corrected' data are reachable as:

$$\rho_t^0(\lambda) = \frac{\rho_t(\mu_s, \mu_v, f_d, \lambda)}{i(\mu_s, \mu_v, f_d, \lambda)}$$

The value of getting to this point is that both  $\rho_t^o(\lambda)$  and  $l(\mu_s, \mu_v, f_o, \lambda)$  can be interfaced with radiative transfer models to obtain parameters for the Earth's surface by inversion. The inversion may be simple (such as end member methods) or sophisticated (such as complete nonlinear modelling).

This physical approach is obviously highly sophisticated but demands a level of data quality and time in value adding that sometimes cannot economically be committed. However, if the end products are of high information value and valued by the client it is clearly an approach that could be pursued.

### A1.2.2 Analytic BRDF models

Assuming the atmospheric correction can be done (and the surface interaction needs to be flagged here as the coupling means the atmospheric correction and surface BRDF estimation are not independent) the surface BRDF can sometimes be defined by an analytic model.

The broad brightness variations in all EO images can be considered as the composite of the spectral properties of component materials combined with the structural effects that result in angular variations (see Volume 1B). Angular effects can be viewed in terms of three factors:

- volume effect—which depends on surface structure;
- occlusion (or hotspot) effect—shadow cast by the Sun occludes parts of the surface; and
- specular (or glint) effect—when surface properties strongly reflect illumination, such as water glint (see Volume 2A—Section 3.2.1).

Among the many models for the Volume effect are the Suits (1972) and SAIL (Scattering by Arbitrary Inclined Leaves; Verhoef, 1984, 1985) models as well as more advanced ones such as the hotspot-based model described (as examples) in Qin and Jupp (1993) and Jupp and Strahler (1991). The field of Geo-optical modelling (GO Models) is now very extensive but not as well served in its entirety yet by any major books or fully inclusive reviews.

The models used for the Volume effect range from the plane parallel layer models (Suits, 1972; Verhoef, 1984, 1985) to sophisticated three-dimensional ones such as the hotspot-based Geometric Optical-Radiative Transfer (GORT) models (Li and Strahler, 1995). Myneni and Ross (1990) provides a formative early overview with many of the original aspirations well presented. Liang *et al.* (2000) reviews the rapid progress during the following decade (see also Chen *et al.*, 2000, and Qin and Liang, 2000). More recently, the Radiation Transfer Model Intercomparison (RAMI) experiment has provided a major influence for stabilizing model development and ensuring maturity (Pinty *et al.*, 2004; Widlowski *et al.*, 2007).

The hotspot effect is a geometric or occlusion effect. From among the many papers that describe the hotspot effect, Excursus A1.1 provides an overview of one implementation of Geometric-Optical (GO) models (extracted from Jupp and Walker, 1996). The GO model approach has been extended into models that combine geometric and radiative transfer components into effective and complete models for canopies. Other GO models include:

- Nilson and Kuusk models (Nilson and Peterson, 1991; Kuusk and Nilson, 2000; and Kuusk, 2001);
- Chen and White models (Chen and Leblanc, 1997, 2001; White *et al.*, 2001, 2002; Chen *et al.*, 2003, 2005);
- Lacaze and Roujean models (Roujean, 2000; Lacaze and Roujean, 2001);
- Li, Strahler and Ni-Meister (GORT) models (Li and Strahler, 1995; Ni *et al.*, 1997, 1999; Yang and Friedl (2003); and
- Huemmrich and Goward (1997), Huemmrich (2001), Hanan (2001) and Haverd *et al.* (2012).

Among these, the GORT model (Li *et al.*, 1995; Ni *et al.*, 1997) has had extensive application, including lidar data inversion, as described (for example) in Ni-Meister *et al.* (2010). However, despite the enthusiasm of the early authors, more pragmatic approaches have been taken since that time based on semi-empirical BRDF models, as outlined below, and on mixture models for vegetation canopies.



## Excursus A1.1—Geometric Optical Models

**Source:** Jupp and Walker (1996)

A simple model for the remote sensing of a canopy is the Geometric Optical (GO) model. The simple GO (or hotspot) model for scenes which describe open forest or woodland areas is based on the one described in Jupp *et al.* (1986), Strahler and Jupp (1991a, 1991b) and Li and Strahler (1992). In this model, there are four kinds of ground cover 'visible' from a given direction. These are referred to as scene components and comprise:

- sunlit canopy (*sc*);
- shaded canopy (*shc*);
- sunlit background (*sb*); and
- shaded background (*shb*).

Each component is assumed to have a characteristic radiance and the radiance of a pixel is modelled as the area weighted combination (or linear mixture) of the characteristic component radiances. That is, the observed radiance of a single pixel ( $r_s$ ) is modelled as:

$$r_s = k_{sc} R_{sc} + k_{shc} R_{shc} + k_{sb} R_{sb} + k_{shb} R_{shb}$$

where

subscripts *sc*, *shc*, *sb*, and *shb* indicate the radiances of the four components as named above;

$R_j$  represents the (mean) radiance of component *j*; and

*k* indicates the sensed proportion of each component within the pixel from the given view direction.

Assuming the view and Sun directions are constant, the mean radiance over the scene ( $R_s$ ) can be written as:

$$R_s = K_{sc} R_{sc} + K_{shc} R_{shc} + K_{sb} R_{sb} + K_{shb} R_{shb}$$

where

capital  $K_j$  represents the mean or expected value of the varying proportions  $k_j$  over the scene for *j* as the components *sc*, *shc*, *sb* or *shb*; and mean value ( $R_s$ ), as a function of Sun and observer position, defines the BRDF of the scene.

In order for the scene BRDF model to be computed, a description of the size and shapes of the objects, their density and how they are distributed over the background is needed and the geometrical relationships between the objects and the expected values of the four components must be established. Jupp *et al.* (1986), Strahler and Jupp (1991a; 1991b) and Li and Strahler (1992) describe such a model for spheroidal crown (not necessarily opaque) volumes,

which is valid for any view or illumination angles using the 'Boolean' model of Serra (1982). In the Boolean model, the object 'centres' are assumed to be randomly distributed in a 'Poisson' distribution. By defining the geometry and the distributions, expressions for  $K_j$  may be derived. Strahler and Jupp (1991a, 1991b) use a simple model for spheroids, which is adequate for moderate Sun and view zenith angles, while Li and Strahler (1992) provide some more general alternative models for resolving the  $K_j$ . These basic scene BRDF models are quite simple and are easily implemented in various forms such as mathematical packages or spreadsheets.

In the woodlands and open forest areas typical of the area of Australia where the model studies have been made, the pixel to pixel behaviour of the image is conveniently (if not as accurately) described by a simpler form of the model in which the shaded background, sunlit (but still relatively dark) tree and shaded tree components are combined into one so that:

$$r_s = R_x + k_{sb} (R_{sb} - R_x)$$

where  $X$  is a composite component combining sunlit and shaded tree and shaded background and  $R_x$  is computed as:

$$R_x = \frac{K_{sc} R_{sc} + K_{shc} R_{shc} + K_{shb} R_{shb}}{1 - K_{sb}}$$

The simpler model has the advantage for this discussion that it shows clearly how, in many woodlands, the image pixel to pixel variation is driven primarily by the variation in the proportion of sunlit background that is visible in the pixels and the contrast between this sunlit background and the other components. It also provides a simple estimate for  $k_{sb}$  from images where  $R_{sb}$  and  $R_x$  are known for an appropriate image channel, or channel combination, as:

$$k_{sb} = \frac{r_s - R_x}{R_{sb} - R_x}$$

For such a model, the mean radiance (that is, BRDF) over all pixels in a patch with the same basic underlying type of cover and structure is therefore:

$$E(r_s) = R_s = R_x + (R_{sb} - R_x) K_{sb}$$

where  $K_{sb}$  is the mean value of  $k_{sb}$ , or the expected proportion of visible sunlit background for the particular Sun and view positions. This simple model has been found to be adequate for describing the data obtained by a Daedalus scanner over woodlands.

Linear end-member analysis is similar to the estimation of components described above. It has been the subject of useful research and application in Australia (Pech *et al.*, 1986, Pickup and Foran, 1987) and at regional scales where all pixels are mixtures of land covers of interest (Cross *et al.*, 1991). End-member analysis assumes each pixel to be a composition, or mixing, of a few base components or 'end-members'. The pixel signature is assumed to be a linear sum of reflectances from each of  $n$  end-members weighted in proportion to its cover ( $k_j$ ) in the pixel:

$$r_s = \sum_{j=1}^n k_j R_j$$

where  $k_j \geq 0$  for  $j=1, n$

$$\sum_{j=1}^n k_j = 1$$

End-member analysis seeks to invert this mixing by deriving the proportions ( $k_j$ ) of each component in the pixel signature. This can feasibly be derived from EO data provided that if there are  $n$  components (trees, shrubs, grass etc.) then there are at least  $(n-1)$  channels of data that separate the end-members spectrally. The key assumptions built into the end-member method are that:

- end-members (pure examples of total cover by trees, shrubs, grass and background) are spectrally consistent between sites; and
- reflectance values for end members ( $R_j$ ) are available from EO data or can be accurately derived by other means (such as field spectral measurements).

There has been considerable work aimed at deriving end-members from the data (a form of Principal Components Analysis; see Boardman, 1990; Appendix A6.3 below; and Volume 2C) and employing high spectral resolution data to effect separation of more than a few components (Adams *et al.*, 1989). However, with a lack of available high resolution spectral data, this linear approach suffers from several significant limitations to its applicability in Australia:

- available broad band signatures of the tree and shrub crowns over much of Australia whilst different, are not markedly spectrally distinct;
- even if spectrally distinct crowns did exist for the available bands, their distinction is confounded by the effects of shadowing within crowns and cast shadow on the background (with bigger plants shading smaller plants). This makes the signature of the end-members difficult to estimate as the signature depends on the proportions of crowns and shadows present and variations in Sun and viewing angles; and

- relatively low covers of trees and shrubs, together with shadowing, introduce such high spectral variance into the data relative to the spectral contrasts between end-members that the numerical methods used in the end-member analysis become highly unstable.

Shadow effects obviously depend primarily on the Sun angle. Although the crown cover is the same, lower Sun angles clearly decrease image brightness. Differences due to shadowing can be taken into account in end-member analysis, provided the end-member values are recalculated for each temporal image and one or more components labelled 'shade' are added to the list. However, its successful application still depends on an assumption of linear scaling along cover gradients due to Sun positional and sensor view angle changes. These assumptions in practice are erroneous in structured vegetation (for example, vegetation with discontinuous cover of trees or shrubs), and this limits the application of such methods to general synoptic estimates of change in cover.

In structured vegetation, that is where the cover of trees and/or shrubs is discontinuous, it is appropriate to model vegetation cover directly as an assemblage of various sizes and shapes of three-dimensional objects (trees, shrubs, grass tussocks, herbs, etc.) scattered on a background that may be uniform or heterogeneous (Li and Strahler, 1985; Jupp *et al.*, 1986). The GO model may then be used to model the bidirectional reflectance of the canopies. In this approach, the effects due to shadowing on the overall reflectance (or infrared temperature) from a scene become important and useful features and the correlated interactions between shaded and sunlit components are built into the analysis—although it now becomes nonlinear. The directional radiance of the vegetation is then a mixture of four components (sunlit and shaded tree crowns, and sunlit and shaded backgrounds) that is seen from a given viewing angle. The areal proportions of these four components, for given illumination and viewing directions (which can be off-nadir), will be a function of the sizes, shapes, orientations and placements of the objects (that is, individual plants) within the scenes.

A GO model is most appropriate to woodlands or vegetation in which the vegetative cover is discontinuous, that is, where tree and shadowed background interactions account for a large proportion of the variance in the image. The further advantage of such models would be that they could also potentially be invertible and provide structural as well as cover information. The invertibility of GO models was demonstrated in principle by Franklin and Strahler (1988) and Wu and Strahler (1993) in which

tree size and density were estimated from reflectance data. When size, shape and orientation are fixed or characterised by distributions of known parameters, and the object centres are randomly distributed, then the proportions of the four components can be estimated using the Boolean model of Serra (1982). This GO model is termed the Boolean version (Strahler and Jupp, 1991a; 1991b; Li and Strahler, 1992). It accounts for the changes in proportions that

occur with random overlapping objects as the density of objects increases and can easily model scale effects and changing Sun and view directions. The GO aspect of the model implies that multiple scattering of radiation in the vegetation layer is neglected. While the evidence of our eyes supports this, there are wavelengths (in particular the near infrared) where multiple scattering is very significant (Li *et al.*, 1995).

## A1.3 Semi-empirical Methods

Semi-empirical BRDF models have also been developed which are empirical but are based on physical models and often contain parameters that relate to surface parameters. They offer the means of using the mosaic approach and ‘empirical’ model fitting but also enable surface parameters to be extracted. The methods are called ‘semi-empirical’ because they generally assume atmospheric correction can be done and also often have forms that involve Earth surface parameters. The main class of these models is that of the ‘kernel’ models.

The following text describing the kernel approach was extracted from the MODIS-Land ATBD (Strahler *et al.*, 1995b).

### A1.3.1 Theoretical description

Kernel-driven models for the bidirectional reflectance distribution function of vegetated land surfaces attempt to describe the BRDF as a linear superposition of a set of kernels that describe basic BRDF shapes, with the coefficients or weights chosen to adapt the sum of the kernels to the given case. Typically, semi-empirical kernels are based either on one of several possible approximations to a radiative transfer scenario of light scattering in a horizontally homogeneous plant canopy (such as a crop canopy), or on one of several approximations feasible in a geometric-optical model of light scattering from a surface covered with vertical projections that cast shadows (such as a forest canopy). Deriving a kernel of this nature requires simplifying and manipulating a model for the BRDF until it reaches the form:

$$R = c_1 k + c_2$$

in which  $k$  is a function only of view and illumination geometry,  $c_1$  and  $c_2$  are constants containing physical parameters, and  $R$  is the modelled value of the true BRDF,  $\rho$ .

Semi-empirical kernels can be of two types. First, they may contain only geometric terms, but no physical parameters. The complete model then is linear, and may be scaled to arbitrary scales even for mixed scenes; neglecting adjacency effects, the weights of the kernels will be linear functions of the areal proportions of the sub-pixel weights. The so-called Ross-kernels, which are approximations to the radiative transfer theory in plant canopies of Ross (1981) described below, belong to this class, as does the so-called Roujean geometric-optical kernel (Roujean *et al.*, 1992). In the second case, kernels contain one or very few physical parameters and thus instead of having one kernel, provide a family of kernels depending on these parameters. The geometric-optical Li-kernels (Wanner *et al.*, 1995) belong to this type. In order to reduce them to the form given in the previous equation, the kernel-internal parameters are fixed to a selection of a few values each. The following discussion presents each of the kernels used in the BRDF/Albedo algorithm. For more complete information on the theory and derivation of the kernels, see Wanner *et al.* (1995).

### A1.3.2 Kernels

The Ross kernels are derived from a formula presented by Ross (1981) for the directional reflectance above a horizontally homogeneous plant canopy calculated from radiative transfer theory in a single scattering approximation. The Ross-thick kernel was derived and described by Roujean *et al.* (1992). It is based on an approximation for large LAI values:

$$k_{thick} = \frac{\left(\frac{\pi}{2} - \xi\right) \cos \xi + \sin \xi}{\cos \theta_i + \cos \theta_v} - \frac{\pi}{4}$$

$$c_1 = \frac{4s}{3\pi} \left(1 - e^{-LAI/B}\right)$$

$$c_2 = \frac{s}{3} + e^{-LAI/B} \left(\rho_s - \frac{s}{3}\right)$$

In the kernel,  $\theta_i$  and  $\theta_v$  are zenith angles for illumination and view, respectively;  $\phi$  is the relative azimuth of illumination and view directions; and  $\xi$  is the phase angle of scattering:

$$\cos \xi = \cos \theta_i \cos \theta_v + \sin \theta_i \sin \theta_v \cos \phi$$

In the constants:

$s$  is leaf reflectance (= leaf transmittance);

$\rho_s$  is the (assumed isotropic) surface reflectance of the soil or understory;

$LAI$  is the leaf area index; and

$B$  is the average of secants of possible view and illumination zenith angles. For this formulation, a spherical leaf angle distribution is assumed.

The Ross-thin kernel simplifies Ross's equation based on an approximation for small  $LAI$  values:

$$k_{thin} = \frac{(\pi/2 - \xi) \cos \xi + \sin \xi}{\cos \theta_i \cos \theta_v} - \frac{\pi}{2}$$

$$c_1 = \frac{2sLAI}{3\pi}$$

$$c_2 = \frac{sLAI}{3} + \rho_s$$

Although this kernel applies primarily to the case of a thin canopy of scatterers over a uniform background, it can also be appropriate for a very dense, uniform canopy of high leaf area, since in that case the leaf layers below the uppermost can act like a uniform background (Strahler *et al.*, 1995a).

The derivation of the Roujean geometric-optical kernel is presented in the appendix to the paper of Roujean *et al.* (1992). The kernel may be regarded as accounting for the scattering of a random arrangement of three-dimensional rectangular solids ('bricks') with isotropic scattering surfaces. Shadows are taken as perfectly black. Mutual shadowing, in which the shadow of one object falls on another object, is not taken into account. The bricks are long with respect to their height  $h$  and width  $b$ . The kernel is:

$$k_{brick} = \frac{1}{2\pi} \left[ (\pi - \phi) \cos \phi + \sin \phi \right] \tan \theta_i \tan \theta_v - \frac{1}{\pi} \left( \tan \theta_i + \tan \theta_v + \sqrt{\tan^2 \theta_i + \tan^2 \theta_v - 2 \tan \theta_i \tan \theta_v \cos \phi} \right)$$

$$c_1 = \rho_s \frac{h}{b}$$

$$c_2 = \rho_s$$

The Li kernels are derived from the modelling approach of Li and Strahler (1986, 1992). In this approach, the surface is taken as covered by randomly-placed projections (such as tree crowns) that are taken to be spheroidal in shape and centred randomly within a layer above the surface. The BRDF is modelled as a function of the relative areas of sunlit and shaded, crown and background that are visible from the viewing position in the hemisphere. For the Li-sparse kernel, it is assumed that shaded crown and shaded background are black, and that sunlit crown and background are equally bright. Under these circumstances, and with some further approximations in the way that view and illumination shadows overlap, the Li-sparse kernel is:

$$k_{sparse} = (O - 1) \left( \sec \theta'_i + \sec \theta'_v \right) + \frac{1}{2} (1 + \cos \xi') \sec \theta'_v$$

where

$$O = \frac{1}{\pi} (t - \sin t \cos t)$$

$$\cos t = \frac{h \sqrt{D^2 + (\tan \theta'_i \tan \theta'_v \sin \phi)^2}}{b \sec \theta'_i + \sec \theta'_v}$$

$$D = \sqrt{\tan^2 \theta'_i + \tan^2 \theta'_v - 2 \tan \theta'_i \tan \theta'_v \cos \phi}$$

$$\cos \xi' = \cos \theta'_i \cos \theta'_v + \sin \theta'_i \sin \theta'_v \cos \phi$$

$$\theta' = \tan^{-1} \left( \frac{b}{r} \tan \theta \right)$$

In these expressions,  $b$  is the vertical radius of the spheroid;  $r$  is the horizontal radius of the spheroid; and  $h$  is the height of the centre of the spheroid. For this model,

$$c_1 = C \lambda \pi r^2$$

$$c_2 = C$$

Here,  $C$  is the brightness of sunlit surface, and  $\lambda$  is the count density of spheroids (number of spheroids per unit area).

The Li-dense kernel differs from the Li-sparse kernel in that it accommodates mutual shadowing. It assumes a random distribution of crown heights to maximize the geometric-optical effect in a dense ensemble of canopies.

$$k_{dense} = \frac{(1 + \cos \xi') \sec \theta'_v}{(1 - O)(\sec \theta'_i + \sec \theta'_v)} - 2$$

$$c_1 = \frac{C}{2}$$

$$c_2 = C$$

These kernels are not yet linear in that they still contain two parameters, namely the ratios  $b/r$  and  $h/b$ , describing crown shape and relative height. The kernels therefore actually represent two families of kernels, governed by the values of these two internal parameters. For the present, we provide two choices for each parameter, thus providing four kernels within each family. For the  $b/r$  ratio, the values 2.5 (prolate shape) and 0.75 (oblate shape) are used; for the  $h/b$  ratio, we use values of 2.5 (tall) and 1.5 (short).

### A1.3.3 Kernel-driven models

A complete kernel-driven semi-empirical model is formulated as a linear combination of kernels. Most suitably it has the form:

$$R = f_{iso} + f_{geo} k_{geo} + f_{vol} k_{vol}$$

which is derived from adding appropriate choices of geometric-optical surface-scattering and radiative-transfer volume-scattering kernels, each multiplied by a proportion  $\alpha$  or  $(1 - \alpha)$  that weights the contribution of each model. These proportions may be regarded as the areal proportions of land cover types exhibiting each type of scattering (neglecting multiple scattering between the two components), or as mixing proportions for land cover types that display both a volume-scattering and a geometric-optical contribution to the BRDF. The quantities  $k_{geo}$  and  $k_{vol}$  are the respective kernels; the factors  $f_{geo}$  and  $f_{vol}$  are their respective weights; and the term  $f_{iso}$  is the isotropic contribution. Using the kernels presented above, there will be two choices for  $k_{vol}$  and three for  $k_{geo}$ , providing six possible models (disregarding here the four sets of internal parameters used for the Li kernels). The formulae for  $f_{iso}$ ,  $f_{geo}$ , and  $f_{vol}$  are shown in Table A1.1.

In the inversion and fitting of a semi-empirical model to data, estimates of the weights  $f$  are retrieved from bidirectional reflectances and specification of viewing and illumination positions. Although this objective satisfies many of the goals of the BRDF/Albedo product, the existence of formulae for the weights  $f$  in terms of physical parameters could provide for direct inference of physical parameters from the weight values fitted.

**Table A1.1** Semi-empirical weight formulas

Model	Weight Formulae
Ross-thin + Roujean	$f_{iso} = \alpha \rho_s = (1 - \alpha) \left( \frac{sLAI}{3} + \rho_s \right)$ $f_{geo} = \alpha \rho_s \frac{h}{b}$ $f_{vol} = (1 - \alpha) \frac{2sLAI}{3\pi}$
Ross-thick + Roujean	$f_{iso} = \alpha \rho_s = (1 - \alpha) \left( \frac{s}{3} + e^{-LAI/B} \left( \rho_s - \frac{s}{3} \right) \right)$ $f_{geo} = \alpha \rho_s \frac{h}{b}$ $f_{vol} = (1 - \alpha) \frac{4s}{3\pi} (1 - e^{-LAI/B})$
Ross-thin + Li-sparse	$f_{iso} = \alpha C = (1 - \alpha) \left( \frac{sLAI}{3} + \rho_s \right)$ $f_{geo} = \alpha C \lambda \pi r^2$ $f_{vol} = (1 - \alpha) \frac{2sLAI}{3\pi}$
Ross-thick + Li-sparse	$f_{iso} = \alpha C = (1 - \alpha) \left( \frac{s}{3} + e^{-LAI/B} \left( \rho_s - \frac{s}{3} \right) \right)$ $f_{geo} = \alpha C \lambda \pi r^2$ $f_{vol} = (1 - \alpha) \frac{4s}{3\pi} (1 - e^{-LAI/B})$
Ross-thin + Li-dense	$f_{iso} = \alpha C = (1 - \alpha) \left( \frac{sLAI}{3} + \rho_s \right)$ $f_{geo} = \alpha \frac{C}{2}$ $f_{vol} = (1 - \alpha) \frac{2sLAI}{3\pi}$
Ross-thick + Li-sparse	$f_{iso} = \alpha C = (1 - \alpha) \left( \frac{s}{3} + e^{-LAI/B} \left( \rho_s - \frac{s}{3} \right) \right)$ $f_{geo} = \alpha \frac{C}{2}$ $f_{vol} = (1 - \alpha) \frac{4s}{3\pi} (1 - e^{-LAI/B})$



### A1.3.4 Modified Walthall model

Empirical models can be understood as being of the kernel-driven model type as well, where the kernels are empirical functions. An example is the modified Walthall model, derived by Walthall *et al.* (1985) and improved by Nilson and Kuusk (1989). It has the form:

$$R = p_0(\theta_i^2 + \theta_v^2) + p_1\theta_i^2\theta_v^2 + p_2\theta_i\theta_v\cos\phi + p_3$$

Note that this is the same form as the semi-empirical models discussed above—it comprises a weighted superposition of functions of angles, and the weights  $p_0$ – $p_3$  are the parameters of the model. As a consequence, models like the modified Walthall model can be processed along with linear semi-empirical models by the same linear inversion scheme.

### A1.3.5 Advantages of linear models

Linear models have a number of advantages in the context of global data processing. The opportunities arising from linearity in BRDF models have been discussed by Barnsley *et al.* (1990) and Strahler *et al.* (1995b). A first advantage is that, provided sufficient observations of BRDF are available, any linear model can be inverted analytically through matrix inversion

for the system of equations obtained by setting the derivative of the error function to zero. This provides direct estimates of the parameters  $f_{iso}$ ,  $f_{geo}$  and  $f_{vol}$  while avoiding numerical inversion problems associated with nonlinear models.

Second, both the directional-hemispherical and bi-hemispherical integrals of the BRDF (black-sky and white-sky albedos) may be pre-calculated for each kernel individually. The albedo of a model then is simply the sum of the kernel albedos, weighted by  $f$  values. By using a look-up table, numerical integration of the models can thus be avoided.

Third, linear BRDF models scale linearly in space if adjacency effects are assumed to be small. This allows for mixed pixel cases, as indicated by the areal proportion parameter  $\alpha$  in the model factors listed above. This feature also allows scaling BRDF and albedo from one spatial resolution up to a coarser one, such as to a particular resolution needed for a climate model. Finally, since some of the parameters driving the models are dependent on wavelength while others (such as structural parameters) are not, it may be possible to extract all or some of them from multiband analysis, making assumptions about the others.

## A1.4 The Issue of Scale

The BRDF models mentioned above and used to develop the kernels are general ‘average’ scene models that apply to spatially averaged data over homogeneous target areas. They do not model the image variance (which is also angle dependent) nor do they apply to a single object in a scene. The BRDF of a single tree is different from that of a stand of trees and the BRDF of a stand of trees is different from that of an open soil area.

Scaling affects the basis for image stratification. The tree and gap structure at one scale provides different functions for a stand and bare area delineation. At a broad scale, the composite of tree and gap merges into stands and at another stands and clearings merge in a ‘woodland’ category. At each level, the BRDF of the separated components will be different from each other and the composite land cover.

This leads to the consideration of some issues most important in an approach to correcting high resolution images for BRDF effects. The atmospheric effect is very smooth having been ‘filtered’ by an MTF with a size of about 200 m in most cases. However, the BRDF is highly variable. Consider a photograph with a wide field of view and the solar hotspot near to its centre. A tree near the nadir view and a similar one near the edge will have different brightness but the difference will not be as great as the overall brightening due

to Volume and hotspot effects. In particular, there will be much more shadow at the edges reached away from the hotspot direction and path lengths for Volume scattering will be longer through the canopy than at the geometric centre of the frame. The ‘colour’ of shadow near the edges will also be slightly darker than at the nadir but not very different. In this case, applying an overall BRDF correction function will tend to change the colour of shadows and tree crowns at the edge in a way that may be quite wrong—even though the overall area average may be ‘balanced’. Obviously, what works at one scale may not help at another.

Even if an H-resolution (high resolution) approach is taken in which the BRDF of the objects is modelled (trees for example), the fact will still be that the image contains different geometric proportions of scene objects and components in a way that depends on the phase angles between the Sun, the target object(s) and the view. Perhaps the only perfect way to resolve the issue is by 3D stereo modelling. This is unlikely to be cost-effective with video or digitised aerial photograph data.

BRDF models are further detailed in Appendix 2.



## A1.5 References

- Adams, J.B., Smith, M.O., and Gillespie, A.R. (1989). Simple models for complex natural surfaces: a strategy for the hyperspectral era of remote sensing. *Proc. IGARSS'89*, Vancouver, B.C., Canada, 16–21.
- Barnsley, M., Morris, K., and Reid, A. (1990). Preliminary analysis of a multiple view angle image data set. *Proc. NERC 1989 Airborne Remote Sensing Campaign Symposium*, Keyworth, England, 18–19 December 1990, 49–68.
- Boardman, J. (1990). Inversion of high spectral resolution data. *Proc. SPIE, vol 1298, Imaging Spectroscopy and Terrestrial Environment*, Orlando, FL, 222–223.
- Chen, J.M., and Leblanc, S.G. (1997). A four-scale bidirectional reflectance model based on canopy architecture. *IEEE Transactions on Geoscience and Remote Sensing*, 35(5), 1316–1337.
- Chen, J.M., and Leblanc, S.G. (2001). Multiple-scattering scheme useful for geometric optical modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 39(5), 1061–1071.
- Chen, J.M., J. Liu, Leblanc, S.G., Lacaze, R., and Roujean, J.-L. (2003). Multi-angular optical remote sensing for assessing vegetation structure and carbon absorption. *Remote Sensing of Environment*, 84(4), 516–525.
- Chen, J.M., Menges, C.H., and Leblanc, S.G. (2005). Global mapping of foliage clumping index using multi-angular satellite data. *Remote Sensing of Environment*, 97(4), 447–457. doi: 10.1016/j.rse.2005.05.003
- Chen, J.M., Li, X., Nilson, T., and Strahler, A. (2000). Recent advances in geometrical optical modelling and its applications. *Remote Sensing Reviews*, 18 (2–4), 227–262, doi: 10.1080/02757250009532391
- Cross, A.M., Settle, J.J. Drake, N.A., and Paivinen, R.T.W. (1991). Subpixel measurement of tropical forest cover using AVHRR data. *International Journal of Remote Sensing*, 12, 1119–1129.
- Franklin, J., and Strahler, A.H. (1988). Invertible canopy reflectance modeling of vegetation structure in semi-arid woodland. *IEEE Transactions on Geoscience and Remote Sensing*, 26, 809–825.
- Hanan, N.P. (2001). Enhanced two-layer radiative transfer scheme for a land surface model with a discontinuous upper canopy. *Agricultural and Forest Meteorology*, 109(4), 265–281.
- Haverd, V., Lovell, J.L., Cuntz, M., Jupp, D., and Newnham, G. (2012). The Canopy Semi-analytic  $P_{\text{gap}}$  and Radiative Transfer (CanSPART) model: Formulation and application. *Agricultural and Forest Meteorology*, 160, 14–35. doi: 10.1016/j.agrformet.2012.01.018
- Huemmrich, K.F., and Goward, S.N. (1997). Vegetation Canopy PAR Absorptance and NDVI: An Assessment for Ten Tree Species with the SAIL Model. *Remote Sensing of Environment*, 61(2), 254–269.
- Huemmrich, K.F. (2001). The GeoSail model: a simple addition to the SAIL model to describe discontinuous canopy reflectance. *Remote Sensing of Environment*, 75(3), 423–431.
- Jupp, D.L.B., and Walker, J. (1996). Detecting structural and growth changes in woodlands and forests: the challenge for remote sensing and the role of geometric optical modelling. In: Gholz, H.L., Nakane, K. and Shimoda, H. (eds). *The Use of Remote Sensing in the Modeling of Forest Productivity at Scales from the Stand to the Globe*. Kluwer Academic Publishers, Dordrecht.
- Jupp, D.L.B., and Strahler, A. (2000). *Image Brightness and BRDF Workshop Issues*. CSIRO Earth Observation Centre, Canberra.
- Jupp, D.L.B., Walker, J., and Penridge, L.K. (1986). Interpretation of vegetation structure in Landsat MSS imagery: a case study in disturbed semi-arid eucalypt woodland. Part 2. Model based analysis. *Journal of Environmental Management*, 23, 35–57.
- Kuusk, A., and Nilson, T. (2000). A directional multispectral forest reflectance model. *Remote Sensing of Environment*, 72(2), 244–252.
- Kuusk, A. (2001). A two-layer canopy reflectance model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 71, 1–9.
- Lacaze, R., and Roujean, J.L. (2001). G-function and HOt SpoT (GHOST) reflectance model—Application to multi-scale airborne POLDER measurements. *Remote Sensing of Environment*, 76(1), 67–80.
- Li, X., and Strahler, A.H. (1986). Geometric-optical bidirectional reflectance modeling of a coniferous forest canopy. *IEEE Transactions on Geoscience and Remote Sensing*, GE24(6), 906–919.
- Li, X., and Strahler, A.H. (1992). Geometric-optical bidirectional reflectance modeling of the discrete-crown vegetation canopy: Effect of crown shape and mutual shadowing, *IEEE Transactions on Geoscience and Remote Sensing*, 30, 276–292.

- Li, X., Strahler, A.H., and Woodcock, C.E. (1995). A hybrid geometric optical-radiative transfer approach for modeling albedo and directional reflectance of discontinuous canopies, *IEEE Transactions on Geoscience and Remote Sensing*, 33, 466–480.
- Liang, S., Strahler, A.H., Barnsley, J., Borel, C.C., Gerstl, S.A.W., Diner, D.J., Prata, A.J., and Walthall, C.L. (2000). Multiangle Remote Sensing: Past, Present and Future. *Remote Sensing Reviews*, 18, 83–102.
- Myneni, R.B., and Ross, J. (1990). *Applications in Optical Remote Sensing and Plant Ecology*. Heidelberg, FRG: (Springer-Verlag).
- Ni, W., Li, X., Woodcock, C.E., Roujean, J.-L., and Davis, R.E. (1997). Transmission of solar radiation in boreal conifer forests: Measurements and models. *Journal of Geophysical Research*, 102(D24), 29555–29566.
- Ni, W., Li, X., Woodcock, C.E., Caetano, M.R., and Strahler, A.H. (1999). An analytical hybrid GORT model for bidirectional reflectance over discontinuous plant canopies. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2), 987–999.
- Ni-Meister, W., Lee, S., Strahler, A.H., Woodcock, C., Schaaf, C., Yao, T., Ranson, K.J., Sun, G., Bryan, B.J. (2010). Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. *Journal of Geophysical Research*, 115. doi: 10.1029/2009JG000936.
- Nilson, T., and Kuusk, A. (1989). A reflectance model for the homogeneous plant canopy and its inversion. *Remote Sensing of Environment*, 27, 157–167.
- Nilson, T., and U. Peterson (1991). A forest canopy reflectance model and a test case. *Remote Sensing of Environment*, 37(3), 131–142.
- Ong, C., Hick, P., Craig, M., Warren, P., and Newman, C. (1995). A correlative technique for correction of shading effects in digital multispectral video imagery. *Proc. ISSSR*, Melbourne, November 1995.
- Pech, R.P., Davis, A.W., and Graetz, R.D. (1986). Reflectance modelling and the derivation of vegetation indices for an Australian semi-arid shrubland. *International Journal of Remote Sensing*, 7, 389–412.
- Pickup, G., and Foran, B.D. (1987). The use of spectral and spatial variability to monitor cover change on inert landscapes. *Remote Sensing of Environment*, 23, 351–63.
- Pickup, G., Chewings, V.H., and Pearce, G. (1995). Procedures for correcting high resolution airborne video imagery. *International Journal of Remote Sensing*, 16, 1647–1662.
- Pinty, B., Widlowski, J.-L., Taberner, M., Gobron, N., Verstraete, M. M., Disney, M., Gascon, F., Gastellu, J.-P., Jiang, L., Kuusk, A., Lewis, P., Li, X., Ni-Meister, W., Nilson, T., North, P., Qin, W., Su, L., Tang, S., Thompson, R., Verhoef, W., Wang, H., Wang, J., Yan, G., Zang, H. (2004). Radiation Transfer Model Intercomparison (RAMI) exercise: Results from the second phase. *Journal of Geophysical Research*, 109, D06210, doi: 10.1029/2003JD004252.
- Qin, W., and Jupp, D.L.B. (1993). An Analytical and Computationally Efficient Reflectance Model for Leaf Canopies. *Agricultural and Forest Meteorology*, 66, 31–64.
- Qin, W., and Liang, S. (2000). Plane parallel canopy radiation transfer modeling: Recent advances and future directions, *Remote Sensing Reviews*, 18(2-4), 281–305. doi: 10.1080/02757250009532393
- Ross, J.K. (1981). *The radiation regime and architecture of plant stands*. Junk Publ., The Hague. 391pp.
- Roujean, J.L. (2000). A parametric hot spot model for optical remote sensing applications. *Remote Sensing of Environment*, 71(2), 197–206.
- Roujean, J.L., Latoy, M., and Deschamps, P.Y. (1992). A bidirectional reflectance model of the earth's surface for the correction of remote sensing data. *Journal of Geophysical Research*, 97, 20455–20468.
- Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press, London, New York.
- Strahler, A.H., and Jupp, D.L.B. (1991a). Modeling bidirectional reflectance of forests and woodlands using boolean models and geometric optics. *Remote Sensing of Environment*, 34, 153–166.
- Strahler, A.H., and Jupp, D.L.B. (1991b). Geometric-Optical Modeling of Forests as Scenes Composed of Three-Dimensional Discrete Objects. In *Photon-Vegetation Interactions: Applications in Optical Remote Sensing and Plant Ecology*, (Eds: R.B Myneni and J. Ross), Springer-Verlag, Heidelberg, FRG, 415–440.
- Strahler, A., Wanner, W., Zhu, Q., and Jin, X. (1995a). Bidirectional reflectance modeling of data from vegetation obtained in the Changchun solar simulation laboratory, *Proc. 15th Int. Geosci. and Remote Sensing Symp.*, Florence, Italy, July 10-14, 1995, vol. 3. pp. 1965–1967.

- Strahler, A.H., Barnsley, M.J., d'Entremont, R., Hu, B., Lewis, P., Li, X., Muller, J.-P., Barker Schaaf, Wanner, W., and Zhang, B. (1995b). *MODIS BRDF/Albedo Product: Algorithm Theoretical Basis Document Version 3.2*. NASA EOS, May 1995.
- Suits, G.H. (1972). The calculation of the directional reflectance of a vegetative canopy. *Remote Sensing of Environment*, 2, 117–125.
- Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. *Remote Sensing of Environment*, 16, 125–141.
- Verhoef, W. (1985). Earth observation modeling based on layer scattering matrices. *Remote Sensing of Environment*, 17, 165–178.
- Walker, J., Jupp, D.L.B., Penridge, L.K., and Tian, G. (1986). Interpretation of vegetation structure in Landsat MSS imagery: a case study in disturbed semi-arid eucalypt woodland. Part 1. Field data analysis. *Journal of Environmental Management*, 23, 19–33.
- Walthall, C.L., Norman, J.M., Welles, J.M., Campbell, G., and Blad, B.L. (1985). Simple equation to approximate the bidirectional reflectance from vegetation canopies and bare soil surfaces. *Applied Optics*, 24, 383–387.
- Wanner, W., Li, X., and Strahler, A.H. (1995). On the derivation of kernels for kernel-driven models of bidirectional reflectance, *Journal of Geophysical Research*, 21077–21090.
- White, H.P., Miller, J.R., and Chen, J.M. (2001). Four-Scale Linear Model for Anisotropic Reflectance (FLAIR) for plant canopies - Part I: Model description and partial validation. *IEEE Transactions on Geoscience and Remote Sensing*, 39(5), 1072–1083.
- White, H.P., Miller, J.R., and Chen, J.M. (2002). Four-Scale Linear Model for Anisotropic Reflectance (FLAIR) for plant canopies - Part II: Validation and inversion with CASI, POLDER, and PARABOLA data at BOREAS. *IEEE Transactions on Geoscience and Remote Sensing*, 40(5), 1038–1046.
- Widlowski, J. L., Taberner, M., Pinty, B., Bruniquel-Pinel, V., Disney, M., Fernandes, R., Gastellu-Etchegorry, J.-P., Gobron, N., Kuusk, A., Lavergne, T., Leblanc, S., Lewis, P.E., Martin, E., Möttus, M., North, P.R.J., Qin, W., Robustelli, M., Rochdi, N., Ruiloba, R., Soler, C., Thompson, R., Verhoef, W., Verstraete, M.M., Xie, D. (2007). Third Radiation Transfer Model Intercomparison (RAMI) exercise: Documenting progress in canopy reflectance models, *Journal of Geophysical Research*, 112, D09111. doi: 10.1029/2006JD007821.
- Wu, Y., and Strahler, A.H. (1993). Remote estimation of crown size, stand density and foliage biomass on the Oregon transect. *Ecological Applications*, 4, 299–312.
- Yang, R.Q., and Friedl, M.A. (2003). Modeling the effects of three-dimensional vegetation structure on surface radiation and energy balance in boreal forests. *Journal of Geophysical Research*, 108(D16).



# Appendix 2

## BRDF Models

This Appendix reviews a number of Bidirectional Reflectance Distribution Functions (BRDF) that are commonly used in conjunction with EO imagery<sup>3</sup>.

### A2.1 Kernel Models

These are (semi-)empirical models based on linear combinations of ‘kernels’:

$$\rho = f_{iso} + f_{geo} k_{geo} + f_{vol} k_{vol}$$

which represent surface reflectance ( $\rho$ ) as a function of component reflectances ( $f_x$ ) and the kernels ( $k_x$ ), which are mathematical functions that depend on Sun (or incident) and view (or observer) angles  $\theta_i$  and  $\theta_v$ . The subscripts ‘geo’ and ‘vol’ refer to the physical bases for some kernels in which there is an identification of a ‘geometric’ or hotspot factor and a ‘volume’ or path length and scattering factor.

By convention, kernel models are expressed in way is that when Sun and observer are at zenith:

$$k_{geo}(0,0) = k_{vol}(0,0) = 0$$

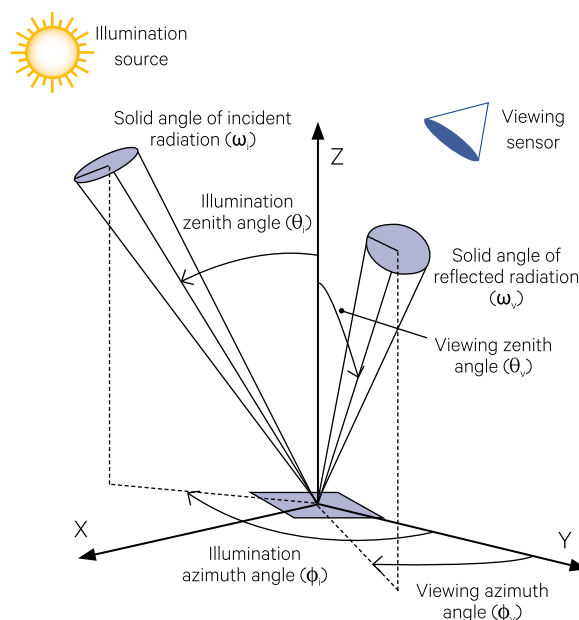
so that  $\rho(0,0) = f_{iso}$ . Some published models do not enforce this convention but all can if needed.

To compute the kernel models, we need to specify which choice of the angular description is being used for Sun and view directions. Sometimes this is ‘target based’ and sometimes it is ‘sensor based’. Relative to an airborne scanner, the position of the hotspot is dependent on flight direction, aircraft attitude, scan read-out order, etc.

In an aircraft and scanner situation, the pointing vector for a pixel may be defined pointing ‘up’ or ‘down’ and there are important issues of whether the scanner scans left to right or right to left! In atmospheric correction routines, the conventions of the scanner and aircraft model frame must be very carefully specified. Any data on angles that came with images being investigated with kernel models must follow a single convention or be carefully and fully specified.

For the following, the consistent convention being used is the ‘target based’ one. That is, the geometry is as illustrated in Figure A2.1, where  $\theta_i$  is the Sun or incident ray zenith angle,  $\theta_v$  is the view or observed zenith angle and the azimuth ( $\phi$ ) is relative to the Sun position. This is the convention used by the MODIS MODLAND team in their product ATBD (Strahler *et al.*, 1996) and also in AVHRR pathfinder.

Figure A2.1 Geometry of remote observer



<sup>3</sup> In this appendix, most text and equations have been sourced from Jupp (2000).

**Background image:** The appearance of identical metal cylinders, illuminated from a fixed light source, changes with their position relative to the viewer. This rendered image is based on the Ward Duer model, an anisotropic BRDF model for brushed metal effects (Ward, 1992; Walter, 2005).

Two important functions of the view and Sun positions, which are used often in the following, are the angle between the Sun and view vectors ( $\xi$ ) and the related 'distance' between the Sun and view vectors ( $D$ ) as defined by:

$$\cos \xi = \cos \theta_i \cos \theta_v + \sin \theta_i \sin \theta_v \cos \phi$$

and

$$D = \left[ \tan^2 \theta_i + \tan^2 \theta_v - 2 \tan \theta_i \tan \theta_v \cos \phi \right]^{1/2}$$

## A2.2 Primary AMBRALS Kernel Models

The base components of the AMBRALS kernel models (Wanner *et al.*, 1995) can either be used as individual kernels of combinations or as described in Appendix A2.2.8. They are as follows:

### A2.2.1 Ross-thin (vol) model

Derived in Wanner *et al.* (1995) based on Ross (1981):

$$\cos \xi = \cos \theta_i \cos \theta_v + \sin \theta_i \sin \theta_v \cos \phi$$

$$k_{thin} = \frac{(\pi / 2 - \xi) \cos \xi + \sin \xi}{\cos \theta_i \cos \theta_v} - \frac{\pi}{2}$$

### A2.2.2 Ross-thick (vol) model

Derived in Roujean *et al.* (1992) based on Ross (1981):

$$k_{thick} = \frac{(\pi / 2 - \xi) \cos \xi + \sin \xi}{\cos \theta_i + \cos \theta_v} - \frac{\pi}{4}$$

### A2.2.3 Roujean (geo) model

Original Roujean model was 'Roujean' + Ross-thick kernel model. Here 'Roujean' refers to the geometric kernel:

$$D = \left[ \tan^2 \theta_i + \tan^2 \theta_v - 2 \tan \theta_i \tan \theta_v \cos \phi \right]^{1/2}$$

$$k_{brick} = \frac{1}{2\pi} \left[ (\pi - \phi) \cos \phi + \sin \phi \right] \tan \theta_i \tan \theta_v - \frac{1}{\pi} (\tan \theta_i + \tan \theta_v + D)$$

Clearly, if the Sun and view positions align (so that  $\theta_v = \theta_i$  and  $\phi = 0$ ) then both  $\xi$  and  $D$  equal zero. To see how they operate away from this 'hotspot' point it is useful to consider the 'Principal azimuth' (or almucantar) for which  $\theta_v = \theta_i$ . That is left to you.

### A2.2.4 Li-sparse (geo) model

The Li-sparse and Li-dense kernels are derived in Wanner *et al.* (1995). A detailed mathematical outline of the computation (the derivation is discussed in Appendix A2.4.3) is as follows:

- sequence of steps for calculations:

$$\theta'_i = \tan^{-1} \left( \frac{b}{r} \tan \theta_i \right)$$

$$\cos \xi' = \cos \theta'_i \cos \theta'_v + \sin \theta'_i \sin \theta'_v \cos \phi$$

$$D' = \left[ \tan^2 \theta'_i + \tan^2 \theta'_v - 2 \tan \theta'_i \tan \theta'_v \cos \phi \right]^{1/2}$$

$$\text{cost} = \min \left\{ 1, \frac{h}{b} \frac{\left[ D'^2 + (\tan \theta'_i \tan \theta'_v \sin \phi)^2 \right]^{1/2}}{\sec \theta'_i + \sec \theta'_v} \right\}$$

$$t = \cos^{-1}(\text{cost})$$

$$O\left(\frac{b}{r}, \frac{h}{b}, \theta_i, \theta_v, \phi\right) = \frac{1}{\pi} (t - \sin t \cos t) (\sec \theta'_i + \sec \theta'_v)$$

- The Li-sparse kernel follows from these definitions as:

$$Li\_sp = O - \sec \theta'_i - \sec \theta'_v + \frac{1}{2} (1 + \cos \xi') \sec \theta'_v$$

### A2.2.5 Li-dense (geo) model

The Li-dense approximation refers to dense vegetation canopies. Using the same definitions as for the Li-sparse kernel it is:

$$Li\_dn = \frac{(1 + \cos \xi') \sec \theta'_v}{\sec \theta'_i + \sec \theta'_v - O} - 2$$



### A2.2.6 Cox-Munk (geo) model

Based on the Cox-Munk (1954) model for sunglint and presented in Strahler *et al.* (1996):

$$k_{\text{spec}} = \begin{cases} \frac{1}{\cos \theta_i} \left( 1 - \frac{\tan^2 \theta_n}{\sigma^2} \right) - 1 & \text{if } \frac{\tan^2 \theta_n}{\sigma^2} \leq 1 \\ -1 & \text{else} \end{cases}$$

where  $\sigma^2$  is the wave slope variance (nonlinear parameter):

$$\sigma^2 = 0.003 + 0.00512 w$$

$w$  is the wind speed and the parameter  $\theta_n$  is defined as:

$$\cos^2 \theta_n = \frac{1}{2} \left( \frac{(\cos \theta_i + \cos \theta_v)^2}{1 + \cos \phi} \right)$$

### A2.2.7 Walthall (vol) model

This empirical model (also called the ‘limaçon’ function) was due to Walthall *et al.* (1985) and improved by Nilson and Kuusk (1989) to include reciprocity. It has four linear parameters:

$$\rho(\theta_i, \theta_v, \phi) = p_0 (\theta_i^2 + \theta_v^2) + p_1 \theta_i^2 \theta_v^2 + p_2 \theta_i \theta_v \cos \phi + p_3$$

### A2.2.8 MODIS products

Best fitting model from a selection of kernels (AMBRALS) include:

- Ross-thin + Li-sparse;
- Ross thin + Li-dense;
- Ross-thick + Li-sparse;
- Ross-thick + Li-dense;
- Cox-Munk + Li-sparse; and
- Walthall.

The first four of these have parameters that can be converted to ‘physical’ parameters.

## A2.3 Extensions and Alternatives

### A2.3.1 Reciprocal Li models

The proposed MODIS AMBRALS approach and the later choice of a single pair of Ross and Li models for operational MODIS BRDF products also used modified definitions of the kernels. The change was to make the models ‘reciprocal’ in that if the Sun and observer positions are interchanged the resulting

$$k_{\text{Sparse, Recip}} = (O - 1) \left( \sec \theta'_i + \sec \theta'_v \right) + \frac{1}{2} (1 + \cos \xi') \sec \theta'_i \sec \theta'_v$$

$$k_{\text{Dense, Recip}} = \frac{(1 + \cos \xi') \sec \theta'_i \sec \theta'_v}{(1 - O) (\sec \theta'_i + \sec \theta'_v)} - 2$$

BRDF value is unchanged. Reciprocity of the BRDF models is a fundamental property of the models (see Appendices A2.4.3 and A2.4.4), from which the Li kernels derive. Although the changes do not have a physical basis, the models are thought to be ‘better’ for having reciprocity. The changes were (for the same definitions of terms as before):

### A2.3.2 RPV model

Rahman, Pinty and Verstraete (RPV) model derived in Rahman *et al.* (1993) based on Pinty *et al.* (1990) and Pinty and Verstraete (1991):

$$\rho(\theta_i, \theta_v, \phi) = \rho_0 M(\theta_i, \theta_v, k) P(g, \xi) H(\rho_0, D)$$

$$M(\theta_i, \theta_v, k) = \frac{\cos^{k-1} \theta_i \cos^{k-1} \theta_v}{(\cos \theta_i + \cos \theta_v)^{1-k}}$$

$$P(g, \xi) = \frac{1 - g^2}{[1 + g^2 - 2g \cos(\pi - \xi)]^{1.5}}$$

$$H(\rho_0, D) = 1 + \frac{1 - \rho_0}{\delta + D}$$

where the linear parameter is  $\rho_0$  and  $g, k$  [and  $\delta$ ] are nonlinear parameters (generally  $\delta \approx 1$ ).  $M$  is the Minnaert model,  $P$  is the Henyey Greenstein phase function and  $H$  is the hotspot function.

### A2.3.3 MRPV model

Multiangle Imaging Spectroradiometer (MISR) version of RPV model as suggested by Martonchik *et al.* (1998):

$$P(g, \xi) = P_M(\xi) = e^{-b_M \cos \xi}$$

$$H(\bar{\rho}, D) = 1 + \frac{1 - \bar{\rho}}{1 + D}$$

where  $\bar{\rho}$  is an average reflectance assumed known if the algorithm is 'linearised' by taking logarithms; otherwise a simple nonlinear iteration can be used.

### A2.3.4 Staylor and Suttles model

As given in Cosnefroy *et al.* (1996), following Staylor and Suttles (1986):

$$\rho(\theta_i, \theta_v, \phi) = \frac{1}{\cos \theta_i \cos \theta_v} \left[ Y_0 + Y_1 \left( \frac{\cos \theta_i \cos \theta_v}{\cos \theta_i + \cos \theta_v} \right)^N \right] \times \frac{1 + C_{sw} \cos^2 \xi}{1 + C_{sw} [\cos^2 \theta_i \cos^2 \theta_v + (\sin^2 \theta_i \sin^2 \theta_v) / 2]}$$

where  $Y_0$  and  $Y_1$  are two linear parameters and  $C_{sw}$  and  $N$  are two nonlinear parameters. This model is very flexible and quite useful. The nonlinear parameters seem to converge quite quickly.

### A2.3.5 Shibayama and Weigand model

Developed in Shibayama and Weigand (1985, 1986) and used in Qi *et al.* (1995):

$$\rho(\theta_i, \theta_v, \phi) = \rho_0 (1 + [\beta_0 + \beta_1 \sin(\phi / 2) + \beta_2 (1 / \cos \theta_i)] \sin \theta_v)$$

### A2.3.6 Dymond and Qi model

Derived by Dymond and Qi (1997):

$$S(\theta_i, \theta_v) = \frac{\cos \theta_i}{\cos \theta_i + (\sigma_{\theta_i} / \sigma_{\theta_v}) \cos \theta_v}$$

$$H(\xi, \theta_i) = \begin{cases} 2 e^{-\tan(\xi/2)/h\theta_i} & \xi < \pi/2 \\ 2 e^{-\tan(\pi/4)/h\theta_i} & \xi \geq \pi/2 \end{cases}$$

$$B(\xi) = \frac{4\rho_0}{3\pi^2} (\sin \xi + (\pi/2 - \xi) \cos \xi)$$

$$\rho(\theta_i, \theta_v, \xi) = S(\theta_i, \theta_v) H(\xi, \theta_i) B(\xi)$$

Parameters are one linear ( $\rho_0$ ) and two nonlinear ( $h$  and  $R$ ) where  $R = \sigma_i / \sigma_v$

### A2.3.7 Chen modification to Roujean model

In Chen and Cihlar (1997) a modification of the Roujean model is proposed based on a simplification of Chen's more complex canopy model. The simple model reduces to:

$$\rho(\theta_i, \theta_v, \phi) = (f_{iso} + f_{geo} k_{brick} + f_{vol} k_{thick}) (1 + C_1 e^{-C_2 \frac{\xi}{\pi}})$$

where the constants  $C_1$  and  $C_2$  must be determined from the data. This added flexibility allows the Roujean model to approximate the hotspot effect much more successfully than before. Perhaps choosing a better model would be even better!

The Chen modification could, of course, be applied to any of the Kernel models to 'sharpen' the hotspot. However, there are two extra parameters to estimate.

### A2.3.8 Liang modification to Walthall model

In a similar way to the modification of the Roujean model by Chen to improve modelling of the hotspot, Liang and Strahler (1994) modified the Walthall model to include a hotspot effect. They increased the number of parameters from four to six by proposing a linear sum of the two models in the form:

$$\rho(\theta_i, \theta_v, \phi) = \rho_1(\theta_i, \theta_v, \phi) + \rho_2(\theta_i, \theta_v, \phi)$$

where the first is the Walthall model described above and:

$$\rho_2(\theta_i, \theta_v, \phi) = C_1 e^{-C_2 \tan(\pi - \xi)}$$

where, as above,  $C_1$  and  $C_2$  are adjustable parameters. In this case, one is non-linear and the hotspot shape is added rather than used multiplicatively.

Jupp (1997) described how a Walthall model plus a Li-sparse kernel could also effectively model a woodland BRDF with the Walthall accommodating the Volume effect and the Li-sparse modelling the hotspot.

### A2.3.9 Pickup and Chewings model

Pickup *et al.* (1995a,b) developed and extended an approach similar to one originally reported by Royer *et al.* (1985) to correct Video images for BRDF and other angular effects.

A simple linear ‘kernel’ model of the form:

$$\rho(\theta_i, \theta_v, \phi) = p_0 + p_1 \xi + p_2 \xi^2 + p_3 \cos^4 \theta_v$$

Was fitted to a line of Video data frames and then used to normalise the individual frames in the line. The  $\cos^4$  term was introduced to account for lens effects but can also model a Volume effect. The authors claim the power is insensitive to choices between 2 and 4.

## A2.4 Slightly More Complex Models: Not in ‘Kernel’ Form

### A2.4.1 Otterman model

Used in Deering *et al.* (1990) and based on previous work by Otterman (such as Otterman and Weiss, 1984), where the geometry is based on thin vertical cylinders:

$$\begin{aligned} \rho &= R_g e^{-s(\tan \theta_i + \tan \theta_v)} + R_p (1 - e^{-s(\tan \theta_i + \tan \theta_v)}) \\ R_g &= (1 - f) \\ &\times \frac{r(\sin \phi - \phi \cos \phi) + t((\phi - \pi) \cos(\phi - \pi) - \sin(\phi - \pi))}{4(\cot \theta_i + \cot \theta_v)} \\ &+ fr_0 \\ R_p &= \frac{r_p(\sin \phi - \phi \cos \phi)}{4(\cot \theta_i + \cot \theta_v)} \end{aligned}$$

where the parameters to be modelled are:

$s$  or ‘cylinder area index’; and  
 $r_p$  or plant reflectance.

and three soil reflectance parameters:

$$\begin{aligned} L &= fr_0 \\ B &= (1 - f)r \\ T &= (1 - f)t \end{aligned}$$

where

$f$  is the Lambertian fraction;  
 $r_0$  is the Lambertian reflectance;  
 $r$  is facet reflectance; and  
 $t$  is facet transmittance.

That is, five parameters in all. Four of the parameters are linear but  $s$  is nonlinear.

### A2.4.2 Verstraete, Pinty and Dickinson (VPD) model

Derived in Verstraete *et al.* (1990) and Pinty *et al.* (1990). This is a more complex form than the RPV model given above.

$$\rho(\theta_i, \theta_v, \phi) = \frac{\omega}{4 \kappa_i \mu_v + \kappa_v \mu_i} \times \left[ P_v(D)P(g, \xi) + H\left(\frac{\mu_i}{\kappa_i}, \omega\right)H\left(\frac{\mu_v}{\kappa_v}, \omega\right) - 1 \right]$$

where

$$\mu_i = \cos \theta_i$$

$$\mu_v = \cos \theta_v$$

Following Dickinson *et al.* (1990) after Goudriaan (1977):

$$\kappa_x = \Psi_1 + \Psi_2 \mu_x$$

$$\Psi_1 = 0.5 - 0.489 \chi_i - 0.33 \chi_i^2$$

$$\Psi_2 = 1 - 2\Psi_1$$

where

$\chi_i$  is a function of the leaf angle distribution of the canopy and varies from -0.4 for an erectophile canopy to 0.6 for a planophile canopy. Random orientation is zero.

$$P_v(D) = \frac{1}{1 + V_p(D)}$$

$$V_p(D) = 4 \left( 1 - \frac{4}{3\pi} \right) \frac{D}{2r\Lambda} \frac{\mu_v}{\kappa_v}$$

For Isotropic:  $P(g, \xi) = 1$

For Henyey and Greenstein:

$$P(g, \xi) = \frac{1 - g^2}{[1 + g^2 - 2g \cos(\pi - \xi)]^{1.5}}$$

For Legendre:

$$P(g, \xi) = 1 + g \cos \xi + \frac{L_2}{2} (3 \cos^2 \xi - 1)$$

$$H(x, \omega) = \frac{1 + x}{1 + x \sqrt{1 - \omega}}$$

The parameters for this model that need to be fitted in inversion or supplied for a given land surface are:

- $\omega$  is the single scattering albedo;
- $g$  is the asymmetry of the phase function (and  $L_2$  if Legendre);
- $\chi_i$  is the scatterer orientation parameter (used to obtain  $\kappa_i$  and  $\kappa_v$ ); and
- $2r\Lambda$  is the structural parameter ( $r$  is the sunfleck radius and  $\Lambda$  is the scatterer area density).

### A2.4.3 Strahler and Jupp simple model

This is derived as in Strahler and Jupp (1991) but here using overlap function derived for the Li kernels (see Appendices A2.2.4 and A2.4.4). In this model, there are four kinds of ground cover 'visible' from a given direction. These are referred to as scene components and consist of sunlit canopy (C), shaded canopy (T), sunlit background (G), and shaded background (Z). Each component is assumed to have a characteristic reflectance and the reflectance of a pixel is modelled as the area weighted combination (or linear mixture) of the characteristic component reflectances. That is, the observed reflectance of a single pixel is modelled as:

$$\rho(\theta_i, \theta_v, \phi) = k_C R_C + k_T R_T + k_G R_G + k_Z R_Z$$

where C, T, G, and Z indicate the reflectances of the four components as named above,  $R_x$  represents the (mean) radiance of component 'x' and  $k$  indicates the sensed proportion of each component within the pixel from the given view direction.

Obviously, with these definitions:

$$k_x \geq 0$$

$$\sum_{x=C,T,G,Z} k_x = 1$$

If the proportions are replaced by their expectations then the various components can be derived from the equations:

$$k_G = e^{-\lambda \bar{A} (\sec \theta'_i + \sec \theta'_v - 0)}$$

$$k_C + k_T = 1 - e^{-\lambda \bar{A} \sec \theta'_v}$$

which, given the closure above, means there is one extra condition to complete the model. This was done one way in Strahler and Jupp (1991) and in a second way in Li and Strahler (1992).

In Strahler and Jupp (1991) it was assumed that the proportion of visible sunlit tree was the same as for a single tree. That is:

$$\frac{k_C}{k_C + k_T} = \frac{1}{2} (1 + \cos \xi')$$

This model is the basis for the simplified Li-sparse formulation. In the full model, the four signatures must be determined as well as the ratios  $b/r$  and  $h/b$  and the vertical projected 'crown area index', or cover,  $\lambda A$ .

There are some simplifications. For example, in cases where the background is bright and trees dark it is possible to write:

$$\rho(\theta_i, \theta_v, \phi) \approx R_x + k_g (R_g - R_x)$$

where  $X$  refers to a composite of tree plus shade, which is assumed dark. Obviously,

$$R_x = \frac{k_c R_c + k_t R_t + k_z R_z}{1 - k_g}$$

This simple two-component model can often do a very good job. If, however, the variation is dependent on the differences between sunlit and shaded tree then the model will not do so well. It is often better just to assume that shaded tree and shaded background have the same colour. This reduces the number of parameters and is usually a reasonable assumption.

In the Li-sparse kernel given above and derived in Wanner *et al.* (1995), it is assumed that  $R_t = R_z = 0$  and  $R_c = R_g$ . Then:

$$\begin{aligned} \rho(\theta_i, \theta_v, \phi) &\approx (k_g + k_c) R_c \\ &= R_c \left[ \frac{1}{2} (1 + \cos \xi') (1 - e^{-\lambda \bar{A} \sec \theta'_v}) + e^{-\lambda \bar{A} (\sec \theta'_i + \sec \theta'_v - O)} \right] \\ &\approx R_c \lambda \bar{A} \left[ \frac{O - \sec \theta'_i - \sec \theta'_v}{2} + \frac{1}{2} (1 + \cos \xi') \sec \theta'_v \right] + R_c \\ &= c_1 k_{\text{sparse}} + c_2 \end{aligned}$$

which is the form previously given.

#### A2.4.4 Li's 'top easy seen' modification

Li and Strahler (1992) modified the simple model to take account of the fact that when the density of trees increases then the shaded crown tends to be 'hidden' and the view becomes dominated by the sunlit crown tops.

Of a number of choices, the one used for the simple dense canopy models assumes that:

$$f = \frac{k_c}{1 - k_g} = F$$

where  $F$  is the ratio for a single crown or rather the very dense case. It follows that:

$$F = \frac{\frac{1}{2} (1 + \cos \xi') \sec \theta'_v}{\sec \theta'_v + \sec \theta'_i - O}$$

The model may now be resolved and even inverted as it stands or it may be approximated in various ways with simpler forms. This model is used in dense vegetation and the previous one in sparse vegetation.

Note that this is equivalent to:

$$\begin{aligned} \frac{k_c}{k_c + k_t} &= F \frac{1 - k_g}{k_c + k_t} \\ &= \frac{\frac{1}{2} (1 + \cos \xi') \sec \theta'_v}{\sec \theta'_v + \sec \theta'_i - O} \frac{1 - e^{-\lambda \bar{A} (\sec \theta'_v + \sec \theta'_i - O)}}{1 - e^{-\lambda \bar{A} \sec \theta'_v}} \end{aligned}$$

which asymptotes to the same result as before when the density is low ( $1 - e^{-x} \approx x$ ).

To derive the Li-dense kernel (Wanner *et al.*, 1995), suppose  $R_g$  is neglected since there is so little sunlit soil showing in dense vegetation and also the forest floor may be dark material. Assume all shadow and  $R_g$  is zero. Then:

$$\begin{aligned} \rho(\theta_i, \theta_v, \phi) &= k_c R_c \\ &\approx f R_c \\ &= \left[ \frac{(1 + \cos \xi') \sec \theta'_v}{\sec \theta'_v + \sec \theta'_i - O} - 2 \right] \frac{R_c}{2} + R_c \\ &= c_1 k_{\text{dense}} + c_2 \end{aligned}$$

which is again in the form of the Li-dense kernel approximation. But again, it is possible also to solve the complete model or a model with some less restrictive assumptions if the linear kernel structure is not enforced.

#### A2.4.5 Hapke BRDF model for soils

Hapke (1981, 1986) derived a model for dimensionless particles, which has principally been applied to Soil data and is used also in the RPV and VPD models. The model was:

$$\begin{aligned} \rho(\theta_i, \theta_v, \phi) &= \frac{\omega}{4} \frac{1}{\mu_v + \mu_i} \\ &\times \left[ \frac{1 + B(\xi)}{1 + H(\mu_i, \omega) H(\mu_v, \omega) - 1} P(g, \xi) \right] \end{aligned}$$

where

$$\begin{aligned} \mu_i &= \cos \theta_i \\ \mu_v &= \cos \theta_v \end{aligned}$$

$B(\xi)$  is a backscattering function that accounts for the hotspot effect:

$$B(\xi) = \frac{S(0)}{\omega P(g, 0)} \frac{1}{\left[1 + (1/h) \tan(\xi/2)\right]}$$

Where  $S(0)$  defines the magnitude of the 'hotspot' and  $h$  defines the width.

$P(g, \xi)$  is the phase function for the particle collection with asymmetry  $g$ :

$$P(g, \xi) = \frac{1 - g^2}{\left[1 + g^2 - 2g \cos(\pi - \xi)\right]^{1.5}}$$

(Heney and Greenstein)

$$P(g, \xi) = 1 + g \cos \xi + \frac{L_2}{2} (3 \cos^2 \xi - 1)$$

(Legendre)

In the Legendre case there is an extra parameter  $L_2$ .

$H(x, \omega)$  is a function to account for multiple scattering:

$$H(x, \omega) = \frac{1 + 2x}{1 + 2x\sqrt{1 - \omega}}$$

The parameters for this model that need to be fitted in inversion or supplied for a given land surface are:

- $\omega$  is the single scattering albedo;
- $g$  is the asymmetry of the phase function (plus  $L_2$  for the Legendre phase function);
- $S(0)$  is a parameter defining the height of the hotspot function at the hotspot; and
- $h$  is a parameter that controls the width of the hotspot function.

#### A2.4.6 Ross' simplified vegetation canopy formula

The 'turbid medium' formula by Ross (1981) which was used by Roujean *et al.* (1992) to derive a Volume kernel (the Ross-Thick) and by Wanner *et al.* (1995) to derive another (the Ross-Thin) kernel was simplified by both groups of writers to the following form as an initial step in their formulation:

$$\rho(\theta_i, \theta_v, \phi) = \frac{4\rho_L (\pi/2 - \xi) \cos \xi + \sin \xi}{3\pi (\cos \theta_i + \cos \theta_v)} \times \left(1 - e^{-\frac{LAI}{2}(\sec \theta_i + \sec \theta_v)}\right) + \rho_s e^{-\frac{LAI}{2}(\sec \theta_i + \sec \theta_v)}$$

where

- $\rho_L$  is the reflectance of a single leaf;
- $\rho_s$  is the background soil reflectance; and
- $LAI$  is Leaf Area Index

If these three parameters are estimated subject to the reflectances being in the range (0,1) and  $LAI > 0$  then this provides a model that includes both the Ross-thin and the Ross-thick (Wanner *et al.*, 1995). As with all these non-kernel models, however, the fit will need to be nonlinear. However, this often presents little problem.

#### A2.4.7 Minnaert's original model

Minnaert (1941) proposed a model that has been used for modelling topographic shading and has been used as a component in a number of semi-empirical models. Its form is:

$$\rho(\theta_i, \theta_v) = \rho_L \frac{k+1}{2} \mu_i^{k-1} \mu_v^{k-1}$$

This function is a reciprocal 'volume' scattering kernel in its behaviour with  $k=1$  being Lambertian,  $k=0$  being a 'bowl' with highest reflectances away from nadir and  $k=2$  being an inverted 'bowl' with darkening away from nadir.

Parameters to estimate are  $\rho_L$  and  $k$ .

Topographic shading and large pixels, although not described in detail here, can also be modelled as a BRDF effect. In addition to the Minnaert model, the model by Hapke (Hapke, 1984) has been used successfully in this case.



## A2.5 References

- Cosnefroy, H., Leroy, M., and Briottet, X. (1996). Selection and characterisation of Saharan and Arabian Desert sites for the calibration of optical satellite sensors. *Remote Sensing of Environment*, 58, 101–114.
- Chen, J.M., and Cihlar, J. (1997). A hotspot function in a simple bidirectional reflectance model for satellite applications. *Journal of Geophysical Research*, 102(D22), 25907–25913. doi: 10.1029/97JD02010.
- Cox, C., and Munk, W. (1954). Measurement of the roughness of the sea surface from photographs of the sun's glitter. *Journal of the Optical Society of America*, 44, 838–850.
- Deering, D.W., Eck, T.F., and Otterman, J. (1990). Bidirectional reflectances of selected desert surfaces and their three-parameter soil characterisation. *Agricultural and Forest Meteorology*, 52, 71–93.
- Dickinson, R.E., Pinty, B., And Verstraete, M. (1990). Relating surface albedos in GCM to remotely sensed data. *Agricultural and Forest Meteorology*, 52, 109–131.
- Dymond, J.R., and Qi, J. (1997). Reflection of visible light from a dense vegetation canopy - a physical model. *Agricultural and Forest Meteorology*, 86, 143–155.
- Goudriaan, J. (1977). *Crop micrometeorology: a simulation study*. Wageningen: Wageningen Centre for Agricultural Publishing and Documentation.
- Hapke, B.W. (1981). Bidirectional reflectance spectroscopy. 1. Theory. *Journal of Geophysical Research*, 86, 3039–3054.
- Hapke, B.W. (1984). Bidirectional reflectance spectroscopy. 3. Correction for macroscopic roughness. *Icarus*, 59, 41–59.
- Hapke, B.W. (1986). Bidirectional reflectance spectroscopy. 4. The extinction coefficient and the opposition effect. *Icarus*, 67, 264–280.
- Jupp, D.L.B. (1997). Some Research and Applications in the CSIRO (Australia) Earth Observation Centre on Scene Brightness due to BRDF. *ENAMORS'97*, Tuusula, Finland.
- Jupp (2000). *A Compendium of Kernel and Other (Semi-) Empirical BRDF Models*. CSIRO Earth Observation Centre, Canberra.
- Li, X., and Strahler, A.H. (1992). Geometrical-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: Effect of crown shape and mutual shadowing. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 276–292.
- Liang, S., and Strahler, A.H. (1994). Retrieval of surface BRDF from multiangle remotely sensed data. *Remote Sensing of Environment*, 50, 18–30.
- Martonchik, J.V., Diner, D.J., Kahn, R., Verstraete, M.M., Pinty, B., Gordon, H.R., and Ackerman, T.P. (1998). Techniques for the retrieval of aerosol properties over land and ocean using multiangle imaging. *IEEE Transactions on Geoscience and Remote Sensing*, 36, 1212–1227.
- Minnaert, M. (1941). The reciprocity principle in lunar photometry. *Astrophysical Journal*, 93, 403–410.
- Nilson, T., and Kuusk, A. (1989). A reflectance model for the homogeneous plant canopy and its inversion. *Remote Sensing of Environment*, 27, 157–167.
- Otterman, J., And Weiss, G.H. (1984). Reflections from a field of randomly located vertically protrusions. *Applied Optics*, 23, 1931–1936.
- Pickup, G., Chewings, V.H. and Pearce, G. (1995a). Procedures for correcting high resolution airborne video imagery. *International Journal of Remote Sensing*, 16(9), 1647–1662.
- Pickup, G., Bastin, G.N., Chewings, V.H. and Pearce, G. (1995b). Correction and classification procedures for assessing rangeland vegetation cover with airborne video data. *Proc. 15th Biennial Workshop on Color Photography and Video Graphy in Resource Assessment*, 2–3 May 1995, USA.
- Pinty, B., Verstraete, M.M. and Dickinson, R.E. (1990). A physical model of the bidirectional reflectance of vegetation canopies. 2. Inversion and validation. *Journal of Geophysical Research*, 95, 11767–11775.
- Pinty, B., And Verstraete, M.M. (1991). Extracting information on surface properties from bidirectional reflectance measurements. *Journal of Geophysical Research*, 96, 2865–2874.
- Qi, J., Moran, M.S., Cabot, F., and Dedieu, G. (1995). Normalization of Sun/View Angle Effects Using Spectral Albedo-based Vegetation Indices. *Remote Sensing of Environment* 52, 207–217.
- Rahman, H., Pinty, B., and Verstraete, M.M. (1993). Coupled surface-atmosphere reflectance (CSAR) model. 2. Semiempirical surface model usable with NOAA Advanced Very High resolution Radiometer data. *Journal of Geophysical Research*, 98, 20791–20801.
- Ross, J.K. (1981). *The radiation regime and architecture of plant stands*. Junk Publ., The Hague. 391pp.

- Roujean, J.L., Latoy, M., and Deschamps, P.Y. (1992). A bidirectional reflectance model of the earth's surface for the correction of remote sensing data. *Journal of Geophysical Research*, 97, 20455–20468.
- Royer, A., Vincent, P., and Bonn, F. (1985). Evaluation and correction of viewing angle effects on satellite measurements of bi-directional reflectance. *Photogrammetric Engineering and Remote Sensing*, 51, 1899–1914.
- Shibayama, M., and Wiegand, C.L. (1985). View azimuth and zenith, and solar angle effects using spectral albedo-based vegetation indices. *Remote Sensing of Environment*, 18, 91–103.
- Shibayama, M., and Wiegand, C.L. (1986). Diurnal patterns of bidirectional vegetation indices for wheat canopies. *International Journal of Remote Sensing*, 7, 233–246.
- Staylor, W.F., and Suttles, J.T. (1986). Reflection and emission models for deserts derived from Nimbus-7 ERB scanner measurements. *Journal of Climate and Applied Meteorology*, 25, 196–202.
- Strahler, A.H., and Jupp, D.L.B. (1991). Modeling bidirectional reflectance of forests and woodlands using boolean models and geometric optics. *Remote Sensing of Environment*, 34, 153–166.
- Strahler, A.H., Wanner, W., Barker Schaaf, C., Li, X., Hu, B., Muller, J-P., Lewis, P., and Barnsley, M.J. (1996). *MODIS BRDF/Albedo Product: Algorithm Theoretical Basis Document Version 4.0*. MODIS Product ID: MOD43, 252pp.
- Verstraete, M.M., Pinty, B., and Dickinson, R.E. (1990). A physical model of the bidirectional reflectance of vegetation canopies. 1. Theory. *Journal of Geophysical Research*, 95, 11755–11765.
- Walthall, C.L., Norman, J.M., Welles, J.M., Campbell, G., and Blad, B.L. (1985). Simple equation to approximate the bidirectional reflectance from vegetation canopies and bare soil surfaces. *Applied Optics*, 24, 383–387.
- Wanner, W., Li, X., and Strahler, A.H. (1995). On the derivation of kernels for kernel-driven models of bidirectional reflectance, *Journal of Geophysical Research*, 100, 21077–21089.

# Appendix 3

## Colour Systems and Sequences

Colour science is introduced in Volume 2A—Section 5. The following sub-sections elaborate on additional aspects of colour systems and sequences, namely:

- colour measurement (see Appendix A3.1);
- colour systems (see Appendix A3.2); and
- colour coordinates (see Appendix A3.3).

### A3.1 Colour Measurement

An interesting aspect of human vision is that it is the mixture of wavelengths that are perceived and not the individual components. This means that a given same visual sensation could be produced by different physical stimuli. (Colours which appear to be identical but have different spectral composition are called metamers). This phenomenon is used as the basis for psychophysical colour matching experiments in which a colour is defined as a mixture of three primary colours. An infinite number of sets of primary colours may be defined subject to the condition that any one may not be produced by a mixture of the other two. The red, green and blue lights used in the additive system were selected to obtain the maximum range of mixture colours. It should be noted however that this range is still less than the full range of colours that can be perceived by humans (see Volume 2A—Section 5).

In colour matching experiments, a reference colour is defined by the proportions of the primary lights required to reproduce it. Due to the limited mixture range of the primaries however, some colours cannot be reproduced this way. These colours are matched by adding one or two of the primaries to the reference colour and defining the reference by subtracting the proportion of those primaries added from the mixture used to match it. These proportions can then be plotted on a chromaticity diagram.

To avoid negative terms in these definitions, three artificial primaries (that is imaginary colours) were determined by the CIE (Commission Internationale d’Eclairage) which are referred to as X, Y and Z (see Volume 2A—Section 5.1.2 and Appendix A3.3). While real primaries are used to perform the actual colour matching then the proportions are transformed algebraically to X, Y, and Z values for plotting on the standardised CIE diagram, this system has the advantage of allowing the measurement to be made with different sets of primaries.

---

*Even in modern art, artists have used methods based on calculation, inasmuch as these elements, alongside those of a more personal and emotional nature, give balance and harmony to any work of art.*  
(Max Bill)

---

## A3.2 Colour Systems

### A3.2.1 RGB and HSI

The RGB (Red,Green,Blue) and HSI (Hue,Saturation,Intensity) models for colour specification were introduced in Volume 2A—Section 5.1.2. These models are related by their intensity axes. For the RGB cube, the intensity axis is defined by the line:

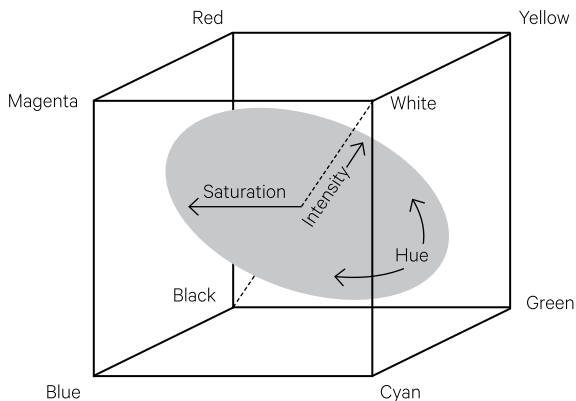
$$R = G = B$$

as illustrated in Figure A3.1a. Colours in the RGB colour space can also be defined as triangular coordinates on equal intensity planes across the cube (see Figure A3.1a). These coordinates can be normalised within each plane such that the values along each axis from the triangle centre to an apex vary from 0 to 1 (see Figure A3.1b).

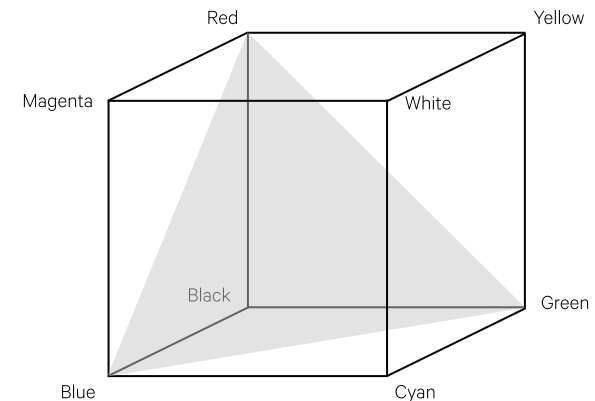
**Figure A3.1** RGB colour coordinates

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.

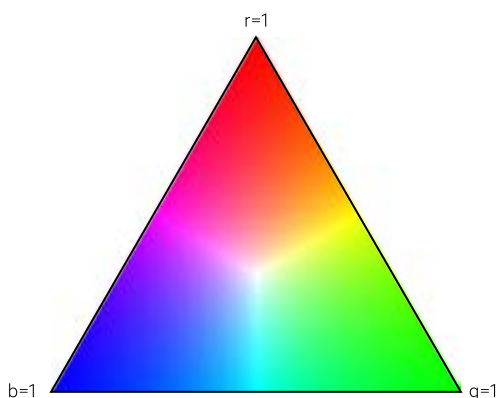
a. Orientation of RGB colour cube



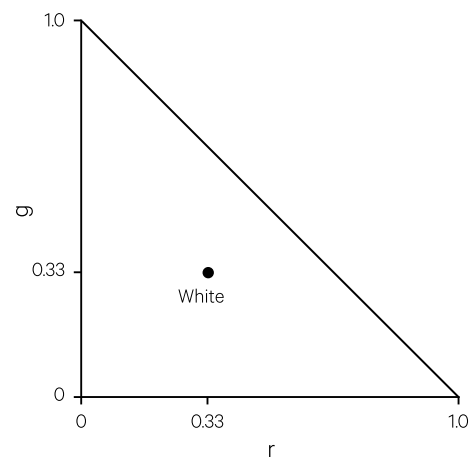
b. Definition of triangular planes in RGB cube



c. Normalised triangular coordinates



d. Normalised chromaticity coordinates. Since, for any colour  $r+g+b=1$ , only two components need to be known to deduce a third. For example at the white point,  $r=1/3$ ,  $g=1/3$ , so  $b=1-r-g=1/3$ .



Source: a. Harrison and Jupp (1990) Figures 11 and 102

Cartesian coordinates in the RGB cube can be converted to normalised triangular coordinates by the equations:

$$r = \frac{R}{R + G + B}$$

$$g = \frac{G}{R + G + B}$$

$$b = \frac{B}{R + G + B}$$

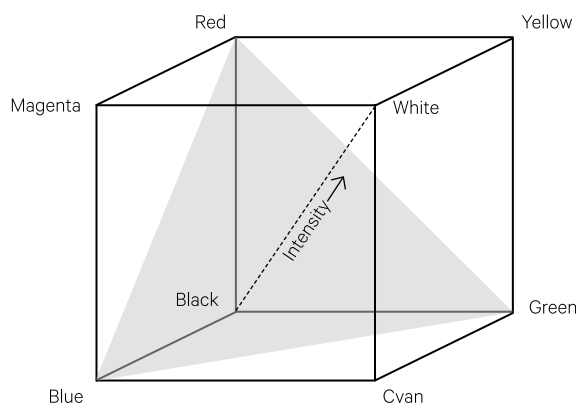
By definition, these normalised coordinates ( $r, g, b$ ) sum to one, with the intensity of the plane being equal to the sum of the RGB values:

$$I = R + G + B$$

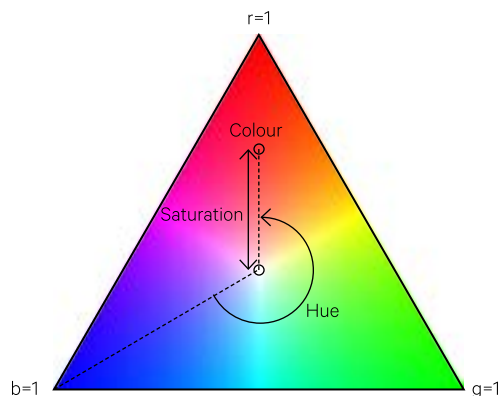
The HSI model defines colours on the equal intensity planes as coordinates of hue and saturation, with hue being measured as an angle around the plane and saturation as the radial distance from the centre of the plane (see Figure A3.2). Measurement of these parameters is detailed in Volume 2C—Section 9.3.2.

**Figure A3.2** HSI coordinates

a. Intensity is measured along the diagonal axis between black and white.



b. Saturation is the linear distance from the intensity axis to the selected colour on an equal intensity plane. Hue is the angular distance from a primary colour (usually blue) to the colour on an intensity plane.



Source: Harrison and Jupp (1990) Figure 104

### A3.2.2 XYZ and UVW

The rgb chromaticity coordinates were defined above as normalised proportions of R, G and B values. Since the normalised coordinates sum to one, only two are required to define a given colour—the third coordinate can be deduced from these two, for example:

$$b = 1 - r - g$$

The values of any two coordinates (usually  $r$  and  $g$ ) can then be usefully represented as a chromaticity diagram as shown in Figure A3.3.

CIE have defined colour-matching functions for monochromatic stimuli of wavelengths 700 nm ( $R_o$ ), 546.1 nm ( $G_o$ ) and 435.8 nm ( $B_o$ ), which can be used to match the spectral power distribution curve of any colour. These stimuli were defined to enclose the chromaticity diagram as illustrated in Figure A3.3.

For 2° standard observer data (that is, standard colour measurement data based on a 2° field of view) the CIE tristimulus values, X, Y and Z (see Volume 2A—Section 5.1.2), are related to  $R_o$ ,  $G_o$  and  $B_o$  (Hunt, 1987) as:

$$X = 0.49 \times R_o + 0.31 \times G_o + 0.20 \times B_o$$

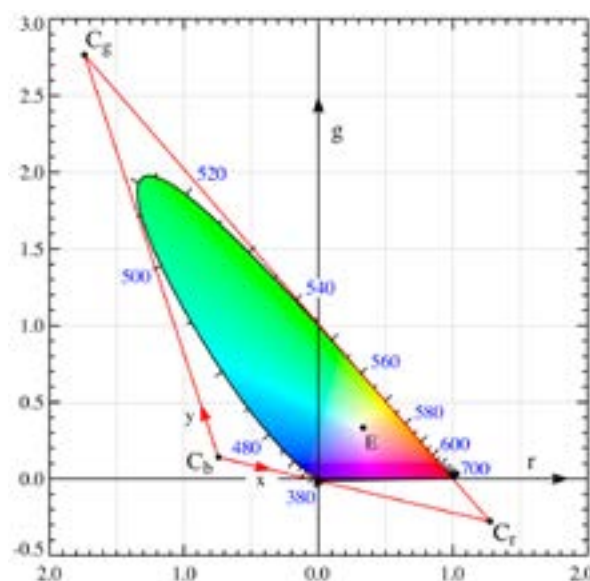
$$Y = 0.17697 \times R_o + 0.81240 \times G_o + 0.01063 \times B_o$$

$$Z = 0.0 \times R_o + 0.01 \times G_o + 0.99 \times B_o$$

For diagrammatic simplicity, the XYZ values are usually normalised in the same way as the RGB coordinates and expressed as xyz values. A chromaticity diagram of x and y can then be used to represent the colour gamut on Cartesian coordinates.

**Figure A3.3** CIE rg chromaticity space

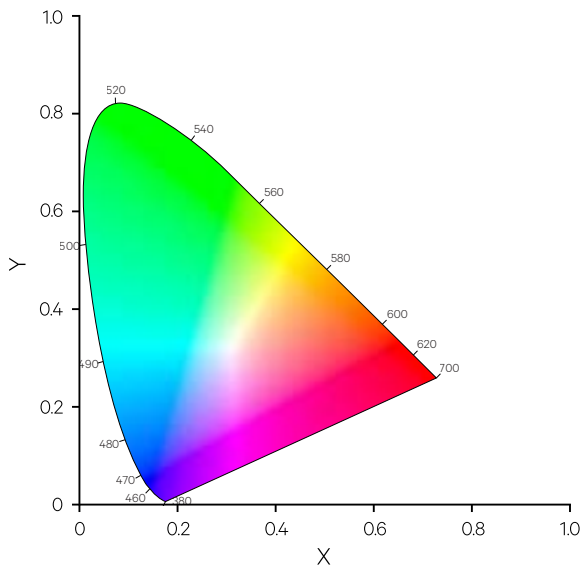
The triangle  $C_b$ - $C_g$ - $C_r$  is equivalent to the triangle (0,0)-(0,1)-(1,0) in the CIE xy chromaticity space.



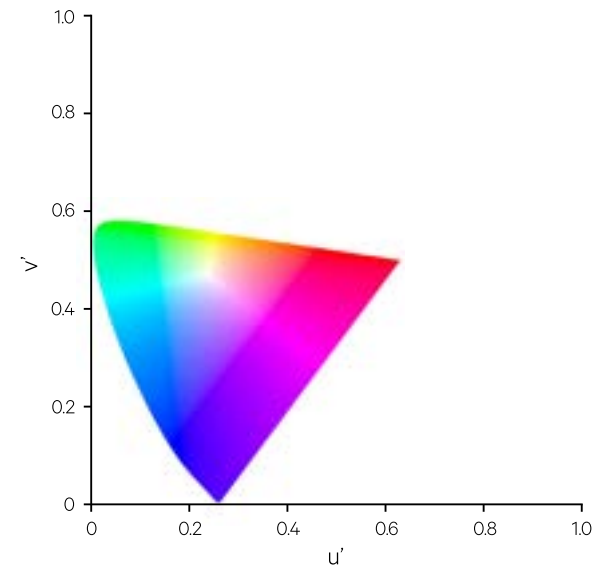
Retrieved from: [https://upload.wikimedia.org/wikipedia/commons/1/16/CIE1931\\_rgxy.png](https://upload.wikimedia.org/wikipedia/commons/1/16/CIE1931_rgxy.png)

**Figure A3.4** Perceptual colour differences

a. CIE 1931 xyz colour space



b. CIE 1976 uvw colour space. The linear distance between colours in this coordinate system can be used to indicate the perceptual difference between them.



Source: Colour space retrieved from [https://upload.wikimedia.org/wikipedia/commons/8/83/CIE\\_1976\\_UCS.png](https://upload.wikimedia.org/wikipedia/commons/8/83/CIE_1976_UCS.png)

As illustrated in Figure A3.4a, this xyz space is not uniform in terms of perceived colour differences. MacAdam (1937) proposed a 'uniform chromaticity' triangle using the coordinates U, V and W (see Figure A3.4b). The normalised coordinates u and v of this system are related to the xy coordinates by:

$$u = 2x / (6y - x + 1.5)$$

$$v = 3y / (6y - x + 1.5)$$

Conversion from xy to uv coordinates can be simplified by use of a nomogram (Hunt, 1975). The uv chromaticity diagram approximates a uniform colour space, so allows relative perceptual differences between two colours to be computed as the Euclidean distance between the points that represent the colours on this diagram.

### A3.3 Colour Coordinates

In many image processing systems, colours for image painting are specified as numeric values, such as in the range 0–511. Such numbers generally represent the proportion of the three additive primaries used in each colour. Using the standard CIE equations given in Appendix A3.2, these numbers can be transformed into uvw coordinates to show their distribution within the perceptible colour space. This conversion assumed that the proportions of each primary are directly related to the rgb coordinate scale, so do not account for the effect of device-specific gamuts.

To achieve maximum visual separation within a set of colours, the CIE uvw space needs to be sampled evenly before transforming the sample coordinates back to their RGB components. Given the extensive spectral resolutions of many colour display devices, the resulting range of colours should be well-represented. The reproduction of the display colour on hardcopy devices, however, can still be challenging (see Volume 2A—Sections 5.1.3 and 6).

### A3.4 References

- Harrison, B.A., and Jupp, D.L.B. (1990). *Introduction to Image Processing. Part TWO of the microBRIAN Resource Manual*. CSIRO, Melbourne. 256pp.
- Hunt, R.W.G. (1975). *The Reproduction of Colour*. Fountain Press, London.
- Hunt, R.W.G. (1987). *Measuring Colour*. Ellis-Horwood Ltd. Chichester, UK.

- MacAdam, D.L. (1937). Projective Transformations of I.C.I. Color Specifications. *Journal of the Optical Society of America*, 27(8), 294–299. doi: 10.1364/JOSA.27.000294



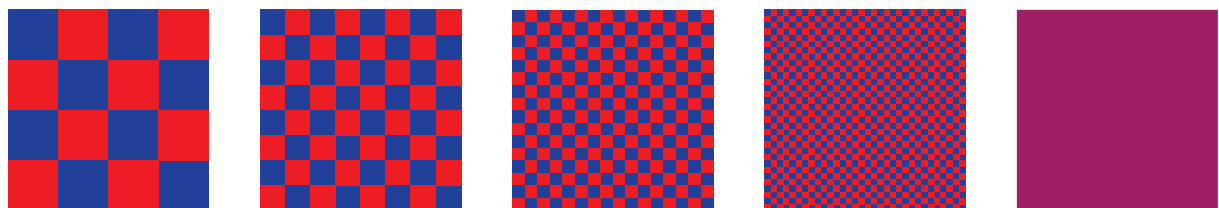
# Appendix 4

## Dithering

In computer graphics, to extend the perceived range of colours that is actually available from a limited colour palette on an output device, a technique known as ‘dithering’ can be used. This approach relies on the human eye blurring adjacent colours to perceive a mixed colour (see Figure A4.1). Dithering methods are comparable to halftone printing techniques and are commonly used to expand the range of colours represented by display hardware.

**Figure A4.1** Dithering example

This sequence of image use just two colours, red and blue, but as the area of each colour is reduced, the human eye blurs them to appear as the single colour purple.

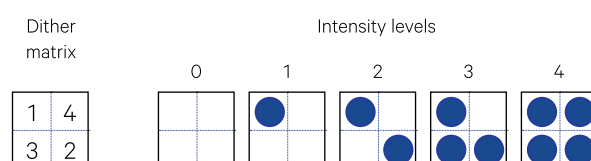


### A4.1 Dither Matrices

One form of dithering relies on using a ‘dither matrix’. The dithering technique used for inkjet printers, for example, is implemented by grouping together square or rectangular blocks of dots (called ‘nibs’), then using the blended colour of each group of dots to represent a single pixel in the image. The size of a dither block can vary and affects the number of colours that can be represented. A non-square image pixel may be more closely represented using a rectangular dither matrix such as 3×4 dots.

**Figure A4.2** Dithering sequence

For example, to print value 3 with an output value range 0–4 (such as in a 2×2 dither), dots are printed at positions 1, 2 and 3 in the dither matrix.



Source: Harrison and Jupp (1990) Figure 35

If we use a 2×2 dither block (or ‘printer pixel’) for each image pixel, we have increased the intensity variations for the pixel from 2 (one dot ON or OFF) to 5 (all 4 dots OFF, one dot ON, two dots ON, three dots ON or all 4 dots ON) as shown in Figure A4.2. The three primary inks allow  $2^3 = 8$  different colour possibilities for a single dot. In a 2×2 dither we can represent  $5^3 = 125$  different colours. Other dither sizes include:

- 2×3 (7 intensity levels:  $7^3 = 343$  colours);
- 3×3 (10 intensity levels:  $10^3 = 1,000$  colours);
- 3×4 ( $13^3 = 2,197$  colours); and
- 4×4 ( $17^3 = 4,913$  colours).

For these colour ranges to be effectively used in a printed image however, the digital image must have a correspondingly diverse range of data values.

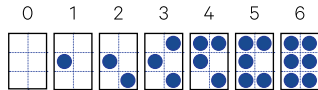
**Figure A4.3** Example dither matrices

a. 2×3

Dither matrix

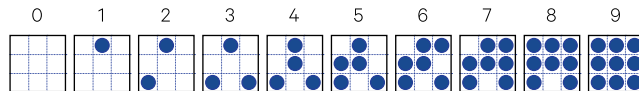
4	3
1	6
5	2

Intensity levels



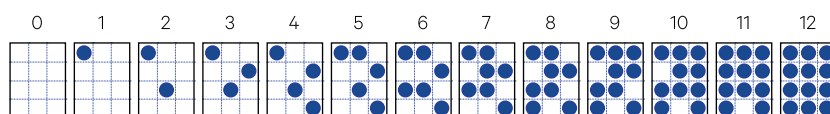
b. 3×3

8	1	6
5	4	7
2	9	3



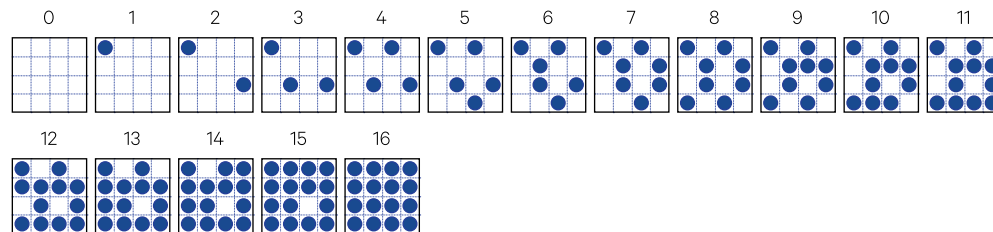
c. 3×4

1	5	9
11	7	3
6	2	10
8	12	4



d. 4×4

1	15	4	14
12	6	9	7
13	3	16	2
8	10	5	11



Source: Harrison and Jupp (1990) Figure 36

The ‘cost’ of this increased colour range is the number of pixels that may be plotted across the page. With a 2×2 dither we are using two dots across the page and two dots down to represent one printed pixel. If the printer allows only 1,024 dots across the page, then using a 2×2 or 2×3 dither we can only print 512 pixels. Similarly, a 3×3 or 3×4 dither would only allow 1024/3 or 341 pixels across the page.

The order of intensity changes within a dither block are usually coded in matrix format (hence the term dither matrix). The values at each dot position in the matrix indicate the intensity at which that dot will become activated. Any other dots with values below this value are also used. For example, in Figure A4.2, to print intensity level 3 we use the dots at positions 1, 2 and 3 in the dither. Some other dither matrices that have been used in image processing software are given in Figure A4.3. As discussed, the dither size determines the intensity range of a printed image. To make best use of the available contrast, enhancement techniques (as described in Volume 2A—Section 4.2) are usually applied to image data before mapping to this range.

## A4.2 Designed Order

The ordering within a dither matrix needs to be carefully selected to avoid geometric patterns (such as basket-weave, twill, or herringbone tweed) in areas of uniform intensity or low frequency blotchiness. Colour transitions in an image need to be smooth to give the impression of continuous variations in tones. One method for achieving this effect is to use dither matrices in which the rows and columns sum to approximately the same value. This criterion can only be achieved for square matrices but satisfactory results may still be obtained in other shapes.

The 4×4 dither matrix shown in Figure A4.3 is based on a pan-diagonal magic square or 'nasik' in which any pair of adjacent cells (horizontal, vertical or diagonal) do not contain sequential numbers. In this case all 32 four-element groups within the matrix (that is, four columns and rows, eight diagonals—including broken diagonals such as [3+5+12]+[10]) and sixteen 2×2 squares—all sum to 34 (Lippel and Kurland, 1971). This pattern avoids horizontal, vertical and diagonal line structures within the dither or across two adjacent dithers as illustrated in Figure A4.4.

The final result of applying dithering and inverse mapping to the complementary primaries results in the ability to plot apparently continuous tone imagery on inkjet technology. This operation is demonstrated in Figure A4.5.

**Figure A4.4** Formation of a balanced dither matrix

The 4×4 matrix above is based on a 'nasik': all groups of four adjacent cells, both within one dither and across adjacent dithers, sum to the same value.

16	2	13	3	16	2	13	3
5	11	8	10	5	11	8	10
4	14	1	15	4	14	1	15
9	7	12	6	9	7	12	6
16	2	13	3	16	2	13	3
5	11	8	10	5	11	8	10
4	14	1	15	4	14	1	15
9	7	12	6	9	7	12	6

Source: Harrison and Jupp (1990) Figure 37

**Figure A4.5** Colour representation using dithering techniques with a 2×2 dither.


The inverse intensity values of three channels are assigned to the subtractive primaries then plotted using dither matrices.

Example colour	Blue value	Yellow value	Yellow dither	Green value	Magenta value	Magenta dither	Red value	Cyan value	Cyan dither	Composite dither	Viewed colour
Blue	4	0		0	4		0	4			
Yellow	0	4		4	0		4	0			
White	4	0		4	0		4	0			
Orange	0	4		2	2		4	0			
Dark brown	1	3		0	4		2	2			
Aqua	2	2		4	0		0	4			

### A4.3 References

---

- Harrison, B.A., and Jupp, D.L.B. (1990). *Introduction to Image Processing. Part TWO of the microBRIAN Resource Manual*. CSIRO, Melbourne. 256pp.
- Lippel, B., and Kurland, M. (1971). The effect of dither on luminance quantisation of pictures. *IEEE Transactions on Communication Technology*, 19, 879–888.



# Appendix 5

## Filters

The use of spatial filters is introduced in Volume 2C. The following sub-sections detail a range of filters that have been used effectively with EO imagery:

- averaging filters—used to smooth imagery (see Appendix A5.1);
- directional filters—for edge enhancement, including exposure, insolation and curvature (see Appendix A5.2); and
- derivature filters—commonly used for geophysical applications (see Appendix A5.3)

### A5.1 Averaging Filters

Balanced average filters are formed from B-spline formulae to ensure a smooth graduation of weighting values (Jupp 1976). For example, a good 3×3 filter is generated from the function:

$$\begin{bmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{bmatrix}$$

as the outer product with its transpose. Larger filters are convolution products of the smaller ones (see Volume 2C—Section 3). For all sizes, the sum of all weighting values is used as a divisor to preserve scaling. A set of example filters is shown in Figure A5.1.

**Figure A5.1** Averaging filters

a. 3×3 filter with divisor 36

1	4	1
4	16	4
1	4	1

b. 5×5 filter with divisor 256

1	4	6	4	1
4	16	24	16	4
6	24	36	24	6
4	16	24	16	4
1	4	6	4	1

c. 7×7 filter with divisor 9,216

1	8	23	32	23	8	1
8	64	184	256	184	64	8
23	184	529	736	529	184	23
32	256	736	1024	736	256	32
23	184	529	736	529	184	23
8	64	184	256	184	64	8
1	8	23	32	23	8	1

d 9×9 filter with divisor 589,824

1	16	76	176	230	176	76	16	1
16	256	1216	2816	3680	2816	1216	256	16
76	1216	5776	13376	17480	13376	5776	1216	76
176	2816	13376	30976	40480	30976	13376	2816	176
230	3680	17480	40480	52900	40480	17480	3680	230
176	2816	13376	30976	40480	30976	13376	2816	176
76	1216	5776	13376	17480	13376	5776	1216	76
16	256	1216	2816	3680	2816	1216	256	16
1	16	76	176	230	176	76	16	1

e. 11×11 filter with divisor 58,982,400

1	32	237	832	1682	2112	1682	832	237	32	1
32	1024	7584	26624	53824	67584	53824	26624	7584	1024	32
237	7584	56169	197184	398634	500544	398634	197184	56169	7584	237
832	26624	197184	692224	1399424	1757184	1399424	692224	197184	26624	832
1682	53824	398634	1399424	2829124	3552384	2829124	1399424	398634	53824	1682
2112	67584	500544	1757184	3552384	4460544	3552384	1757184	500544	67584	2112
1682	53824	398634	1399424	2829124	3552384	2829124	1399424	398634	53824	1682
832	26624	197184	692224	1399424	1757184	1399424	692224	197184	26624	832
237	7584	56169	197184	398634	500544	398634	197184	56169	7584	237
32	1024	7584	26624	53824	67584	53824	26624	7584	1024	32
1	32	237	832	1682	2112	1682	832	237	32	1



## A5.2 Directional Filters

Various edge enhancement filters are presented in Volume 2C. Some additional directional filters are given below.

Prewitt (1970) suggests a bi-Laplacian operator

(a fourth derivative bidirectional filter:  $\frac{\partial^4}{\partial x^2 \partial y^2}$ ) as:

$$\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \times \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}$$

This 3×3 implementation of the operator is not recommended as its own spatial frequency is quite abrupt and would cause 'ringing' in the processed image. A set of second derivative filters with directional properties is shown in Figure A5.2.

Various directionally selective filters have been used to detect edges. These generally rely on the eight compass directions (North, North-east, East, South-east, South, South-west, West, North-west), which can be easily aligned with the image grid. These filters

can be applied in sets of eight with the maximum (absolute) filtered value being selected as the output value and the direction associated with the selected filter being the edge direction.

Frei and Chen (1977) defined a set of directional, edge-detecting filters with the weights shown in Figure A5.3, where  $a=\sqrt{2}$  (listed as anti-clockwise from north).

The Sobel and Prewitt operators could similarly be implemented in these eight directions by using a value of 'a' equal to 2 or 1 respectively. Another set of directional filter weights proposed by Prewitt (1970) are shown in Figure A5.4.

Kirsch (1971) defined a slightly different set of directional filters (see Figure A5.5).

Any of these filters could be used individually to enhance the edge information towards a particular direction. The exposure transformation (see Volume 2C and Appendix A5.2.1) implements directional filtering of imagery from a selected compass direction.

Figure A5.2 Derivative filters

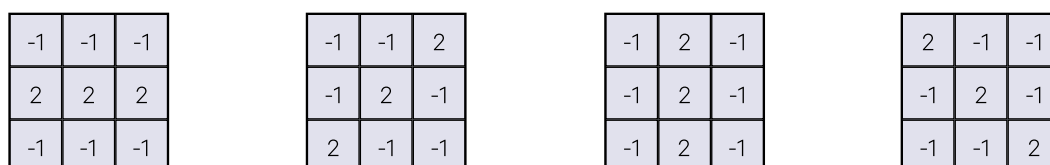


Figure A5.3 Directional, edge-detecting filters

$a=\sqrt{2}$

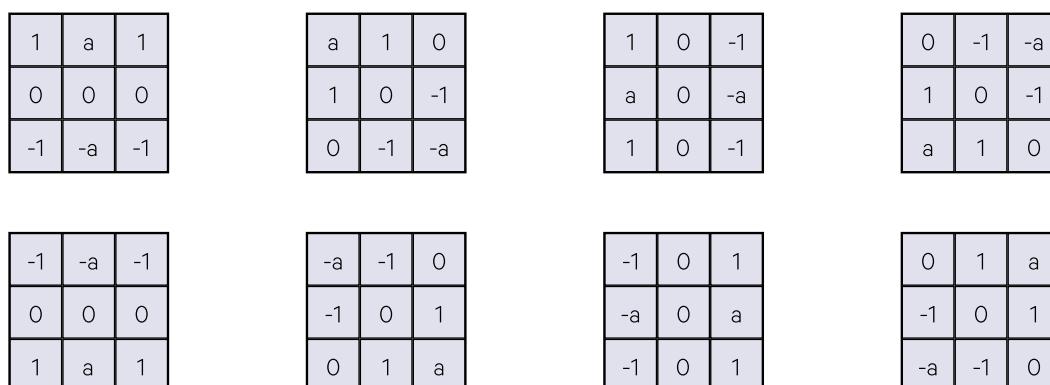


Figure A5.4 Directional filters

1	1	1
1	-2	1
-1	-1	-1

1	1	1
1	-2	-1
1	-1	-1

1	1	-1
1	-2	-1
1	1	-1

1	-1	-1
1	-2	-1
1	1	1

-1	-1	-1
1	-2	1
1	1	1
-1	-1	1
-1	-2	1
1	1	1
-1	1	1
-1	-2	1
-1	1	1
1	1	1
-1	-2	1
-1	-1	1

Figure A5.5 Kirsch directional filters

5	5	5
-3	0	-3
-3	-3	-3

5	5	-3
5	0	-3
-3	-3	-3

5	-3	-3
5	0	-3
5	-3	-3

-3	-3	-3
5	0	-3
5	5	-3

-3	-3	-3
-3	0	-3
5	5	5
-3	-3	-3
-3	0	5
-3	5	5
-3	-3	5
-3	0	5
-3	-3	5
-3	5	5
-3	0	5
-3	-3	-3

### A5.2.1 Exposure

As detailed in Volume 2C—Section 6.3, the exposure transformation uses two smoothing, differentiating filters to compute the exposure of each 'pixel' to a selected direction. The 3×3 smoothing derivative filters, with divisor 12 and offset 127, are defined as:

$$\frac{\partial}{\partial X} = \begin{bmatrix} -1 & 0 & 1 \\ -4 & 0 & 4 \\ -1 & 0 & 1 \end{bmatrix}$$

$$\frac{\partial}{\partial Y} = \begin{bmatrix} 1 & 4 & 1 \\ 0 & 0 & 0 \\ -1 & -4 & -1 \end{bmatrix}$$

These are the product of a B-spline filter:

$$\begin{bmatrix} 1 & 4 & 1 \end{bmatrix}$$

and a central difference filter:

$$\begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$$

### A5.2.2 Relief shading

To produce an relief shading filter, two differential filters are used to compute the first derivatives in two directions: along-line (X) and down pixel columns (Y) (see Volume 2C). These filters are the product of a B-spline filter:

$$\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \end{bmatrix}$$

(divisor 16), with a central difference filter:

$$\begin{bmatrix} -1 & -2 & 0 & 2 & 1 \end{bmatrix}$$

(divisor 8).

The 5×5 average filter shown in Figure A5.1b may be applied to the elevation channel before using differential filters.

$\frac{\partial}{\partial X}$  filter with divisor 128:

-1	-4	-6	-4	-1
-2	-8	-12	-8	-2
0	0	0	0	0
2	8	12	8	2
1	4	6	4	1

$\frac{\partial}{\partial Y}$  filter with divisor 128:

1	2	0	-2	-1
4	8	0	-8	-4
6	12	0	-12	-6
4	8	0	-8	-4
1	2	0	-2	-1

### A5.2.3 Curvature

Three differential filters are used in the microBRIAN Curvature transformation to compute the second derivatives in three directions: along-line (X) and down pixel columns (Y) and along the diagonal (XY) (see Section 12.9). The X and Y filters are the product of a B-spline filter (divisor 96):

1	8	23	32	23	8	1
---	---	----	----	----	---	---

with a second derivative filter (divisor 4):

1	2	-1	-4	-1	2	1
---	---	----	----	----	---	---

The XY filter is formed as the outer product of the first derivative filter (divisor 4):

-1	-4	-5	0	5	4	1
----	----	----	---	---	---	---

with its transpose.

$\frac{\partial^2}{\partial X^2}$  filter with divisor 1536:

1	8	23	32	23	8	1
2	16	46	64	46	16	2
-1	-8	-23	-32	-23	-8	-1
-4	-32	-92	-128	-92	-32	-4
-1	-8	-23	-32	-23	-8	-1
2	16	46	64	46	16	2
1	8	23	32	23	8	1

$\frac{\partial^2}{\partial Y^2}$  filter with divisor 1536:

1	2	-1	-4	1	2	1
8	16	-8	-32	46	16	8
23	46	-23	-92	-23	46	23
32	64	-32	-128	-92	64	32
23	46	-23	-92	-23	46	23
8	16	-8	-32	46	16	8
1	2	-1	-4	23	2	1

$\frac{\partial^2}{\partial X \partial Y}$  filter with divisor 1024:

-1	-4	-5	0	5	4	1
-4	-16	-20	0	20	16	4
-5	-20	-25	0	25	20	5
0	0	0	0	0	0	0
5	20	25	0	-25	-20	-5
4	16	20	0	-20	-16	-4
1	4	5	0	-5	-4	-1

The Laplacian can be computed in the curvature transformation as the average of the second derivatives in X and Y. A Laplacian filter that is equivalent to this

operation can be derived from the  $\frac{\partial^2}{\partial X^2}$  and  $\frac{\partial^2}{\partial Y^2}$  filters given above:

1	5	11	14	11	5	1
5	16	19	16	19	16	5
11	19	-23	-62	-23	19	11
14	16	-62	-128	-62	16	14
11	19	-23	-62	-23	19	11
5	16	19	16	19	16	5
1	5	11	14	11	5	1

### A5.3 Geophysical Filters

---

For geophysical applications, Fuller (1966) reports a number of derivative and other filters (including second and fourth vertical derivations, residual and analytic continuation), which implement useful frequency filters in the spatial domain. Several derivatives (or 'residuals') of geomagnetic fields were proposed by Henderson and Zeitz (1949) to highlight magnetic anomalies. Their equations 10, 13 and 15 were represented as spatial filters by Fuller (1966).

Peters (1949) computed second and fourth derivatives of magnetic data. Second derivative equations for interpreting gravity data were reported by Elkins (1951). Rosenbach (1953) also proposed equations for computing the second vertical derivative from gravity data.

### A5.4 References

---

- Elkins, T.A. (1951). The second derivative method of gravity interpretation. *Geophysics*, 16, 29–50.
- Frei, W., and Chen, C.C. (1977). Fast boundary detection: A generalisation and a new algorithm. *IEEE Transactions on Computers* C-26, 988–98.
- Fuller, B.D. (1966). Two-dimensional frequency analysis and design of Grid Operators. In *Society of Exploration Geophysicists Mining Geophysics. Vol II. Theory*. (Ed: D.A. Hansen), p. 658–708.
- Henderson, R.G., and Zeitz, I. (1949). The computation of second vertical derivatives of geomagnetic fields. *Geophysics* 14, 508–16.
- Jupp, D.L. (1976). B-Splines for smoothing and differentiating data sequences. *Mathematical Geology*, 8, 243–66.
- Kirsch, R.A. (1971). Computer determination of the constituent structure of biological images. *Computers and Biomedical Research* 4, 315–28.
- Peters, L.J. (1949). The direct approach to magnetic interpretation and its practical application. *Geophysics* 14, 290–319.
- Prewitt, J.M.S. (1970). *Object Enhancement and Extraction*. In *Picture Processing and Psychopictorics*. (Eds: B.S. Lipkin and A. Rosenfeld.) Academic Press, New York. pp 75–149.
- Rosenbach, O. (1953). A contribution to the computation of the 'second derivative' from gravity data. *Geophysics* 18, 894–909.

# Appendix 6

## Matrix Algebra

An excellent explanation of various mathematical analyses, including matrix algebra, is given in Sawyer (1955). Unlike most mathematical texts, this unique account was specifically written to interest people in mathematics and exposes various branches of the subject using lucid, commonplace examples and only assumes ‘a confused recollection of School Certificate mathematics’.

Using material from this source, Appendix A6.1 describes the operation of matrices in terms of their geometric effects on image spaces. Some

characteristics of matrix algebra which may be used for representing specific geometrical transformations are detailed in Appendices A6.2 and A6.3.

### A6.1 Geometric Effect of Matrix Operations

The geometrical effect of the general matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

on the image data space can be understood using example points along the X and Y axes. In Figure A6.1, we use three equally spaced points on each axis: P, Q and R on the X axis and U, V and W on the Y axis. The matrix operation then creates two new axes X', where the three X axis points have been moved to P', Q' and R', and Y', where the Y axis points are now located at U', V' and W'. Because this is a linear operation, straight lines remain straight and parallel lines remain parallel. For geometrically representable operations, the matrix coefficients can be simply determined as the transformed values of the points P (that is a,c at P') and U (namely b,d at U'), that is:

$$\begin{bmatrix} X' \\ Y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} X \\ Y \end{bmatrix}$$

For example in Figure A6.1 the original coordinates of P (1,0) are transformed to the values (a,c) at P'.

Image data axes may be reflected about any line L by the matrix coefficients:

$$\begin{bmatrix} \cos 2a & \sin 2a \\ \sin 2a & -\cos 2a \end{bmatrix}$$

where a is the angle (in degrees) between L and the X axis (see Figure A6.2). For:

$$Y=X \\ a=45^\circ$$

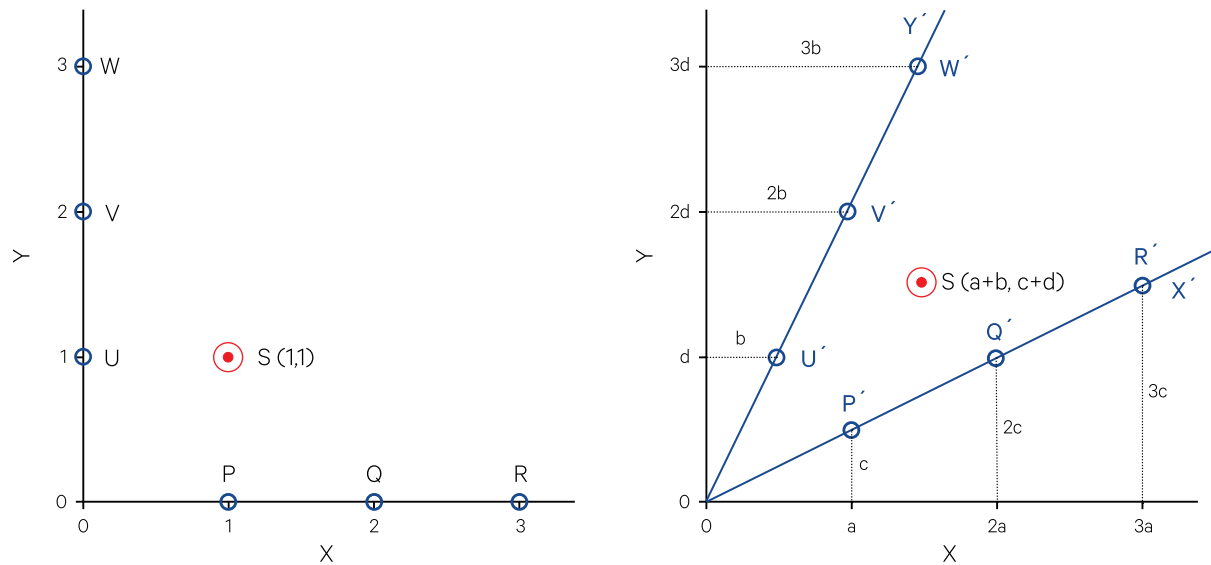
then

$$2a=90^\circ \\ \sin 2a=1 \\ \cos 2a=0$$

to produce the coefficients used in the example of Figure A6.3.

**Figure A6.1** Effect of matrix operation on data geometry

The transformation from  $X, Y$  to  $X', Y'$  can be described by the matrix equations:  $X' = aX + bY$  and  $Y' = cX + dY$ .

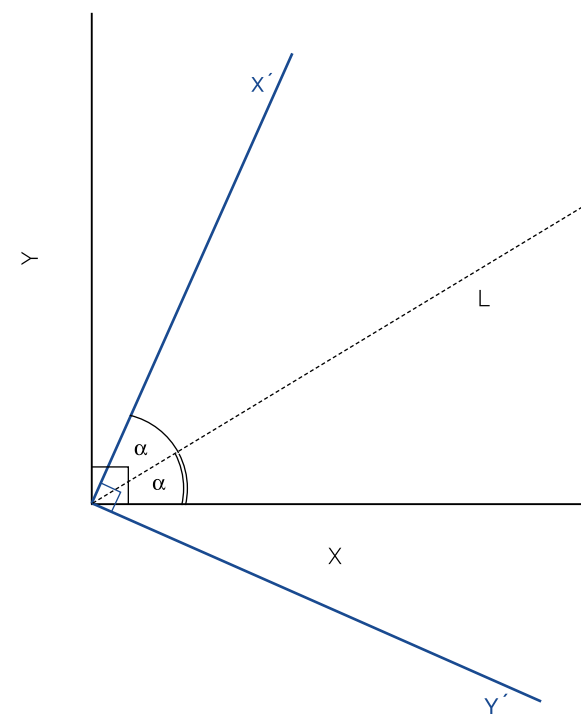


Source: Harrison and Jupp (1990) Figure 108

**Figure A6.2** Reflecting axes

$$X' = \cos 2\alpha X + \sin 2\alpha Y$$

$$Y' = \sin 2\alpha X - \cos 2\alpha Y$$



Source: Harrison and Jupp (1990) Figure 109

Linear rescaling of image values (see Volume 2A and Figure A6.4) can be simply implemented by multiplying  $X$  and/or  $Y$  values by some constant value(s). This can be represented by the matrix:

$$\begin{bmatrix} m & 0 \\ 0 & n \end{bmatrix}$$

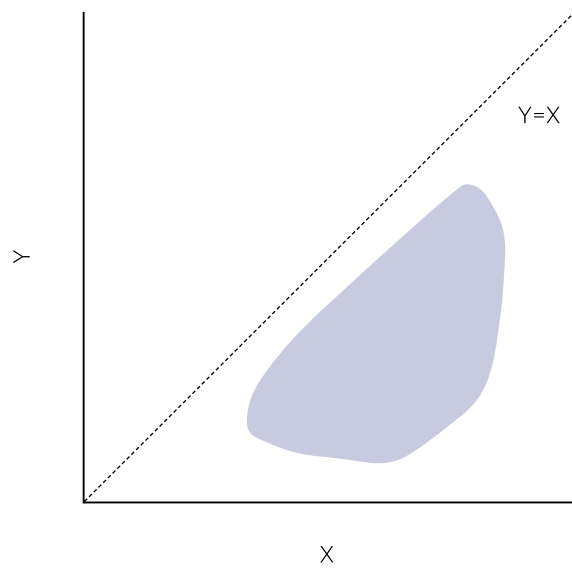
where, if  $m = n$ , then both  $X$  and  $Y$  channels are being rescaled by the same amount. A special case of channel rescaling involves inverting the channel values (that is, changing a positive increasing order to a negative decreasing order). This can be effected by using  $m$  and  $n$  values equal to  $-1$ . In image processing, an offset value would usually need to be applied to this result to preserve positive pixel values.

*By three methods we may learn wisdom:  
First, by reflection, which is noblest;  
Second, by imitation, which is easiest; and  
Third by experience, which is the bitterest.  
(Confucius)*

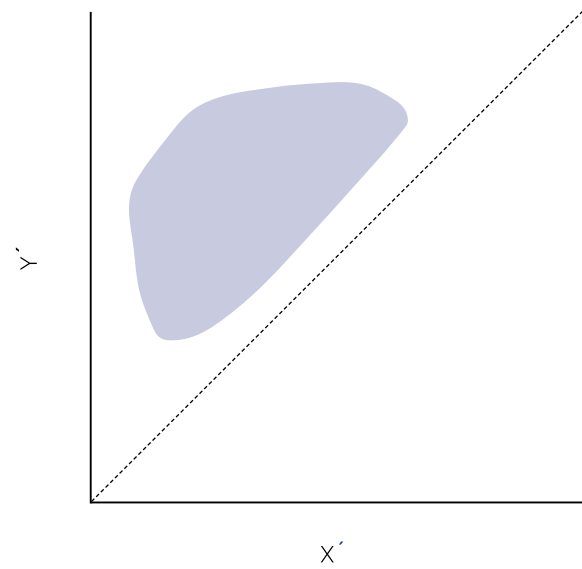


**Figure A6.3** Reflecting axes about the line  $Y=X$ 

a. Original axes



b. Reflected axes

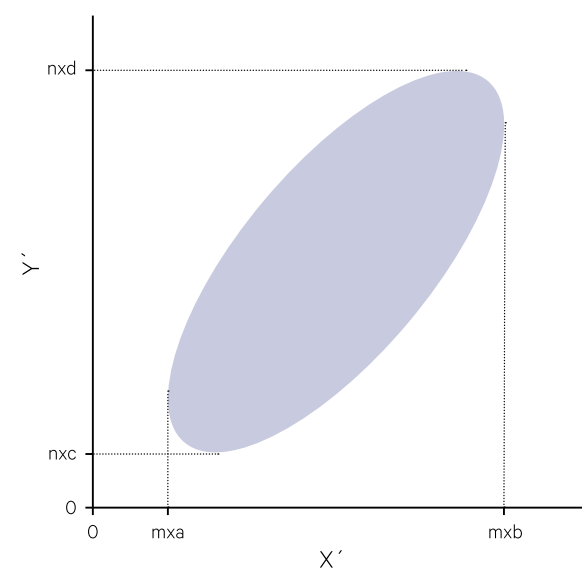
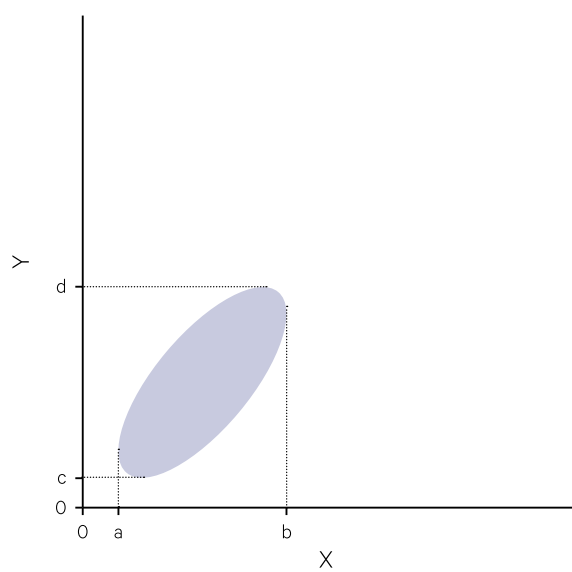


Source: Harrison and Jupp (1990) Figure 85

**Figure A6.4** Rescaling axes

$$X' = mX$$

$$Y' = nY$$



Source: Harrison and Jupp (1990) Figure 110

Rotation of the image data space (actually a combination of skewing and scaling) can be represented by the matrix:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

where  $\theta$  is defined as the clockwise angle between  $X$  and  $X'$  (see Figure A6.5). Some special cases of this operation are:

- $\theta=90^\circ$ ,  $\cos\theta=0$  and  $\sin\theta=1$ , which produces a matrix of reflection about  $Y=X'$  with inverted scaling in  $X'$ ; and
- $\theta=180^\circ$ ,  $\cos\theta=-1$  and  $\sin\theta=0$ , so the rotation matrix simply inverts scaling on both the  $X'$  and  $Y'$  axes.

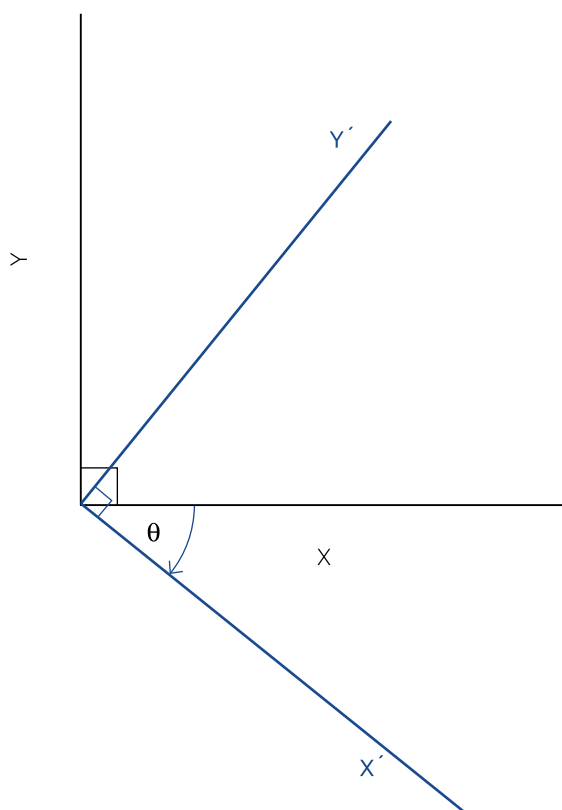
A skewing operation in the  $X$  direction as illustrated in Figure A6.6 can be represented as the matrix:

$$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$$

**Figure A6.5** Rotating axes

$$X' = \cos\theta X - \sin\theta Y$$

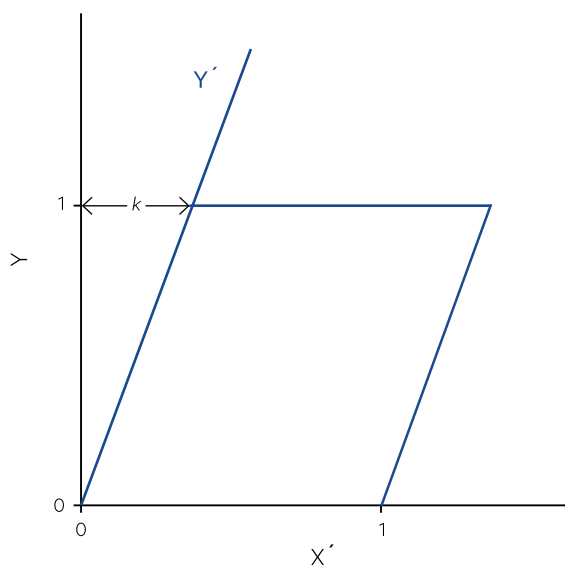
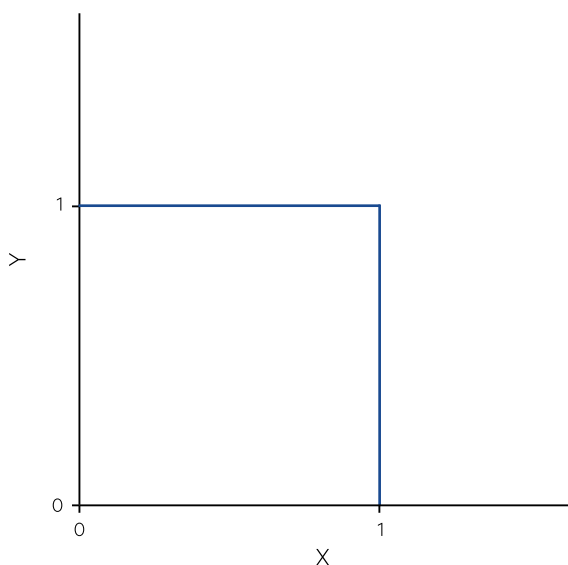
$$Y' = \sin\theta X + \cos\theta Y$$



Source: Harrison and Jupp (1990) Figure 111

**Figure A6.6** Skewing axes

$$X' = X + kY$$



Source: Harrison and Jupp (1990) Figure 112

## A6.2 Matrix Arithmetic and Properties

A Zero matrix (all elements equal to 0) is defined as:

$$\begin{bmatrix} 0 \end{bmatrix}$$

For a 2×2 matrix, this would be:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

A Unit or Identity matrix (all diagonal elements are 1 and off-diagonals equal 0) is represented as:

$$\begin{bmatrix} 1 \end{bmatrix}$$

For a 2×2 matrix, this would be:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

For two 2×2 matrices:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$B = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$$

### A6.2.1 Trace

The Trace of a matrix equals the sum of its diagonal elements:

$$\text{Trace}(A) = a + d$$

### A6.2.2 Matrix equality

$$A = B$$

only if:

$$a = p; \text{ and}$$

$$b = q; \text{ and}$$

$$c = r; \text{ and}$$

$$d = s.$$

### A6.2.3 Matrix addition

$$A + B = \begin{bmatrix} a+p & b+q \\ c+r & d+s \end{bmatrix} = B + A$$

### A6.2.4 Scalar multiplication

For a constant  $n$ :

$$n \times A = \begin{bmatrix} n \times a & n \times b \\ n \times c & n \times d \end{bmatrix} = \begin{bmatrix} n & 0 \\ 0 & n \end{bmatrix} \times \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

### A6.2.5 Matrix multiplication

$$A \times B = \begin{bmatrix} a \times p + b \times r & a \times q + b \times s \\ c \times p + d \times r & c \times q + d \times s \end{bmatrix}$$

$$B \times A = \begin{bmatrix} p \times a + q \times c & p \times b + q \times d \\ r \times a + s \times c & r \times b + s \times d \end{bmatrix}$$

Note: matrix multiplication is not commutative:  $AB$  is not equal to  $BA$ .

### A6.2.6 Transpose of a matrix

$$A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

### A6.2.7 Additive inverse of a matrix

$$A + (-A) = (-A) + A = 0$$

$$-A = \begin{bmatrix} -a & -b \\ -c & -d \end{bmatrix}$$

### A6.2.8 Multiplicative inverse of a matrix

$$A \times A^{-1} = A^{-1} \times A = 1$$

If  $(ad-bc)$  is not equal to 0, then  $A$  is called an orthogonal matrix and

$$A^{-1} = \begin{bmatrix} \frac{d}{a \times d - b \times c} & \frac{-b}{a \times d - b \times c} \\ \frac{-c}{a \times d - b \times c} & \frac{a}{a \times d - b \times c} \end{bmatrix}$$

$$= \frac{d}{a \times d - b \times c} \times \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

If  $ad-bc$  equals 0,  $A$  is called a singular matrix and  $A^{-1}$  does not exist.

$$(ad-bc)$$

is called the determinant of  $A$ , and

$$\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

is called the adjoint of  $A$ .

## A6.2.9 Determinant of a matrix

In Figure A6.1, we applied a matrix to a set of points located on the X and Y axes. The effect of moving points P, U and S changed the shape of a unit square to that of a parallelogram. The dimensions of this parallelogram are described in Figure A6.7. The area of this parallelogram equals the determinant of **A** and can be computed as:

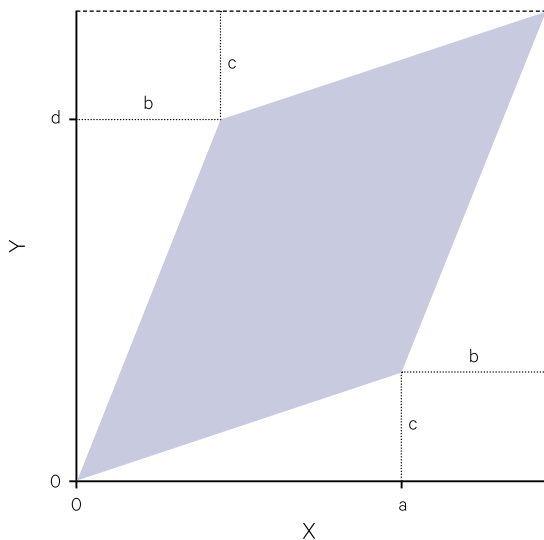
$$(a+b) \times (c+d) - 2(b \times c) - (a \times c) - (b \times d) \\ = a \times d - b \times c$$

Since this Figure originated from a unit square (area 1), the determinant represents the change in area caused by the matrix operation. The determinant is usually written as:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

**Figure A6.7** Effect of matrix operation on unit area

The area of a parallelogram resulting from a matrix operation is described by the determinant of the matrix. In this case the determinant equals  $ad-bc$ . For a  $2 \times 2$  matrix this is defined as the ratio in which the matrix changes areas (see Figure A6.1).



Source: Harrison and Jupp (1990) Figure 113

For a  $2 \times 2$  matrix, the is defined as the ratio in which the matrix changes area. The determinant of a  $3 \times 3$  matrix is defined as:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & k \end{vmatrix} = a \begin{vmatrix} e & f \\ h & k \end{vmatrix} - b \begin{vmatrix} d & f \\ g & k \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ = a \times e \times k + b \times f \times g + c \times d \times h \\ - a \times h \times f - b \times d \times k - c \times e \times g$$

and represents the constant ratio in which the matrix changes volume in a three-dimensional data space. The concept can be used with higher-dimensional dataset although the geometric process cannot be visualised.

The determinant is a useful summary of many matrix characteristics such as:

$$\begin{vmatrix} c & d \\ a & b \end{vmatrix} = - \begin{vmatrix} a & b \\ c & d \end{vmatrix} \\ \begin{vmatrix} a+k \times c & b+k \times d \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \\ \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} \\ \begin{vmatrix} a & b \\ c & d \end{vmatrix} \times \begin{vmatrix} p & q \\ r & s \end{vmatrix} = \begin{vmatrix} a \times p + b \times r & a \times q + b \times s \\ c \times p + d \times r & c \times q + d \times s \end{vmatrix}$$

These operations can be used to determine the overall data scaling action of a sequence of matrix transformations on a dataset. The reflection, rotation and skewing operations described above do not alter the area of the dataset so have determinants equal to one in magnitude. The reflection transformations, however, all have negative determinants which characteristically indicates that there is a reversal of axes involved with such matrices. For example, the skewing operation can be represented as:

$$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+k \times c & b+k \times d \\ c & d \end{bmatrix}$$

and

$$\begin{vmatrix} a+k \times c & b+k \times d \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

or

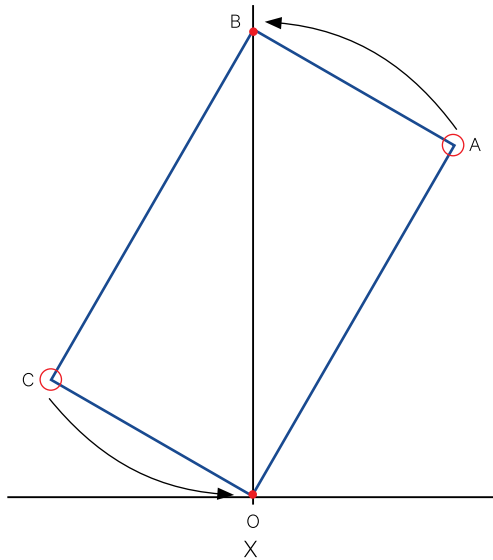
$$\begin{vmatrix} 1 & k \\ 0 & 1 \end{vmatrix} = 1 \times 1 - k \times 0 = 1$$

so does not change area.

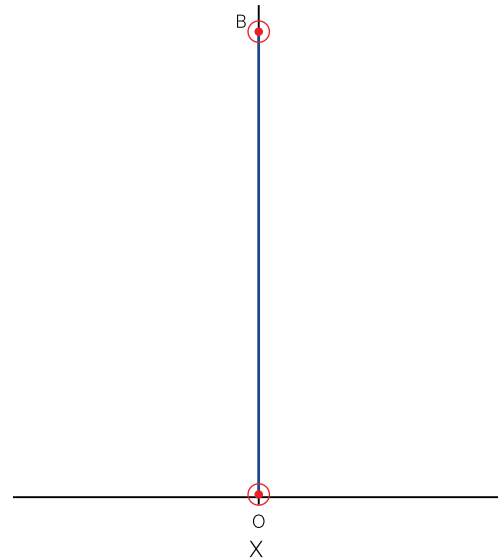
**Figure A6.8** Effect of singular matrix operation

A matrix whose determinant equals zero represents an operation which collapses a two-dimensional area to one dimension (see Figure A6.7). In this example, point C is collapsed onto O and point A is collapsed onto point B.

a. Original area



b. Two dimensional area collapsed to one dimensional line



Source: Harrison and Jupp (1990) Figure 114

Finally a special property of some matrices is that their determinant value is zero. Such matrices are called singular and operate by collapsing areas in the data space into lines (of zero area) as illustrated in Figure A6.8. In image processing, this matrix operation loses a dimension so effectively 'destroys' data.

#### A6.2.10 Characteristic equation

The Characteristic Equation of any square matrix  $A$  is:

$$|A - k \times 1| = 0$$

or

$$\left| \begin{bmatrix} a & b \\ c & d \end{bmatrix} - \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix} \right| = 0$$

or

$$\begin{vmatrix} a-k & b \\ c & d-k \end{vmatrix} = 0$$

Thus

$$(a-k) \times (d-k) - b \times c = 0$$

#### A6.2.11 Eigenvalues of a matrix

The eigenvalues of a matrix are defined as the solutions of its characteristic equation. The eigenvalues  $k$  of matrix  $A$  may be written as:

$$A \times \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = k \times \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$$

where

$r_1$  and  $r_2$  are points representing a vector (called the eigenvector) from the origin of the original coordinate system, then

$$A \times \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} - k \times 1 \times \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = 0$$

or

$$(A - k \times 1) \times \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = 0$$

that is

$$\begin{bmatrix} a-k & b \\ c & d-k \end{bmatrix} \times \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Solutions of this equation are:

$$(a-k) \times r_1 + b \times r_2 = 0$$

and

$$c \times r_1 + (d-k) \times r_2 = 0$$

### A6.2.12 Eigenvectors of a matrix

The eigenvectors  $r$  of matrix  $A$  are defined by:

$$(A - k_1 \times 1) \times r = 0$$

and

$$(A - k_2 \times 1) \times r = 0$$

where

$k_1$  and  $k_2$  are the roots of the characteristic equation (also called the eigenvalues).

## A6.3 Derivation of the Principal Components Transformation

In an image  $X$ , the covariance of channels  $j$  and  $k$  is defined as:

$$V_{ij} = \frac{\sum_{k=1}^n (x_{jk} - \bar{x}_j) \times (x_{ik} - \bar{x}_i)}{n}$$

where

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

see Volume 2A—Section 8.1.4). The covariance matrix,  $C_x$ , can be defined as:

$$\frac{1}{n} \times \sum_{i=1, n} (x_i - \bar{x}) \times (x_i - \bar{x})^T$$

or

$$E \left\{ (x - \bar{x}) \times (x - \bar{x})^T \right\}$$

As defined in Volume 2C, a PCA (Principal Components Analysis) image has a diagonal covariance matrix, that is all off-diagonal elements are zero. The PCA image  $Y$  can be computed from the image  $X$  by a rotation transformation  $P$  where:

$$y = P \times (x_i - \bar{x})$$

The covariance matrix of a PCA image  $Y$  can be similarly defined as:

$$C_y = E \left\{ (y - \bar{y}) \times (y - \bar{y})^T \right\}$$

By substituting the equation relating  $Y$  and  $X$  we get:

$$\begin{aligned} C_y &= E \left\{ (P \times x - P \times \bar{x}) \times (P \times x - P \times \bar{x})^T \right\} \\ &= P \times E \left\{ (x - \bar{x}) \times (x - \bar{x})^T \right\} \times P^T \\ &= P \times C_x \times P^T \end{aligned}$$

Since  $C_y$  is diagonal, its diagonal elements can be defined as the eigenvalues of  $C_x$  and the rows in the orthogonal matrix  $P$  are the eigenvectors of  $C_x$  (Gonzalez and Wintz, 1977). The trace of  $C_x$  also equals the trace of  $C_y$ .

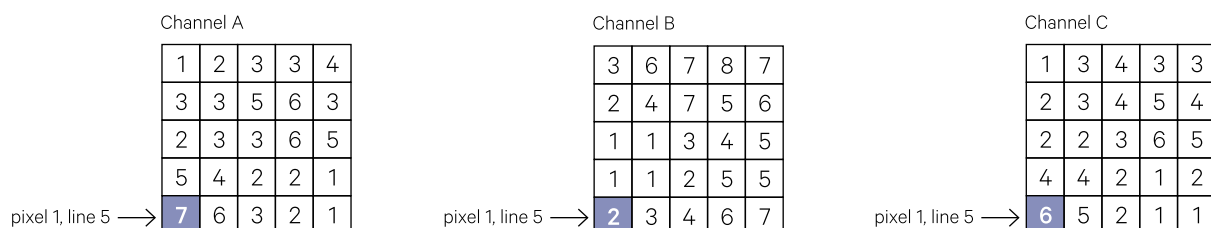
The eigenvalues  $k$  can be computed as defined in Section A6.2.10 above, that is:

$$|C_x - k \times 1| = 0$$

For the example image from Volume 2A (see Figure A6.9), the covariance matrix is:

$$C_x = \begin{bmatrix} 2.9380 & -0.3840 & 2.3000 \\ -0.3840 & 4.7200 & -0.0144 \\ 2.3000 & -0.0144 & 2.2656 \end{bmatrix}$$

Figure A6.9 Example image channels



Source: Harrison and Jupp (1990) Figures 41



The eigenvalues of the matrix are derived as:

$$\begin{vmatrix} 2.938-k & -0.384 & 2.3 \\ -0.384 & 4.72-k & -0.0144 \\ 2.3 & -0.0144 & 2.2656-k \end{vmatrix} = 0$$

or

$$\begin{aligned} & (2.938-k) \times (4.72-k) \times (2.2656-k) \\ & + (-0.384 \times -0.0144 \times 2.3) + (2.3 \times -0.384 \times -0.0144) \\ & - ((2.938-k) \times -0.0144^2) - (-0.384^2 \times (2.2656-k)) \\ & - (2.3 \times (4.72-k) \times 2.3) \end{aligned}$$

which gives eigenvalues of 5.1, 4.5 and 0.3. The covariance matrix of  $Y$  can then be written as:

$$\begin{bmatrix} 5.1 & 0 & 0 \\ 0 & 4.5 & 0 \\ 0 & 0 & 0.3 \end{bmatrix}$$

The trace of this matrix (5.1+4.5+0.3) equals the trace of  $C_x$  (2.938+4.72+ 2.2656). The percentage variation accounted for by each principal component (PC) can be determined using the equation:

$$\frac{\sigma^2 \times 100}{\text{trace}}$$

The variance and percent variation results are shown in Table A6.1.

A noise-to-signal ratio (NSR) percentage summarises the variance of a component  $j$  relative to those components above it:

$$NSR_j \% = 100 \times \sqrt{\frac{\sum_{i=j+1,n} \left( \frac{\sigma_i^2}{n-j} \right)}{\sum_{i=1,j} \left( \frac{\sigma_i^2}{j} \right)}}$$

where

$n$  is the number of components;  
 $j$  is the selected component for NSR; and  
 $i$  is the component number.

**Table A6.1** Principal component percentage variation

PC	Variance	Percentage
1	5.1	51.8
2	4.5	45.5
3	0.3	2.7
Total	9.9	100

This can be used to identify the number of data dimensions, or rank, of the image matrix. Typically useful dimensions have a high variance value and ‘noise’ components have low variance so the number of data dimensions can be determined from the component with the minimum noise to signal ratio value. (This ratio is generally observed as a trend of decreasing values through the data components then marginally increasing in the noise components.) For example, in an eight channel image we have eight components  $a, b, c, d, e, f, g, h$  in which the variances of ( $a, b, c, d, e$ ) are much greater than the variances of components  $f, g$  and  $h$ . The NSR % at component  $e$  can be computed as:

$$NSR_e \% = 100 \times \sqrt{\frac{(\sigma_f^2 + \sigma_g^2 + \sigma_h^2) / 3}{(\sigma_a^2 + \sigma_b^2 + \sigma_c^2 + \sigma_d^2 + \sigma_e^2) / 5}}$$

Since the variance of  $f$  is much smaller than that of  $e$ , this percentage will be larger than the NSR % at component  $f$ :

$$NSR_e \% = 100 \times \sqrt{\frac{(\sigma_g^2 + \sigma_h^2) / 2}{(\sigma_a^2 + \sigma_b^2 + \sigma_c^2 + \sigma_d^2 + \sigma_e^2 + \sigma_f^2) / 6}}$$

If the variances of  $f, g$  and  $h$  are all small relative to components  $a$  to  $e$ , the NSR % at  $g$  and  $h$  will be marginally larger than  $f$  since a similarly small numerator is being divided by a slightly smaller denominator (the additional noise channel reduces the arithmetic mean).

The eigenvectors of  $C_x$  can also be computed to form the rows of the transformation matrix  $P$ :

$$\begin{bmatrix} p_{ii} & p_{ij} & p_{ik} \\ p_{ji} & p_{jj} & p_{jk} \\ p_{ki} & p_{kj} & p_{kk} \end{bmatrix}$$

using the equations:

$$\begin{aligned} (C_x - k_1 \times 1) \times p_{.i} &= 0 \\ (C_x - k_2 \times 1) \times p_{.j} &= 0 \\ (C_x - k_3 \times 1) \times p_{.k} &= 0 \end{aligned}$$

For  $k_1$  this expands to the three simultaneous equations:

$$\begin{aligned} (2.938 - 5.1) \times p_{ii} - 0.384 \times p_{ji} + 2.3 \times p_{ki} &= 0 \\ 0.384 \times p_{ii} (4.72 - 5.1) \times p_{ji} - 0.0144 \times p_{ki} &= 0 \\ 2.3 \times p_{ii} - 0.0144 \times p_{ji} + (2.2656 - 5.1) \times p_{ki} &= 0 \end{aligned}$$

Similarly, two more sets of three equations can be generated by substituting the values of  $k_2$  and  $k_3$ . Since matrix  $\mathbf{A}$  must be orthogonal ( $\mathbf{A}^T = \mathbf{A}^{-1}$ ), the eigenvectors are also normalised such that:

$$\begin{aligned} p_{ii}^2 + p_{ji}^2 + p_{ki}^2 &= 1 \\ p_{ij}^2 + p_{jj}^2 + p_{kj}^2 &= 1 \\ p_{ik}^2 + p_{jk}^2 + p_{kk}^2 &= 1 \end{aligned}$$

For our example, the final transformation matrix is computed as:

$$\begin{bmatrix} 0.6306 & -0.5879 & 0.5068 \\ 0.4153 & 0.8072 & 0.4195 \\ 0.6556 & 0.0541 & -0.7531 \end{bmatrix}$$

As detailed in Volume 2C, this matrix can now be used to rotate the data space of image  $X$  to the uncorrelated PCA image space. This is implemented by the equations:

$$\begin{aligned} \text{PC1} &= 0.6306 \times \text{ch a} - 0.5879 \times \text{ch b} + 0.5068 \times \text{ch c} \\ \text{PC2} &= 0.4153 \times \text{ch a} + 0.8072 \times \text{ch b} + 0.4195 \times \text{ch c} \\ \text{PC3} &= 0.6556 \times \text{ch a} + 0.0541 \times \text{ch b} - 0.7531 \times \text{ch c} \end{aligned}$$

The determinant of this matrix can be computed using the equation given in Appendix A6.2.9:

$$\begin{aligned} & (0.6306 \times 0.8072 \times -0.7531) + (-0.5879 \times 0.4195 \times 0.6556) \\ & + (0.5068 \times 0.4153 \times 0.0541) - (0.6306 \times 0.0541 \times 0.4195) \\ & + (-0.5879 \times 0.4153 \times 0.7531) - (0.5068 \times 0.8072 \times 0.6556) \\ & = -1 \end{aligned}$$

which indicates that there is a reversal of axes involved with the transformation but no change in the 'area' occupied by the data.

The degree of correlation of each channel  $i$  with each principal component  $j$  can also be determined using the equation (Short, 1982):

$$p_{ij} \times \frac{\sqrt{k_j}}{\sigma_{ii}}$$

to give the normalised matrix:

$$\begin{bmatrix} 0.8307 & -0.6109 & 0.7604 \\ 0.5136 & 0.7875 & 0.5909 \\ 0.2096 & 0.0197 & -0.2742 \end{bmatrix}$$

The elements of this matrix indicate the degree of correlation between the original channels and each PC. For example, channels  $a$  and  $c$  are strongly correlated with PC1 (0.8307 and 0.7604 respectively) while channel  $b$  is negatively correlated with PC1 (-0.6109) and more strongly correlated with PC2 (0.7875).

## A6.4 References

- Gonzalez, R.C., and Wintz, P. (1977). *Digital Image Processing*. Addison-Wesley Pub. Co. Reading, Mass.
- Harrison, B.A., and Jupp, D.L.B. (1990). *Introduction to Image Processing. Part TWO of the microBRIAN Resource Manual*. CSIRO, Melbourne. 256pp.
- Sawyer, W.W. (1955). *Prelude to Mathematics*. Penguin Books Ltd. London.
- Short, N.M. (1982). *The Landsat Tutorial Workbook. Basics of Satellite Remote Sensing*. NASA Reference Publication 1078.

# Appendix 7

## Predictive Error and Variance

### A7.1 Predictive Error for Least Squares

The predictive error (or Predictive Residual Error Sum of Squares, PRESS) described below is the one defined by Allen (1974) and is a variation on the 'Jackknife' statistic discussed in Tuckey (1967).

The Least Squares (LS) situation for fitting transformation models by LS can be written by defining the error at Ground Control Point (GCP) 'i' as:

$$e_i = \varphi_i^T \alpha - z_i$$

where  $\varphi_i$  is the vector of 'monomials' evaluated at the  $i$ 'th GCP ( $x_i, y_i$ ) in the 'from' coordinate system,  $\alpha$  is the vector of coefficients and  $z_i$  is the coordinate ( $x$  or  $y$ ) being predicted in the 'to' coordinate system (see Volume 2B).

If there are  $M$  GCPs then the complete expression for the residuals at the GCP points is:

$$\underline{e} = \underline{z} - A\alpha$$

The LS solution vector  $\alpha^*$  minimises the sum of squared residuals or vector 'norm':

$$\text{Min}_{\alpha} \|\underline{e}\|^2 = \underline{e}^T \underline{e} = \sum_{i=1}^M e_i^2$$

The solution is well known and can be written in this notation as:

$$\begin{aligned} \alpha^* &= (A^T A)^{-1} A^T \underline{z} \\ \underline{e}^* &= \underline{z} - A\alpha^* \\ &= \left[ I - A(A^T A)^{-1} A^T \right] \underline{z} \end{aligned}$$

The consequent estimate for the error called the Residual Mean Square error (RMS) error is:

$$\begin{aligned} RMS &= \frac{1}{M} \sum_{i=1}^M |e_i^*|^2 \\ &= \frac{1}{M} RSS \end{aligned}$$

where RSS is termed the Residual Sum of Squares, which is not normalised by the number of observations  $M$ .

For reasons that will become clearer in the following we will denote the matrix:

$$\begin{aligned} A(A^T A)^{-1} A^T &= H \\ &= [h_{ij}] \\ h_{ij} &= \varphi_i^T (A^T A)^{-1} \varphi_j \end{aligned}$$

The (symmetric) matrix  $H$  is a special matrix in that powers of  $H$  simply result in  $H$  again as is the case with  $(I-H)$ . It follows that the RSS can be written:

$$\begin{aligned} RSS &= \underline{z}^T (I - H) \underline{z} \\ &= \underline{z}^T \underline{z} - \underline{z}^T H \underline{z} \\ &= TSS - MSS \end{aligned}$$

which simply expresses the 'analysis of variance' associated with LS of the total Sum of Squares (TSS) into the sum of Model Sum of Squares (MSS) and Residual sum of Squares (RSS).

The 'Predictive' error is developed by noting that the errors at the GCPs after the model is fitted are too small as estimates for the actual statistical error. This comes about as they are used to fit the model. The consequent bias in the RMS is more than the normal allowance for the model degrees of freedom to obtain the estimated variance (EVAR) as in:

$$EVAR = \frac{1}{M-p} \sum_{i=1}^M |e_i^*|^2$$

(where  $p$  is the order of the model) will allow.

As an alternative and unbiased estimate of the error, if each GCP is removed from the model fit in turn and the error between the point removed and its prediction by the fitted model without its influence is used in place of the LS error at that point then the Predictive Error (PRESS) is obtained.

The method is really a variation of retaining some points as ‘test points’ and some as ‘training points’. However, the existence of the test points is always a problem in such a scheme in that people wish to use all of the information they have to train and to test. The Predictive Error is a good compromise.

Let us write for the predictive error at the  $i$ 'th GCP:

$$e_i^{(i)} = z_i - \phi_i^T \alpha^{(i)}$$

where  $\alpha^{(i)}$  is the set of model parameters obtained if the point  $z_i$  is not used to fit the model. Then:

$$\text{PRESS} = \frac{1}{M} \sum_{i=1}^M |e_i^{(i)}|^2$$

It may seem that this quantity involves  $M$  LS solutions and is quite messy. However, it turns out that the result is immediately available following the LS solution since:

$$\text{PRESS} = \frac{1}{M} \sum_{i=1}^M w_i |e_i^*|^2$$

$$w_i = \frac{1}{(1 - h_{ii})^2}$$

where  $e_i^*$  is the full LS residual obtained above and  $h_{ii}$  is the  $i$ 'th diagonal element of the matrix  $H$  above.

The proof of this result can be found in Golub *et al.* (1979) but is easily and usefully derived as follows:

- noting that we can write the components of the LS solution as:

$$A^T A = \sum_{i=1}^M \phi_i \phi_i^T$$

$$A^T z = \sum_{i=1}^M z_i \phi_i$$

- it follows that we can express the  $i$ 'th component of the PRESS as:

$$e_i^{(i)} = z_i - \phi_i^T \alpha^{(i)}$$

where:

$$\alpha^{(i)} = (A^T A - \phi_i \phi_i^T)^{-1} (A^T z - z_i \phi_i)$$

- with the help of the well-known Sherman-Morrison formula (Sherman and Morrison, 1949) for (in this case) a non-singular and symmetric matrix,  $H$ :

$$(H + \lambda a a^T)^{-1} = H^{-1} - \frac{\lambda (H^{-1} a)(H^{-1} a)^T}{1 + \lambda a^T H^{-1} a}$$

- it follows with a little manipulation that:

$$\phi_i^T \alpha^{(i)} = \left( 1 + \frac{h_{ii}}{1 - h_{ii}} \right) (\phi_i^T \alpha^* - h_{ii} z_i)$$

$$= \frac{\phi_i^T \alpha^* - h_{ii} z_i}{1 - h_{ii}}$$

so that

$$e_i^{(i)} = z_i - \phi_i^T \alpha^{(i)}$$

$$= z_i - \frac{\phi_i^T \alpha^* - h_{ii} z_i}{1 - h_{ii}}$$

$$= \frac{z_i - \phi_i^T \alpha^*}{1 - h_{ii}}$$

$$= \frac{1}{1 - h_{ii}} e_i^*$$

QED

## A7.2 Generalised Cross-Validation (GCV)

Golub *et al.* (1979) favoured the Generalised Cross-Validation or GCV statistic over PRESS because they say that PRESS is not ‘rotation invariant’. Perhaps this is not a vital characteristic of a statistic to have as I believe PRESS is an excellent tool that is sensitive to the distribution of the GCPs. This is the main reason for its use.

However, for completeness and because GCV is used to resolve ‘ties’ in the modelling described here, the GCV needs to be derived as well.

We have seen above that PRESS can be written as:

$$\text{PRESS} = \frac{1}{M} \sum_{i=1}^M w_i |e_i^*|^2$$

$$w_i = \frac{1}{(1 - h_{ii})^2}$$

The GCV obtained by writing:

$$\text{GCV} = \frac{\frac{1}{M} \sum_{i=1}^M |e_i^*|^2}{\left(1 - \frac{1}{M} \sum_{i=1}^M h_{ii}\right)^2}$$

It has a similar nature to PRESS but does not separately weight the points. This can be a disadvantage when the location of sensitive points and their stability to outlier data is being assessed as well as simply fitting data.

Both PRESS and GCV have the property that as the model order increases the fit becomes less stable and the statistics (as expressed by how close  $(1 - h_{ii})$  is to zero) increase the errors. As the model order increases, RMS and RSS will both continue to decrease. PRESS and GCV are therefore commonly used to choose the ‘best’ model order as described in this report.

[NOTE: The microBRIAN program Model used PRESS to help select the ‘best’ order of model to use. The SIEVER model (see Appendix 8 below) also has a calculation for the value of  $w_i$  to help identify points where errors will have greatest effect on the LS solution. These points can be the best to use if GCPs are accurate but also the worst to use if GCPs are inaccurate.]

## A7.3 Singular Value Decomposition

The Singular Value Decomposition (SVD; Lanczos, 1958) provides a convenient method to solve the LS problem and also derive PRESS or GCV.

The SVD of the  $M$  row by  $N$  column matrix  $\mathbf{A}$  is the (unique) decomposition of the matrix into factors such that:

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are ortho-normal and  $\mathbf{S}$  is a diagonal matrix such that:

$$\mathbf{U}^T \mathbf{U} = \mathbf{I}_M$$

$$\mathbf{V}^T \mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}_p$$

$$\mathbf{S} = \text{diag}[s_1, s_2, \dots, s_p]$$

$$s_1 \geq s_2 \geq \dots \geq s_p \geq 0$$

The  $p \times p$  matrix  $\mathbf{V}$  is the eigenvector matrix for  $\mathbf{A}^T \mathbf{A}$  and the squares of the singular values ( $s_j$ ) are the eigenvalues. The  $M \times p$  matrix  $\mathbf{U}$  consists of the first  $p$  columns of the eigenvector matrix for  $\mathbf{A}\mathbf{A}^T$ .

If the matrix has Rank ‘ $q$ ’ ( $< p$ ) then:

$$s_q > 0$$

$$s_{q+1} = \dots = s_p = 0$$

Also, if  $\mathbf{u}_j$  is the  $j$ ’th column of  $\mathbf{U}$  and  $\mathbf{v}_j$  is the  $j$ ’th column of  $\mathbf{V}$  then the SVD can also be written as:

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T = \sum_{j=1}^p s_j \mathbf{u}_j \mathbf{v}_j^T$$

The famous theorem of Eckert and Young (1936) concerning the approximation of a matrix  $\mathbf{A}$  by a matrix  $\mathbf{B}$  of lower rank is provided with a solution by the SVD in that the best approximation to  $\mathbf{A}$  by a matrix  $\mathbf{B}$  of rank  $q < p$  in the sense of minimum square norm is simply:

$$\mathbf{B} = \sum_{j=1}^q s_j \mathbf{u}_j \mathbf{v}_j^T$$

$$\|\mathbf{A} - \mathbf{B}\|^2 = \sum_{i=1}^M \sum_{j=1}^p |a_{ij} - b_{ij}|^2$$

$$= \sum_{j=q+1}^p s_j^2$$

The SVD can be used to define a ‘generalised inverse’ for the matrix  $\mathbf{A}$  such that:

$$\mathbf{A}^+ = \mathbf{V}\mathbf{S}^+ \mathbf{U}^T$$

$$s_j^+ = \begin{cases} \frac{1}{s_j} & \text{if } s_j > 0 \\ 0 & \text{if } s_j = 0 \end{cases}$$

The relationship with LS is that when the matrix  $\mathbf{A}$  is the LS matrix above:

$$\begin{aligned}\hat{\alpha}^* &= \mathbf{A}^+ \mathbf{z} \\ \hat{e}^* &= (\mathbf{I} - \mathbf{A}\mathbf{A}^+) \mathbf{z} \\ \mathbf{H} &= \mathbf{A}\mathbf{A}^+ = \mathbf{U}\mathbf{U}^T\end{aligned}$$

where in this case the columns of the matrix  $\mathbf{U}$  are reduced to those corresponding to non-zero singular values. It follows that  $h_{ii}$  may be easily derived if the SVD is used to solve the LS problem since:

$$\begin{aligned}h_{ii} &= \sum_{j=1}^p u_{ij}^2 \\ w_i &= \frac{1}{\left(1 - \sum_{j=1}^p u_{ij}^2\right)^2}\end{aligned}$$

The SVD is also used to check that LS solutions are ‘well posed’ in that singular values are either zero or not ‘close’ to zero. Near-zero singular values correspond to unstable parameters. Various methods exist to modify the LS equations (often called ‘ridge’ regression) to provide more stable (but biased) solutions. These will not be pursued here.

## A7.4 Predictive Variance

Predictive Variance (PV) is a related but different idea of the variation away from control. In this case, the resulting equation expresses how the model may vary at places away from the control due to the variation we know is in our estimate of the model parameters.

We can consider that the data vector values  $\mathbf{z}$  may have a variance, which we can estimate from the residual errors or (better) the PRESS. We will assume (although mainly because it is rare to have enough information to do otherwise—the expressions can be derived under different knowledge) that the error model is:

$$\begin{aligned}\mathbf{z} &= \mathbf{z}' + \delta\mathbf{z} \\ E(\delta\mathbf{z}) &= \mathbf{0} \\ \text{Var}(\delta\mathbf{z}) &= E(\delta\mathbf{z}\delta\mathbf{z}^T) = \sigma^2 \mathbf{I}\end{aligned}$$

Even accepting that the LS model predicted data values are an estimate for  $\mathbf{z}'$  there will be an expected variation around the model parameters ( $\delta\alpha$ ) with mean zero and variance:

$$\begin{aligned}\delta\alpha &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \delta\mathbf{z} \\ E(\delta\alpha) &= \mathbf{0} \\ \text{Var}(\delta\alpha) &= E(\delta\alpha\delta\alpha^T) = \sigma^2 (\mathbf{A}^T \mathbf{A})^{-1}\end{aligned}$$

The variation of the model in the ‘to’ coordinate system at a point predicted from the point  $(x,y)$  in the ‘from’ system is therefore:

$$\begin{aligned}\delta z(x,y) &= \phi(x,y)^T \delta\alpha \\ E(\delta z) &= 0 \\ \text{Var}(\delta z) &= \sigma^2 \phi(x,y)^T (\mathbf{A}^T \mathbf{A})^{-1} \phi(x,y) \\ &= \sigma^2 h(x,y)\end{aligned}$$

The function  $h(x,y)$  is related to the Predictive Error above since at the GCP points:

$$h(x_i, y_i) = h_{ii}$$

The function  $h(x,y)$  may be plotted over the ‘from’ space coordinate system to indicate how variations may increase away from control. To allow for the possibility of higher order models being present this is often done using PRESS for the data variance and a higher order of model than the one used for the fitting. Areas of high predictive variation are ones where more control should be obtained if possible.

## A7.5 References

- Allen, D.M. (1974). The relationship between variable selection and data augmentation and a method of prediction. *Technometrics*, 16, 125–127.
- Eckert, C., and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1, 211–218.
- Golub, G.H., Heath, M., and Wahba, G. (1979). Generalised cross validation as a method for choosing a good ridge parameter. *Technometrics*, 21, 215–223.
- Lanczos, C. (1958). Linear systems in self-adjoint form. *American Mathematical Monthly*, 65, 665–679.
- Sherman, J., and Morrison, W.J. (1949). Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Annals of Mathematical Statistics*, 21, 124–127.
- Tuckey, J.W. (1967). Discussion of Anscombe (1967). *Journal of the Royal Statistical Society B*, 29, 47–48.





# Appendix 8

## ‘Sieving’ Outliers

### A8.1 Basic Assumptions and Framework

This Appendix explains the basis for an approach to identifying outliers in spatial datasets used for EO image registration (see Volume 2B). Typically such spatial datasets are either a pair of EO images to be coregistered, or an EO image and its corresponding map data, but may also involve more than two datasets. In the microBRIAN image processing system (Harrison and Jupp, 1992), the program SIEVER implemented this approach and the following sub-sections describe it in terms of the functionality of the SIEVER model.

The context of this approach is that there exists an image or a number of images (perhaps overlapping frames or the same area at different times) and a map base or coordinate system within which they will be referenced. Clearly identified points (GCP or Ground Control Points) are selected pairwise between the spatial datasets as accurately as possible. This may be done manually or by computer (such using a correlation program). In the case of the identification of positions from a map base it is generally necessary to manually identify the GCP (see Volume 2B).

For this approach, it is necessary that each identified feature have a unique identifier so that a given feature can be located in all of the images to be coregistered, or to the map base frame. This may be accomplished simply by associating a separate list of GCP with each

image or map base and associating each record in the file with a unique spatial feature. In this way, if a feature is not present in one dataset, its location in the associated file is empty. If this or a similar method of identifying features across multiple images is available then the GCP that are common to all of a group of images can be identified quickly.

This method also assumes that major geometric distortions have been ‘nominally’ removed from EO imagery, that is, panoramic distortion and other artefacts created by the geometry of the EO sensor viewing perspective and acquisition method have been taken into account in the coordinates of the image GCP. The remaining geometric differences between the spatial datasets should then be accounted for (or well described) by an ‘affine’ transformation. An affine transformation is linear such that if one image has coordinates  $(x,y)$  and the other  $(x',y')$  then the relationship between the coordinates has the form:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_{01} \\ a_{02} \end{bmatrix}$$

This six parameter transformation can be thought of as some combination of shift of origin, rotation, skew and possibly a separate ‘stretch’ in the  $x$  and  $y$  directions (see Appendix 6 above and Volume 2B).

### A8.2 Fitting Transformation Models

For any pair of spatial datasets there can be transformations from each coordinate system to the other. If all of the transformations have been determined then any feature in one dataset can be located in each of the others. In the case where all of the transformations can be represented by a simple affine transformation then if the transformation is exactly known in one direction it can be inverted to obtain the transformation in the other direction.

However, transformations are never known exactly. Normally, the GCP are used to estimate the coefficients by least squares or some other measure of goodness of fit between the original coordinates and those predicted by the transformation. The GCP are normally subject to error or ‘noise’ and in this process any outliers present will play a disturbing role.

There will always be some error or noise level that limits the accuracy of location of the GCPs and therefore also limit the accuracy of the estimated coefficients of a transformation. However, ‘outliers’ are points that are inaccurate well beyond this base of error or ‘noise’.

When the transformation is being estimated it is usual to suppose that errors in GCP are only in those of the ‘TO’ side of the transformation. That is in the affine transformation above we are calling the  $(x,y)$  data the FROM side of the transformation (or the coordinates the transformation acts on) and the  $(x',y')$  data the TO side of the transformation (or the coordinates produced by the action of the transformation). In this case, assuming the errors are uncorrelated, it is possible to separate the fit to the  $x$  and  $y$  components of the TO side as:

$$x'_i = a_{10} + a_{11}x_i + a_{12}y_i + \varepsilon_i$$

and

$$y'_i = a_{20} + a_{21}x_i + a_{22}y_i + \eta_i$$

where  $\varepsilon_i$  and  $\eta_i$  and noise variances and the coefficients can be efficiently found by two separate least squares solutions.

However, the errors usually occur in both the FROM and the TO coordinate systems—a fact that is utilised when the inverse transformation is estimated—where the errors are then assumed to be only in the FROM side of the transformation. In the case of an affine transformation it is possible to solve for the coefficients assuming errors in both variables—but only with a simplified error model. This solution is a by-product of the work described here and will be described in passing.

The more significant issue, however, relates to outliers. In this case, if the outlier is in a GCP that is on the FROM side of the system then it is difficult if not impossible to actually identify the outlier. Its presence is often only indicated by a very poor fit to the data.

Moreover, if the outlier is on the TO side of the system then because least squares balances all errors it is likely the fit to the outlier will be better than to other data.

The situation where the outliers are strongly fitted by the model—and hence very difficult to identify as outliers—is closely related to a concept called ‘predictive error’ (see Appendix 7 above). The predictive error is the error at a point between the data and the model prediction based on all data *except* the point in question (see Volume 2B—Section 4). Conversely, it measures the sensitivity of the model to the data at the point or the degree of control that the data point exerts on the model. The sensitivity can be expressed as a ‘predictive error multiplier’ which will be large if the model is very sensitive to the presence of a data point.

High predictive error multipliers are both good and bad. A high value indicates a very important point in the modelling—or one that exerts a lot of influence over the model. However, such points are also the ones that introduce the greatest effects of error and when they are an outlier the effect can be very great but the actual error at the point is usually small—since the model fits the outlier in preference to other points!

The SIEVER algorithm implements three analyses:

- solves for the affine model without assuming the errors are all on one of the FROM and TO sides of the transformation model;
- estimates the predictive errors of the residuals as one means of identifying outliers; and
- plots the squares of the errors (which would have a Chi-square distribution if they were from a normal population or errors) against the Gamma distribution in what is called a Q-Q (or Quantile-Quantile) plot (see Volume 2B—Section 4). This also allows errors that are too large to be explained as simply large but still possible error values to be identified.

### A8.3 The SIEVER Model

The SIEVER model assumes there is a set of  $N_v$  spatial datasets with  $M$  common GCP as identified by their individual GCP files. These GCP sets can be represented as a matrix with  $M$  rows and  $2N_v$  columns:

$$A = \begin{bmatrix} \underline{x}_1, \underline{y}_1, \underline{x}_2, \underline{y}_2, \dots, \underline{x}_N, \underline{y}_N \end{bmatrix}$$

(where the curled underscore indicates a column vector), which is the starting point for the analysis.

Because we have assumed that nominal geometric transformations have been carried out on the data to remove the major image distortions (see Volume 2B—Sections 2 and 3) and that the relationships between the images can (in the absense of noise) be modelled

as affine transformations, the situation is one of the Generalised Linear Model (GLM) or Factor Analysis. That is, we are assuming that there is an underlying coordinate system  $(\mu, \nu)$  such that for each of the image coordinate systems we have:

$$x_{jk} = a_{j0} + a_{j1}\mu_k + a_{j2}\nu_k + \varepsilon$$

and

$$y_{jk} = b_{j0} + b_{j1}\mu_k + b_{j2}\nu_k + \eta$$

where  $(\varepsilon, \eta)$  are the errors in the coordinates and  $(\mu_k, \nu_k)$  are the  $M$  common (and transformed) coordinates.

We are going to assume for convenience that the images coordinates have had the means extracted from them so that the data have ‘zero mean’. The underlying coordinates will also be assumed to have ‘zero mean’. In this case it is possible to express the form of the underlying transformation as:

$$\begin{bmatrix} x_{jk} \\ y_{jk} \end{bmatrix} = \begin{bmatrix} a_{j1} & a_{j2} \\ b_{j1} & b_{j2} \end{bmatrix} \begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix} + \begin{bmatrix} \varepsilon \\ \eta \end{bmatrix} \\ = B_j \begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix} + \begin{bmatrix} \varepsilon \\ \eta \end{bmatrix}$$

It follows that we could express the matrix  $A$  in the form:

$$A = \begin{bmatrix} \mu_1 & \nu_1 \\ \mu_2 & \nu_2 \\ \dots & \dots \\ \mu_M & \nu_M \end{bmatrix} \begin{bmatrix} B_1^T & B_2^T & \dots & B_N^T \end{bmatrix} + E$$

where  $E$  is the matrix of random errors or ‘noise’. That is, The underlying model is one where  $A$  has basic rank 2 or  $A$  is the product of a matrix of size  $M \times 2$  and one of size  $2 \times 2N_v$ .

## A8.4 Solving for Underlying Model and Errors

To separate the underlying model from the errors we will use the theorem of Eckart and Young (1936) which states that the best approximation to a matrix by one of lower rank can be conveniently found using the Singular Value Transformation (SVD) of the matrix. The SVD (introduced by Lancos, 1958) provides an eigenvalue decomposition for an arbitrary matrix. That is, if  $A$  is any  $M \times N$  matrix then there exist matrices  $U$  ( $M \times N$ ),  $S$  ( $N \times N$  and diagonal) and  $V$  ( $N \times N$ ) that satisfy the conditions:

$$S = \text{diag}[s_1, s_2, \dots, s_N] \\ s_1 \geq s_2 \geq \dots \geq s_N \geq 0 \\ U^T U = I_N \\ V^T V = V V^T = I_M \\ A = U S V^T$$

Since, using these rules we have that:

$$A^T A = V S^2 V^T$$

it follows that the squares of the singular values are the eigenvalues of  $A^T A$  and the matrix  $V$  is the matrix of its eigenvectors. An alternative way to write the SVD is in terms of the column vectors of  $U$  and  $V$  in the form:

$$A = \sum_{j=1}^N s_j u_j v_j^T$$

The Eckert-Young theorem can be stated in the way used here in the following form. If  $B$  is the best approximation to  $A$  by a matrix of rank  $p < N$  where ‘best’ means in the sense of minimising the sums of squares of differences between all elements of the matrices then:

$$A = B + E$$

$$B = \sum_{j=1}^p s_j u_j v_j^T$$

$$E = \sum_{j=p+1}^N s_j u_j v_j^T$$

$$\|A - B\|^2 = \sum_{i=1}^M \sum_{j=1}^N (a_{ij} - b_{ij})^2 \\ = \|E\|^2 = \sum_{j=p+1}^N s_j^2$$

As with many cases of least squares there is an Analysis of Variance (AOV) for this generalised system in the form:

$$TSS = MSS + RSS$$

$$\sum_{i=1}^M \sum_{j=1}^N a_{ij}^2 = \sum_{i=1}^M \sum_{j=1}^N b_{ij}^2 + \sum_{i=1}^M \sum_{j=1}^N e_{ij}^2$$

$$\|A\|^2 = \|B\|^2 + \|E\|^2$$

or

$$\sum_{k=1}^N s_k^2 = \sum_{k=1}^p s_k^2 + \sum_{k=p+1}^N s_k^2$$

where TSS stands for the Total Sum of Squares, MSS the Model Sum of Squares and RSS the Residual Sum of Squares.

Of special interest to us is the AOV for a single row of each matrix. Let the  $i$ 'th row of a matrix be denoted by  $r_i(A)$ . It is a row vector that will, in this case, have  $N$  components. Then:

$$TSS_i = MSS_i + RSS_i$$

$$\|r_i(A)\|^2 = \|r_i(B)\|^2 + \|r_i(E)\|^2$$

or

$$\sum_{k=1}^N s_k^2 u_{ik}^2 = \sum_{k=1}^p s_k^2 u_{ik}^2 + \sum_{k=p+1}^N s_k^2 u_{ik}^2$$

From this (or directly) it can be seen that

$$\|r_i(A)\|^2 = \|r_i(US)\|^2$$

It is easy to see that in our specific case, where  $p=2$  and  $N$  as used above is actually  $2N_v$  then the corresponding solution to the SIEVER model is:

$$\begin{bmatrix} \mu_1 & V_1 \\ \mu_2 & V_2 \\ \dots & \dots \\ \mu_M & V_M \end{bmatrix} = \begin{bmatrix} s_1 u_{1\sim 1} & s_2 u_{2\sim 2} \end{bmatrix}$$

$$\begin{bmatrix} B_1^T & B_2^T & \dots & B_N^T \end{bmatrix} = \begin{bmatrix} \tilde{V}_1^T \\ \tilde{V}_2^T \end{bmatrix}$$

Because of the orthonormal properties of  $V$  it therefore follows that the AOV above on  $US$  is a direct analysis of the errors between the underlying two-factor model of the GCP and the selected GCP. Moreover, the row AOV above is an analysis of each row or identifiable spatial feature and our objective is to isolate the outliers to the features and images in which they occur.

Note that these steps have overcome the very serious problems that can occur if it is assumed that the errors are all in one 'side' of the equations. However, we still have a difficulty in that least squares 'spreads' error across all the points and can be dominated by a few points with strong 'control'. If these points are also outliers then their identification can be very difficult. In many cases the best tool to use is the human eye and so in the following there will be a number of suggested plots that are sometimes much more decisive than the accompanying statistics!

## A8.5 Analysing the SIEVER Solution

The first action in SIEVER is to normalise the matrix  $A$  so that the columns have zero mean and the sum of squares of the columns is 1.0. In this case, the matrix  $A^T A$  is in 'correlation' matrix form.

The value of the normalisations is twofold:

- firstly, as noted above, the formulation is more convenient when the columns have zero mean; but
- secondly, the normalisation frees the data of scale changes that different coordinate systems may have.

The SVD of the matrix is then formed to obtain the components discussed above.

### A8.5.1 Basic 'PCA' summary

The SVD of the normalised matrix  $A$  can be interpreted as a Principal Component Analysis or PCA of  $A$  in correlation form (see Appendix A6.3 above and Volume 2C).

For each of the  $N$  components, a simple set of summary statistics comprises

- the Principal Variance ( $P_{var}$ );
- the percent variance ( $Var\%$ );
- the Total Variance ( $T_{var}\%$ ); and
- the Noise to Signal Ratio (as %;  $NSR\%$ );

each of which is defined by:

$$P_{var,k} = s_k^2$$

$$Var_k (\%) = 100 \times \frac{s_k^2}{\sum_{j=1}^N s_j^2}$$

$$T_{var,k} (\%) = 100 \times \frac{\sum_{j=1}^k s_j^2}{\sum_{j=1}^N s_j^2}$$

$$NSR_k (\%) = 100 \times \frac{\sum_{j=k+1}^N s_j^2 / (N-k)}{\sum_{j=1}^k s_j^2 / k}$$

These can be calculated for any PCA or SVD exercise. In our case, if the rank of the system is 2, we should have the Total variance in the first two components as very high and the contribution of the remaining components as very low and NSR should reach its minimum at  $k=2$ .

It may be shown that these statistics are not very reliable for only two images and it is better in every way if SIEVER is used on a number of images in combinations—such as up to five in an initial investigation and down to studies of pairs of images in final outlier searches.

### A8.5.2 Row or feature-based error analysis

The issue, of course, is how to measure what is ‘high’ and ‘low’ with regard to the model fit. To help with this we can also generate some statistics for the rows of  $\mathbf{A}$ —that is for the spatial features that are being mapped as GCP in the images being registered.

For this we will work assuming that the assumption of the rank of  $\mathbf{A}$  ( $p$ ) being 2 is confirmed by the PCA analysis. The consequences of its not being true are discussed in subsequent sub-sections. The assumption we have from this is that the estimated variance of the residual can be written:

$$\tilde{\sigma}^2 = \frac{1}{N-p} \sum_{j=p+1}^N s_j^2$$

For Feature  $i$ , the row or feature statistics we can generate to help to go on from here are:

- the Normalised Error ( $e_i^2$ );
- the feature Chi-square ( $X_i^2$ ); and
- the Predictive Error Weight ( $PE_i(\%)$ ).

as defined by:

$$e_i^2 = \frac{1}{N-p} \sum_{k=p+1}^N s_k^2 u_{ik}^2$$

$$X_i^2 = M \frac{e_i^2}{\tilde{\sigma}^2}$$

$$PE_i = 100 \times \left( \frac{1}{1 - \sum_{k=1}^p u_{ik}^2} - 1 \right)$$

The derivation of the Predictive Error formula and its relationship with ‘jackknifing’ is described fully in Appendix 7 above.

## A8.6 Using the Q-Q Plot

To investigate the distribution of the errors remaining after fitting the general affine model, SIEVER uses a technique that is very useful from exploratory data analysis called Quantile-Quantile plotting (see also Volume 2B—Section 4.2.3).

### A8.6.1 Quantile-Quantile plot

A Quantile-Quantile (Q-Q) plot is constructed to test the statistical distribution of data against a known model. A ‘quantile’ is a fraction (which when expressed as a percentage is called a ‘percentile’) between zero and one. Consider that you have  $M$  data values ( $y_j$ ) which are sorted into increasing order such that:

$$y_1 \leq y_2 \leq \dots \leq y_M$$

Associated with each  $y_i$  is a estimate of the fraction (or quantile) of data values less than or equal to  $y_i$ . This is simply:

$$q_i = \frac{i-1/2}{M}$$

Suppose  $F(x)$  is the cumulative distribution function for a probability distribution function  $P(x)$  such that:

$$F(x) = \int_{-\infty}^x P(x') dx'$$

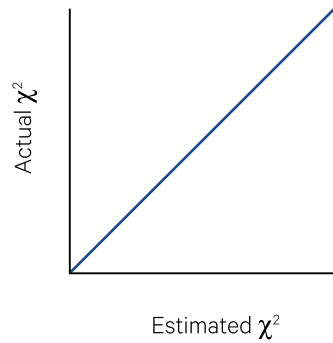
The set of  $x$  values that satisfy:

$$F(x_i) = q_i \quad i = 1, M$$

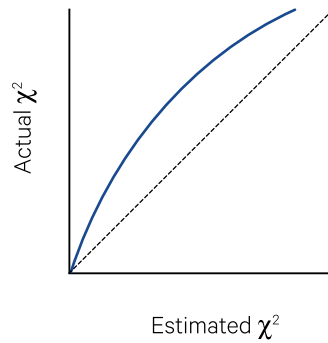
are the set of values for which the expected fraction of values from a sample of  $M$  ‘drawings’ which are less than or equal to  $x_i$  is the quantile  $q_i$ . It follows that if the data  $y_i$  are drawings from the distribution function  $P(x)$  then the plot of  $y_i$  against  $x_i$  for  $i=1, M$  will be close to a straight line with slope one and intercept zero. This plot is the Q-Q plot (see Figure A8.1) and tests whether the data values are samples from the distribution  $P(x)$ .

**Figure A8.1** Q-Q plot

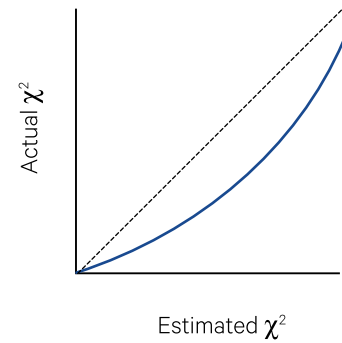
a. If actual and estimated chi-square values agree for all control points then the slope of Q-Q plot equals 1



b. If the slope is greater than 1, then some outliers exist in the set of control points.



c. If the slope is less than 1, there are too few control points.



Source: Harrison and Jupp (1992) Figure 40

## A8.6.2 Gamma probability plotting

To test the distribution of the residual errors from the SIEVER model, SIEVER uses a general method called Gamma probability plotting. This allows the distribution to be tested against a family of possible distributions, one of which is  $\chi^2$ . The hypothesis is that if the residuals are normal or near normal then the statistic  $X_i^2$  defined in the previous section will be distributed as Chi-square with  $(N-p)$  degrees of freedom, or  $\chi_{N-p}^2$ . The cumulative distribution for the incomplete Gamma distribution can be defined as:

$$F(x; \alpha, \lambda, \eta) = \int_{\alpha}^x P(x'; \alpha, \lambda, \eta) dx'$$

$$P(x; \alpha, \lambda, \eta) = \frac{\lambda^{\eta}}{\Gamma(\eta)} (x - \alpha)^{\eta-1} e^{-\lambda(x-\alpha)}$$

If  $x$  is  $\chi_r^2$  then  $\alpha = 0$ ,  $\lambda = 1/2$  and  $\eta = r/2$ . It follows that if the  $y_i$  values above are the sorted,  $X_i^2$  defined as in the previous section, and you obtain the values  $x_i$  by solving:

$$F(x_i; 0, 1/2, (N-p)/2) = q_i$$

then the Q-Q Plot of the  $x_i$  values against the  $y_i$  values should be a straight line with slope one and intercept zero (see Figure A8.1a). In SIEVER this is tested by a linear regression and associated statistics.

## A8.6.3 Interpreting the Q-Q plot

The plot can depart from the ideal for a number of reasons. One is that the errors are not normal so the  $X_i^2$  are not distributed as  $\chi^2$ . It is known that for the general incomplete Gamma distribution the fitted linear regression will have slope  $\lambda/2$  and intercept  $\alpha$ . Also, if the value ( $\eta$ ) for which the data become well approximated by a straight line is different from the theoretical value, it may mean the effective degrees of freedom are different from  $(N-p)$ .

However, any of these reasons simply relates to the distributions and not the extreme events that represent outliers. As discussed further in Appendix A8.7, outliers are often indicated by very high values of the  $y_i$  data for the high quantiles. In this case, the slope is usually much greater than one and the intercept large and negative.

Once outliers are detected they should be traced to the images where the features have been located and checked very closely. Removing points suspected of being outliers is not always justified as in any well-defined distribution there can be large values. SIEVER actually provides a good indication of when there are too few large errors (which might happen if points with large errors are arbitrarily deleted), since in this case the Q-Q plot will often 'flatten' at the higher quantiles indicating there has been truncation of larger but still statistically possible errors (see Figure A8.1b).



## A8.7 Data Plots to Indicate Outliers

In addition to the Q-Q Plot, SIEVER provides for some other plots that can give some assurance that the analysis is working as expected or for the location of outliers. These involve listing and plotting columns of the matrix  $US$  and/or the Error Matrix.

### A8.7.1 Listing and plotting the matrix $US$

As noted above, the matrix first two columns  $US$  constructed from the SVD can be identified with the modelled 'underlying' set of feature coordinates:

$$\begin{bmatrix} \mu_1 & \nu_1 \\ \mu_2 & \nu_2 \\ \dots & \dots \\ \mu_M & \nu_M \end{bmatrix} = \begin{bmatrix} s_{1\sim 1} & s_{2\sim 2} \end{bmatrix}$$

The last  $N-2$  columns can be identified as orthogonal transformations of the errors between the affine model and the data.

If only two spatial datasets are being considered, there will be two columns of the underlying coordinates and two columns of orthogonalised errors. If more than two datasets have been considered then there will be more possible error plots.

An XY plot of the first two columns will indicate if the transformation is working well as the coordinates should be recognisable as linear transformations

(such as rotations and scaling) of the coordinates in any one of the datasets. An XY plot of pairs of the columns of orthogonalised errors can locate outliers or large errors and confirm any findings from the Q-Q plot. These alternative analyses of the data are important since, in many cases, outliers so affect the SIEVER model that the actual location is hidden, apart from the general lack of fit or improbability of the model result.

### A8.7.2 Listing and plotting the Error Matrix

Another way to search for errors in the location of features, and to identify the image from which they are likely to have come, involves listing and plotting of the Error Matrix. The Error Matrix ( $E$ ) is obtained by 'rotating the errors back' into the original system. That is, it is a reconstruction of the data matrix but with the model components left out:

$$E = \sum_{k=p+1}^N s_k u_k v_k^T$$

It is the estimated error matrix in the model for the data matrix  $A$  into model and error as:

$$A = B + E$$

XY plots of the columns of  $E$  are sometimes very helpful, but are also often either hard to interpret or not much more informative than the Q-Q plot and plots of the columns of  $US$ .

## A8.8 Using SIEVER to Locate Outliers

The SIEVER approach provides a set of tools that need to be applied skillfully. The main problem is that outliers, especially when they occur at features with a high Predictive Error weight, can affect the model to a degree where the outlier or outliers may well be fitted better than the majority of the features. Guarding against this possibility is important and is best done by ensuring that the features SIEVER identifies as having a large predictive error (PE) weight are as accurate and well positioned as possible. A high PE weight indicates that a point has a lot of control over the model. This can be good and often some of the most important features are those with high PE weight. However, if a point with high PE weight is in error it can be very bad for the model.

The existence of a number of outliers to the point where there is not a clear distinction between the fitted errors and the outliers is usually indicated by a very poor fitting model and a lack of clear support for the rank of the  $A$  matrix as 2. These situations indicate the need to look carefully among the located features and images for the problem points.

What is more common is for the outliers to be clearly indicated as large errors after the model is fitted. These unusually large errors will show up in the Q-Q plot and in the XY plots of the columns of  $US$  and/or the columns of the estimated error matrix  $E$ .

Because of the possibility that eliminating points with large, but still statistically feasible, errors from the dataset will create a truncated distribution, it is important to improve the locations of features rather than deleting them. The Q-Q plot allows you to see if this has occurred. It is important not to over-SIEVE datasets before models are fitted. The objective is to find outliers and not remove points with large but statistically feasible error values.

The objective of the SIEVER approach is to identify and advise. The feature identifications in the spatial datasets should be revisited before action is taken with a specific point. If this is kept in mind SIEVER provides a very powerful means for screening collections of GCP and preparing the corresponding datasets for resampling.



## A8.9 References

---

- Eckart, C., and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3), 211–218. doi: 10.1007/BF02288367.
- Harrison, B.A., and Jupp, D.L.B. (1992). *Image Rectification and Registration. Part FOUR of the microBRIAN Resource Manual*. CSIRO and MPA, Melbourne.
- Lanczos, C. (1958). Linear systems in self-adjoint form. *American Mathematical Monthly*, 65, 665–679.









Australian Government  
Geoscience Australia



bushfire & natural  
**HAZARDS**CRC