ELSEVIER

Remote Sensing of Environment 77 (2001) 186–196

Remote Sensing of Environment

www.elsevier.com/locate/rse

Calibrating images from different dates to 'like-value' digital counts

S.L. Furby*, N.A. Campbell

CSIRO Mathematical and Information Sciences, Private Bag 5, Wembley, Perth, Western Australia 6913, Australia

Received 25 September 1998; received in revised form 11 July 2000; accepted 25 February 2001

Abstract

Sequences of satellite images are increasingly being used to monitor our environment. An important step in the development of methods for detecting, measuring, and monitoring change through time is the ability to compare images from different dates and sites in different scenes. These comparisons require the digital counts from each scene to be calibrated to common reference values. A method of relative calibration using robust regression and large numbers of likely invariant targets has been developed and implemented operationally. One or more images are calibrated to the digital numbers of a common reference scene. The technique has been used to calibrate over 100 Landsat Thematic Mapper (TM) scenes of the Western Australian wheatbelt. The method and its application are described. The repeatability of the calibrations over a sequence of images of the same scene from different dates is examined. © 2001 Elsevier Science Inc. All rights reserved.

Keywords: Image calibartion; Robust regression

1. Introduction

An important step in the development of methods for detecting, measuring, and monitoring change through time in land condition is the ability to compare images from different dates and sites in different scenes. These comparisons require the digital counts from each scene to be calibrated to common reference values. The raw data values recorded by the sensors on the satellite are not consistent over time or between scenes. The sensors record the energy from the sun reflected by the earth's surface. The amount of incident energy varies throughout the year with changes in the solar zenith angle. Both the incident and reflected energy are affected by atmospheric scattering and absorption. The amount of scattering varies with factors such as temperature, humidity, and haze. Changes in sensor response over time and on-ground preprocessing also affect the raw data values.

With correctly calibrated images, simple displays of sequences of images with the same band combinations and common enhancements immediately convey changes in the images. In classification, training signatures extracted for various condition classes from one or more of the images can be applied from image to image. Areas in which the label changes from one classification to the next can be ascribed to radiometric changes in that region rather than to variations in the training data. Areas that are poorly classified in images other than the one(s) on which the classifier was trained will be indicative of areas that have changed radiometrically to a class not covered by the original training data.

A number of approaches to the calibration of images exist. These include:

(i) calibration to standard reflectance units using atmospheric radiative transfer code (see, e.g., Dave, 1978; Fraser, Ferrare, Kaufman, & Mattoo, 1989; Kaufman, 1988; Kneizys et al., 1983, 1988; Tanre, Deroo, Duhaut, Herman, & Morcrette, 1990; Vermote, Tanre, Deuze, Herman, & Morcrette, 1994),
(ii) calibration to standard reflectance units by relating image spectral curves to laboratory spectral curves (see, e.g., Smith,Ustin, Adams, & Gillespie, 1990),
(iii) calibration to standard reflectance units using dark objects and atmospheric radiative transfer code (see, e.g., Teillet & Fedosejevs, 1995),
(iv) scene correction by dark-object subtraction (see, e.g., Chavez, 1988),
(v) scene-to-scene correction by histogram normalisation (see, e.g., Chavez & MacKinnon, 1994), and
(vi) scene-to-scene correction using invariant targets (see, e.g., Casselles & Garcia 1989; Hall, Strebel, Nickeson, & Goetz, 1991; Schott, Salvaggio, & Volchok, 1988).

* Corresponding author. Tel.: +61-8-9333-6125; fax: +61-8-9333-6121.

E-mail address: suzanne.furby@cmis.csiro.au (S.L. Furby).

PII: S0034-4257(01)00205-X

Only the last two approaches directly compare one or more image pairs.

Ideally, all images would be calibrated to standard reflectance units. As well as comparing images from different dates, image pixel values could then be compared directly with field and laboratory measurements, and changes could be related directly to physical properties of the ground cover. Calibration methods that produce images in standard reflectance units require one (or more) of

- sensor radiometric calibration,
- simultaneous in situ measurement of atmospheric parameters,
- known reflectance signatures of target(s) in the image, and
- the use of simulated atmospheres.

Sensor radiometric calibration information is available (see, e.g., Slater et al., 1987), but the time dependence of such corrections is less well-known. Measurements of atmospheric parameters and ground targets are expensive and impractical for current data and impossible for historical data. Standard atmospheres over agricultural areas in Australia have not been well-characterised. This makes such procedures impossible to use for routine change detection applications.

However, when comparing images to detect change, it is sufficient to convert the raw digital counts to be consistent with the counts for a chosen reference image. Such a 'like-values' calibration procedure is described here. It is an extension of the scene-to-scene correction proposed by Casselles and Garcia (1989) that is designed to be robust against targets that may not be truly invariant and to be less labour-intensive to apply to a multitemporal sequence of images. It has been used to calibrate approximately 150 Landsat Thematic Mapper (TM) scenes from 1987 to 1994 (Bands 1, 2, 3, 4, 5, and 7) covering the southwest agricultural region of Western Australia.

It is assumed that the digital counts for any image are related to the digital counts of a chosen reference image via the following linear relationship (see Casselles & Garcia, 1989 for a derivation):

$$R_i = g_i \mathrm{DN}_i + o_i$$

where R is the digital count of the target in the reference image, DN is the digital count of the target in the raw overpass image, g and o are the gain and offset to calibrate the raw data values, and i denotes the image band.

The gain and offset account for atmospheric, sensor, and on-ground processing differences. Dark-object subtraction procedures only attempt to correct for the additive component (offset) in the above relationship and have been shown to be worse than no correction in some circumstances (Moran, Jackson, Slater, & Teillet, 1992). Other dark-object correction procedures (e.g., Teillet & Fedosejevs, 1995) use dark target(s) to provide an estimate of aerosol optical depth that can be used with atmospheric correction algorithms, essentially providing both additive and multiplicative corrections, but are subject to the problems of using simulated atmospheres mentioned previously.

Since the calibration applied is a relative one between each image in the sequence and the reference scene, it is not necessary to apply sensor radiometric calibrations or atmospheric path length corrections separately, as these operations are linear and are therefore accounted for in the derived calibration coefficients. Tests of the procedure have shown that the match between each image and the reference image is the same whether or not the corrections are applied.

The calibration coefficients are assumed to be constant over the entire image. This requires that any sensor imbalances that lead to striping in any of the images have been corrected prior to applying the calibration procedure.

Pseudoinvariant targets are used to derive the relationship between each image and the reference image. These are image features, such as deep water, bare ground, quarries, and gravel pits, which are likely to have constant reflectance over time. A robust regression procedure for estimating the gains and offsets is proposed so that changing targets are automatically omitted from the calculations. It is assumed that at least 50% of the targets are invariant between a given image and the reference image. That is, up to half the targets may change between the pair of images without affecting the estimated calibration coefficients. In contrast, histogram normalisation approaches to scene-to-scene corrections require that the histogram minima and maxima (and sometimes intermediate percentiles) are invariant between the image pairs. This assumption is often violated between image pairs from different stages within a growing season.

The calibration procedure presented here can be applied to a sequence of images from different dates over the same region or, by making use of the overlap region between adjoining scenes, two images that can be joined together to cover a larger region. It can be applied to any satellite or aircraft scanner imagery where the sensor characteristics and on-ground preprocessing operations are linear, or where known nonlinearities are first corrected, and the atmospheric corrections are linear, or can be approximated by linear functions. The linearity requirement for the atmospheric corrections implies clear atmospheric conditions. As the atmospheric optical depth increases (e.g., increasing haze), the required corrections become increasingly nonlinear.

Section 2 describes the steps necessary to produce a sequence of calibrated images. Section 3 describes the regression procedure used to estimate the calibration coefficients. In Section 4, the repeatability of the calibration procedure is assessed using several images of the same region. Section 5 discusses some issues that need to be considered when applying the procedure.

2. The calibration procedure

Calibrating a sequence of images to like-values consists of the following steps:

(i) Select a reference image.
(ii) Select invariant targets.
(iii) Calculate the calibration coefficients to calibrate each image to the reference image.
(iv) Examine the calibration curves and refine the target selection if necessary.
(v) Use the estimated coefficients to calibrate the image.

Each of these steps will be discussed briefly. The calculation of the calibration coefficients is discussed in detail in Section 3.

The reference image is the scene to which the other images will be corrected. It is important that:

(i) the image is cloud-free;
(ii) the atmosphere is relatively clear;
(iii) the data are within the storage format range for all bands;
(iv) the time of year is appropriate for the application; and
(v) the image has the best possible dynamic range.

Cloud cover causes two problems. Obviously, it reduces the number of potentially invariant targets that are available. Increased scatter has also been observed in the plots of the digital counts from the invariant targets in cloud-affected images, particularly in TM Bands 5 and 7. It is thought that variations in atmospheric moisture across the image may lead to variations in the gains across these images. These moisture variations are expected to be greater in cloud-affected images. TM Bands 5 and 7 are most sensitive to moisture variations. Hall et al. (1991) also suggest that problems with the detectors for TM Bands 5 and 7 could contribute to the increased scatter.

Other variations in the atmospheric conditions of the reference image, such as varying amounts of haze across the scene, will also affect the assumptions of linearity and constancy of the corrections across the whole image.

It is also important that the data are within the range of the storage format for all bands, e.g., Landsat TM has a maximum recordable digital count of 255. Reflected energies higher than an upper threshold are truncated and recorded as a digital count of 255. Similarly, reflected energies less than a lower threshold are recorded as a digital count of 0. Data whiteouts — where the reflectances are greater that the maximum recordable digital count — are common in TM Band 5 in summer images of the Western Australian wheatbelt. Unless very good midrange targets are available, this leads to uncertainty in the upper-end point of the calibration line, as illustrated in Fig. 1(b). Our experience suggests that the recording of zeros is less commonly a problem.

The time of the year for the reference image will often be chosen to be the time at which the majority of the data are acquired. The angle of incident energy affects the reflectance of most naturally occurring surfaces. The repeatability analysis in Section 4 shows that the calibrations are less repeatable when a summer image is included with a series of spring images. This seasonal variation was also observed by Casselles and Garcia (1989).

The application may also influence the choice of reference image. The reflectance of woody vegetation is very low and the range of digital counts recorded over such vegetation is small. Using a summer image as the reference image, where the range of recorded digital counts is greatest, prevents this range of digital counts being compressed even further.

Invariant targets are features that have constant reflectance over time. Potential invariant targets are identified from ground knowledge, maps of the area of interest, and manual inspection of the image data.

A number of targets should be selected to cover the range of bright, midrange, and dark data values. Midrange targets are as important as the bright and dark targets. They provide the means for assessing whether the assumption of a linear relationship between the two images is valid. The approach used by Hall et al. (1991) for automatically selecting invariant pixels only finds bright and dark features. The method effectively fits a line through only two points. It does not allow for finding invariant pixels in the middle of the intensity range to test or confirm the linearity assumption.

The regression procedures also require that the number and size of bright and dark targets should be balanced. Fig. 1(g) illustrates the consequences of including too many dark targets. An even geographic distribution of the targets across the image is also desirable, as this will help to minimise the problems of partial cloud cover when using the same set of targets to calibrate subsequent images of the same area. Each target should be located over a uniform, level area of the feature, and away from the edges of the feature, to minimise any errors due to misregistration. Ideally, targets should contain four to nine pixels, with the target size usually being limited by the size of the feature.

Potential targets from the Western Australian wheatbelt include:

Dark targets	Ocean
	Lakes
	Water in dams and reservoirs
Mid-range targets	Rock outcrops
	Airfields (where runways cross)
	Quarries, gravel scrapes and open mines
Bright targets	Roaded catchments (cleared areas for water catchment)
	Beach sand
	Bare ground

In our experience, vegetated targets should generally be avoided as they tend to show seasonal trends. Targets should

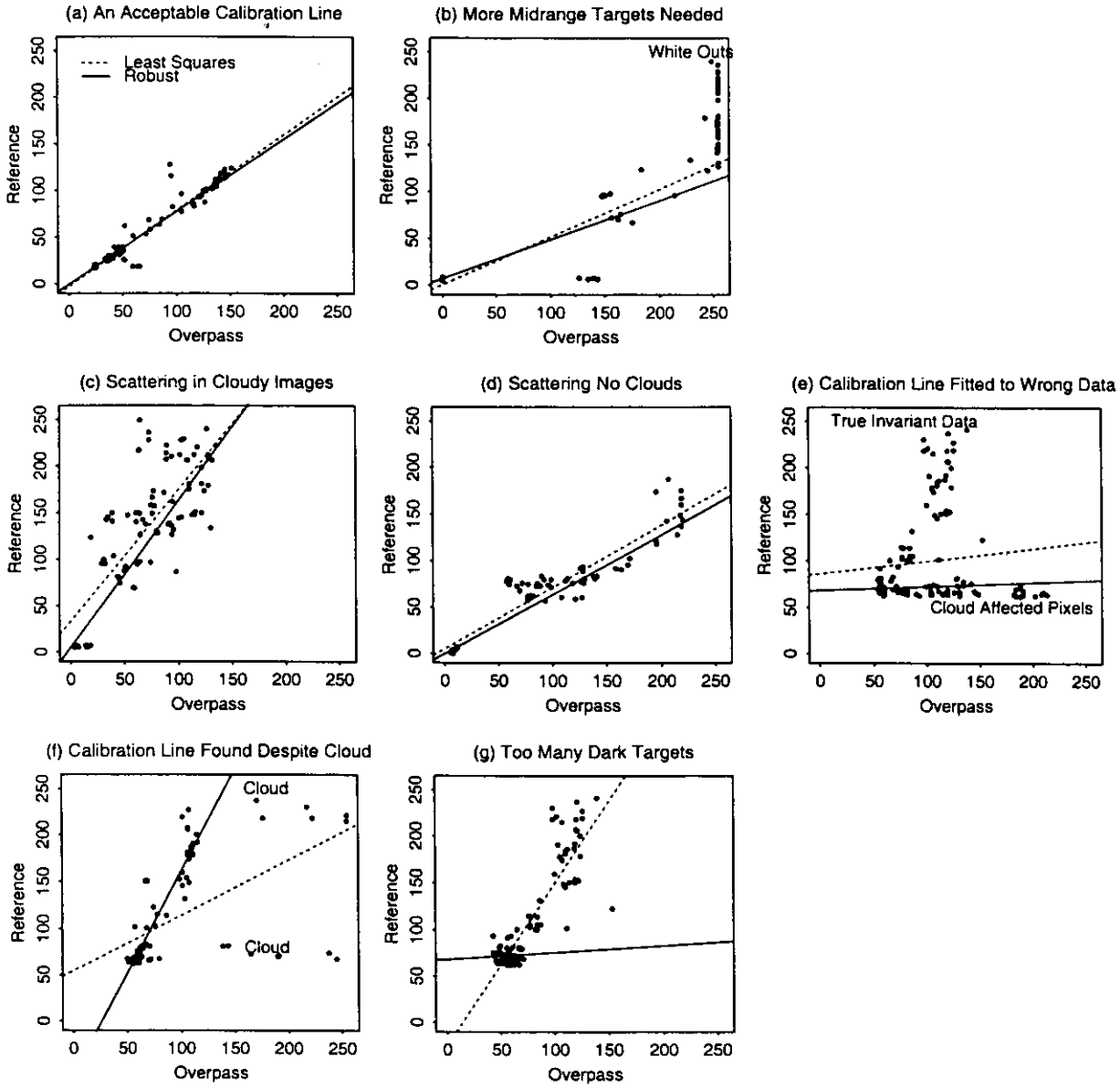


Fig. 1. Sample scatter plots showing the fitted robust (solid) and least squares (dotted) calibration lines. Plots (a) and (f) show acceptable calibration lines. The remainder show some of the problems encountered. The axes are in units of digital counts.

also be located away from terrain features that might cause shadows, such as the walls of open-cut mines. Shadowed pixels in steep terrain have been used as dark targets by others but are not found in the southwest of Western Australia.

It was anticipated that dry salt lakes, which are very prevalent in this region, would be prime candidates for invariant targets. However, our analyses have shown that such targets vary considerably from image to image compared to the targets listed above. The dry salt lakes readily absorb moisture from the atmosphere. The amount of moisture absorbed varies with barometric pressure. Small variations in moisture content cause quite large changes in reflectance in the infrared part of the spectrum, particularly in the regions corresponding to Landsat TM Bands 5 and 7.

Both Hall et al. (1991) and Schott et al. (1988) argue against manual selection of invariant targets. They argue that the process is very labour intensive, is subject to errors in the spatial registration of the images, and that the targets may change over several years if the land use changes. The robust regression approach to the estimation of the calibration coefficients presented here is unaffected by changes in up to half of the selected targets. This means that once a set of targets has been selected for any pair of scenes, any number of scenes of the same region may be calibrated using these targets without the need to manually inspect each image pair. This significantly reduces the manual effort involved when more than two images are to be calibrated. Misregistration problems are minimised by avoiding small or narrow features; however, if misregistration is an issue

between the image pairs, then the whole concept of change detection is compromised.

In the examples presented in this paper from the early investigations of the calibration procedure, as many potentially invariant targets as possible were located (typically 50–80). In the operational application of the procedure to new regions, typically 10–20 targets are used.

Once potential targets have been identified, the next step is to extract the intensity values for the target pixels from the image and calculate the regression coefficients (gains and offsets), which relate the overpass image to the reference image for each band. Coefficients resulting from a least-squares approach, a robust approach based on S-estimation (see Section 3 for a formal definition), and on weighted least-squares estimation procedures are calculated. The S-estimation method fits a line to the data in each band separately by down-weighting atypical or outlying values, and assigns a weight to each point based on its residual from the fitted line. The weighted least-squares method uses the minimum weight over all bands for each point to determine the weighted least-squares line. The computational details of the robust regression techniques are discussed in Section 3.

Robust regression techniques are needed so that the calculated gains and offsets are not unduly influenced by changing targets or changing pixels within otherwise invariant targets. In principle, the correct calibration line will be estimated provided that at least 50% of the pixels, distributed across the range of data values, are invariant. In this paper, the least-squares regression line has been calculated for comparison purposes, and to provide a warning against the routine use of least-squares regression for image-to-image calibration.

The final step in the calculations — combining the weights across all image bands — reflects the fact that if the reflectance of a target pixel has changed sufficiently in one band to warrant it being down-weighted, then something about that target has changed. While the change in reflectance may be less in other image bands, the evidence suggests that the target has changed and, hence, it is arguable that it should be omitted from the calculations (or down-weighted) for all bands. In practice, with a good set of targets, there is very little difference between the S-estimation and weighted least-squares calibration lines.

Once the calibration lines have been calculated, they should be examined on a scatterplot of the data to verify that the line being found is sensible. Fig. 1 shows seven typical plots from calibrations that we have performed. They show acceptable calibration curves, together with some of the common problems that we have encountered.

Fig. 1(a) is an acceptable calibration curve.

Fig. 1(b) shows the problems that occur when the reflectance of bright targets exceeds the range that can be recorded by the sensor. The points with values of 255 have been excluded from the calculations. In this case, more midrange targets are required to get a good estimate of the calibration line.

Fig. 1(c) and (e) illustrate problems when one or both of the images are cloud-affected.

Fig. 1(c) illustrates the typical increase in scatter in TM Band 5 in cloud-affected images. Each target pixel is several kilometers distant from any cloud or cloud shadow. Fig. 1(d) shows a typical scatterplot for an image with no cloud. Although there is more scatter generally in TM Band 5 compared to the other bands, this scatter increases in cloud-affected images. In Fig. 1(e), a significant number of the target pixels are cloud-covered in the reference image. The robust procedures fit the calibration line through the dark targets and the cloud-affected targets instead of through the dark targets and the good bright targets. Omitting the cloud-covered targets would correct the calibration line. In Fig. 1(f), the number of cloud-covered pixels is relatively few and the robust procedures ignore them to fit the correct calibration line. Similar plots are obtained when lakes dry out or bare ground becomes inundated between images. This highlights the importance of using robust regression techniques.

Cloud-screening techniques do not form part of the calibration procedure for two reasons. Firstly, small patches of cloud affecting a few targets can be accommodated by the robust procedures. The cloud-affected targets will be down-weighted in the calculations. Secondly, large areas of cloud cover which affect many sites suggest that the atmospheric conditions would lead to a breakdown of the linearity and constancy assumptions of the calibration procedure and that the whole image should be omitted from routine processing.

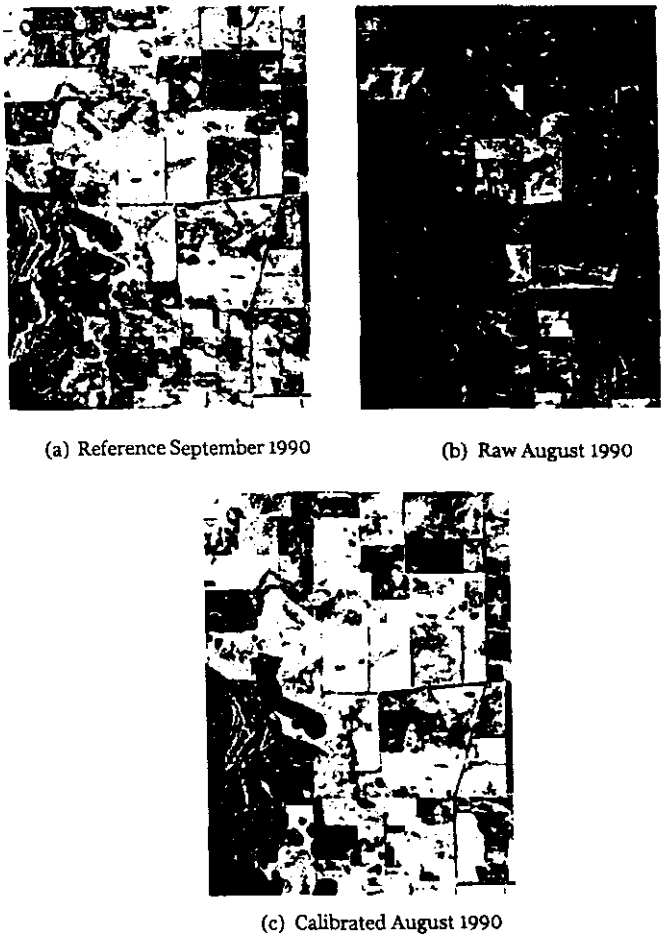
The final scatterplot, Fig. 1(g), shows the robust procedures 'breaking down' because over half the data are from dark targets. The S-estimation procedure can ignore up to half the data points when determining the line of best fit.

The final step is to perform the arithmetic on the raw image to produce the calibrated product. Fig. 2 shows a region from a raw and calibrated image together with the reference image. The darker winter image (August) has been calibrated to the spring (September) reference image. The remaining differences between the calibrated winter image and the reference spring image are due to the progression of the growing season.

3. Robust regression for estimating the calibration coefficients

Estimating the calibration coefficients is the process of finding, for each image band, the line of best fit to the data from the reference image and the image to be calibrated.

Least-squares regression is a procedure commonly used for finding the coefficients of the line of best fit. This procedure finds the line that minimises the sum of the squared residuals of each data point from the fitted line. As can be seen in the plots in Section 2, in particular, Fig. 1(f), the least-squares calibration line is strongly influenced by atypical data values. These atypical values are from targets that have changed between image dates. Ideally,



(a) Reference September 1990 (b) Raw August 1990

(c) Calibrated August 1990

Fig. 2. An example of a calibrated Landsat TM image (c) together with the raw (b) and reference (a) images. Band 4 is displayed with identical enhancements for each image.

we would like to down-weight or omit these targets from the estimation procedure.

Various robust regression procedures are available to counter this problem (see, e.g., Rousseeuw & Leroy, 1987, and references therein). The approach adopted here is based on S-estimation of the regression coefficients (Rousseeuw & Yohai, 1984, pp. 256–272). S-estimation minimises the residual sum of squares, $\sum_{m=1}^{n} r_m^2$, where the residuals are given by $r_m = R_m - o - g\mathrm{DN}_m$ and R_m are the digital counts in the reference image and DN_m are the digital counts in the raw overpass image, subject to the constraint that

$$n^{-1}\sum_{m=1}^{n}\rho(r_m/s) = E_\Phi(\rho) \qquad (1)$$

where $E_\Phi(\rho)$ is the expected value of the function $\rho(.)$ of the scaled residuals.

Here, s is the weighted residual root mean square; s^2 is given by:

$$s^2 = \sum_{m=1}^{n} w(r_m^*)r_m^2 \Big/ \sum_{m=1}^{n} \psi(r_m^*)r_m^*$$

where $r_m^* = r_m/s$ denotes the scaled residuals, $w(r_m^*) = \psi(r_m^*)/r_m^*$ is a weight function, and $\psi(r_m^*) = \rho'(r_m^*)$ the derivative of ρ. relates to the influence of a residual.

We take $\rho(.)$ to be the Tukey bi-weight function,

$$\rho(x) = \frac{x^2}{2} - \frac{x^4}{2c^2} + \frac{x^6}{6c^4} \qquad |x| \le c$$

$$\rho(x) = \frac{c^2}{6} \qquad |x| \ge c$$

so that

$$\psi(x) = x\left\{1 - \left(\frac{x}{c}\right)^2\right\}^2 \qquad |x| \le c$$

$$\psi(x) = 0 \qquad |x| \ge c$$

is a redescending influence function in which observations that have scaled residuals greater than c have zero influence on the regression. The cut-off constant, c, is chosen to ensure that the regression can accommodate a specified proportion of atypical observations and still give sensible results (see, e.g., Table 19 in Rousseeuw & Leroy, 1987, p. 142). We have taken $c = 1.85$.

The constraint in Eq. (1) has the effect that the average of the bounded function, ρ, of the scaled residuals is set equal to the value we would expect to obtain if all the residuals were from a well-behaved Gaussian sample (viz., $E_\Phi(\rho)$). In principle, a large number of possible regression lines are calculated, and the one that best fits the majority of the data is chosen.

An intuitive description of how S-estimators work is given in Campbell, Lopuhaä, and Rousseeuw (1998, Section 4).

The regression procedure is applied separately for each image band. A by-product from this procedure is a weight for each pixel based on its residual from the regression line. A weight near 1 indicates a typical pixel, which lies on or near the fitted calibration line. A weight near 0 indicates an atypical or changing pixel, which is distant from the fitted calibration line. The data for such pixels have been down-weighted in the calculation of the robust calibration line.

The final step in the estimation of the calibration coefficients is to combine the weights from all bands to produce a single weight for each pixel, which is used to perform a weighted least-squares regression for each image band. The new weight for each pixel is the minimum weight over all the image bands. As discussed earlier, if a pixel is detected as atypical in one band and we reduce or omit its influence on the calibration coefficients for that band, we want to down-weight it equally in the calculations for the other bands.

The final detail of the calculation is that pixels with the maximum data value (255 for TM images) in a given band are omitted from the calculations for that band only. Predictions of the true digital counts for the bright targets, based on the estimated calibration line when 255s are

omitted, range from 256 to over 300, depending on the target and the rest of the image. Including such targets would result in large errors in the fitted line.

Similarly, all pixels with the minimum value of 0 should also be omitted, but this has been found to be impractical. In many Landsat TM images, intensity values of 0 are quite commonly recorded in Bands 5 and 7 for water bodies, the only source of dark targets in the Western Australian agricultural areas. If such targets were omitted, the lower end of the calibration curve could not be accurately defined in many regions. Predictions indicate that where these targets differ from zero, it is only less than 5 digital counts. The errors introduced by including those targets with intensity 0 are less of a problem than attempting the calibration without them in areas where they are the only source of dark targets.

4. Repeatability of the calibrations

This section examines the reliability, or repeatability, of the proposed calibration procedure.

One way to investigate the repeatability of the calibrations is to have a separate set of invariant targets that are not used to calculate the calibration coefficients, and to compare the intensities of these targets in the reference and calibrated images. Alternatively, several calibration lines could be calculated from independent sets of invariant targets and compared. Unfortunately, in Landsat TM images of the Western Australian agricultural area, finding enough good targets to be reasonably confident of the calculated calibration line leaves few spare targets for such questions. Leave-one-out calculations could be used, but since outlying targets are down-weighted anyway by the robust calculations, leaving them out will have no effect on the results and leaving out unchanging targets can alter the balance between bright and dark targets and could potentially distort the robust regression.

Another way of assessing the repeatability of the calibration is to ask the following question. Does calibrating Image A to Image B, producing Image A*, then calibrating Image B to Image C and applying this same relationship to Image A*, produce the same calibrated image as comparing Image A directly with Image C? That is, does the path A → B → C produce the same result as the path A → C? This question is examined here using a sequence of several images of the same region from different times of the year. The discussion in this section concludes by considering four invariant targets in the calibrated sequence of images.

Six largely cloud-free scenes are available from 1990 over a typical agricultural study area. These images — from June, July, August, September, October, and December — are from Landsat TM Path 112, Row 081. The August image has a narrow band of cloud across one corner of the image; the remainder is completely cloud free. All images were acquired at least 5 days after any rainfall event. The June image has been calibrated to the September reference

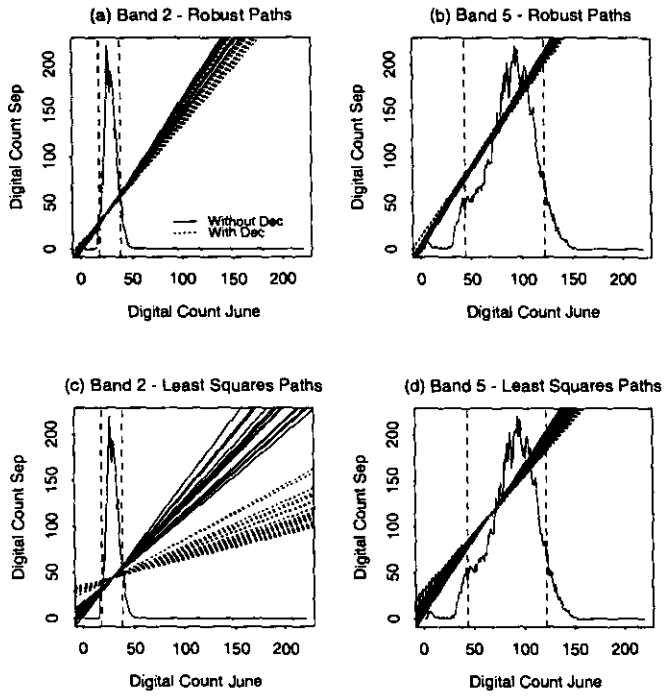


Fig. 3. Calibration lines for the 45 calibration paths for both the robust (a and b) and least squares (c and d) regressions. The dotted lines indicate paths that include the December image.

image. This has been done by comparing the two images directly and by using the other images as intermediate steps in the calibration path. The same set of invariant targets was used in all calibrations.

There are 65 possible calibration paths from June to September, as follows:

One step	1 path	e.g., Jun → Sep
Two steps	4 paths	e.g., Jun → Jul → Sep
Three steps	12 paths	e.g., Jun → Jul → Aug → Sep
Four steps	24 paths	e.g., Jun → Jul → Aug → Oct → Sep
Five steps	24 paths	e.g., Jun → Jul → Aug → Oct → Dec → Sep

The calibration coefficients were calculated for only half of the four- and five-step calibration paths; the five-step paths considered were obtained by adding an extra step to the four-step paths already calculated.

All of the pairwise calibration coefficients were calculated, both forwards, e.g., Jul → Aug, and the reverse, e.g., Aug → Jul. The calibration coefficients for each of the multiple-step paths were calculated by repeated application of the following sequence of equations:

If $y = g_1 x + o_1$

and $z = g_2 y + o_2$

then $z = g_2 g_1 x + g_2 o_1 + o_2$

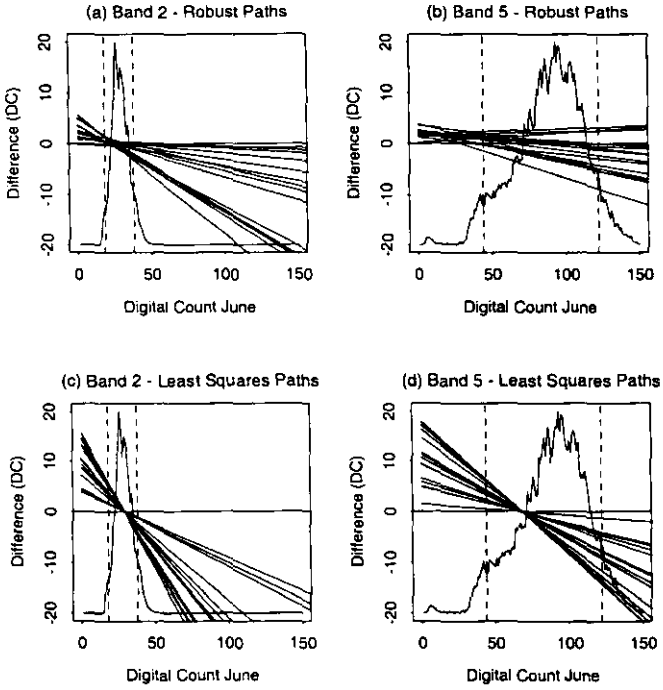


Fig. 4. Difference curves between the multistep calibration paths and the direct calibration path for both robust (a and b) and least squares (c and d) regressions. Paths including the December image have not been plotted.

Both the least-squares and the robust calibration coefficients have been calculated and compared.

Fig. 3(a) and (b) shows the 45 robust calibration lines plotted over a histogram of the June image data for TM Bands 2 and 5. It is the coincidence of the calibration lines over the actual range of the image data, rather than the full 0–255 range, that matters. The dashed vertical lines correspond to the 5% and 95% points of the cumulative histogram of the June data. The plot for TM Band 2 is representative of the variation in calibration lines for the visible bands. The plot for TM Band 5 is representative of the sort of variation in the calibration lines for the infrared bands, although the size of the differences is smaller for TM Bands 4 and 7.

The dotted lines correspond to calibration paths involving the December image. These calibrations show greater differences from the calibration line for the direct calibration path than any of the other calibration paths.

Fig. 3(c) and (d) show the calibration lines derived from standard least-squares regressions. It is immediately apparent that there are much greater differences between the least-squares calibration lines than between the robust calibration lines. This is to be expected since a handful of the targets selected are not invariant between some or all of these images.

The difference, in digital counts, between the calibrated pixel values from the direct path and the calibrated pixel values from each multiple-step path was calculated.

These differences are plotted, again for TM Bands 2 and 5, in Fig. 4 for both the robust and least-squares calibration lines. The horizontal line at difference zero in each plot corresponds to the direct path calibration curve differenced from itself. The vertical distance between the other lines and this line corresponds to the difference in the calibrated pixel intensity value between the calibration curves that would be obtained for the input pixel value indicated on the x axis. The dashed vertical lines on the plots correspond to the 5% and 95% points of the cumulative histogram of the June data. The paths involving the December image have been omitted from these plots.

For all of the least-squares calibration lines and for all of the robust calibration lines for Bands 1, 2, and 3, the multiple-step calibration lines all deviate from the direct-path calibration line in the same direction. For the robust calibration for Bands 4, 5, and 7, the multiple-step calibration lines vary either side of the direct-path calibration line.

The maximum of the ranges, in digital counts, of the calibrated intensity values produced by the different cali-

Table 1
The maximum of the ranges of calibrated values, in digital numbers, at the 5% and 95% points of the histogram for the within-scene image sequence

Calibration and path type	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Least squares — all	16.56	14.35	21.38	11.39	22.63	16.81
Least squares — no December	8.65	5.95	9.32	5.81	14.32	12.30
Robust — all	7.15	5.00	9.98	12.21	17.70	11.32
Robust — no December	3.65	3.65	6.62	5.0	11.76	8.52
Robust — eight 'sensible' paths	1.39	1.13	3.04	3.12	7.28	4.49

Table 2
The maximum of the ranges of calibrated values for the within-scenes image sequence expressed as a percentage of the data range between the 5% and 95% points of the histogram

Calibration and path type	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Least squares — all	53.4	57.4	43.6	16.4	22.8	28.0
Least squares — no December	27.9	23.8	19.0	9.5	14.5	20.5
Robust — all	23.1	20.0	20.4	20.0	17.9	18.9
Robust — no December	11.8	14.6	13.5	8.2	11.9	14.2
Robust — eight 'sensible' paths	4.5	4.5	6.2	5.1	7.4	7.5

Table 3
Two times the standard deviation within typical cover types within the September scene

Cover type	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Water	3.8	2.6	3.2	1.8	3.4	1.8
Bush	3.6	3.2	4.8	4.0	7.4	5.4
Crop	3.0	1.8	2.6	6.6	5.2	3.9
Pasture	3.8	2.8	3.0	8.9	5.9	6.4
Bare	11.8	9.4	22.1	3.9	27.4	20.7
Salt-affected	4.4	2.6	6.8	4.8	7.4	7.1

bration lines for input pixel intensities at the 5% and 95% points of the cumulative histogram are listed in Table 1. The calibration paths for Bands 1, 2, 3, and 4 produce similar results, varying by 4–6 digital counts at the extremes. The calibrations for Bands 5 and 7 show twice this variation. It has been observed for most calibrations that the scatterplot for the 'invariant' targets in Bands 5 and 7 are also more variable. In Table 2, the digital count range has been converted to a percentage of the data range between the 5% and the 95% points of the cumulative histogram.

The final row in the tables lists the range of calibrated values for a subset of eight 'sensible' multiple-step paths. These paths, which include two-, three-, and four-step paths, contain few unordered dates and involve comparing images from similar seasons. For these combinations of 'sensible' dates, the calibration paths are within 3 digital counts for TM Bands 1 to 4, and 7.3 and 4.5 digital counts for TM Bands 5 and 7, respectively. Table 3 shows two

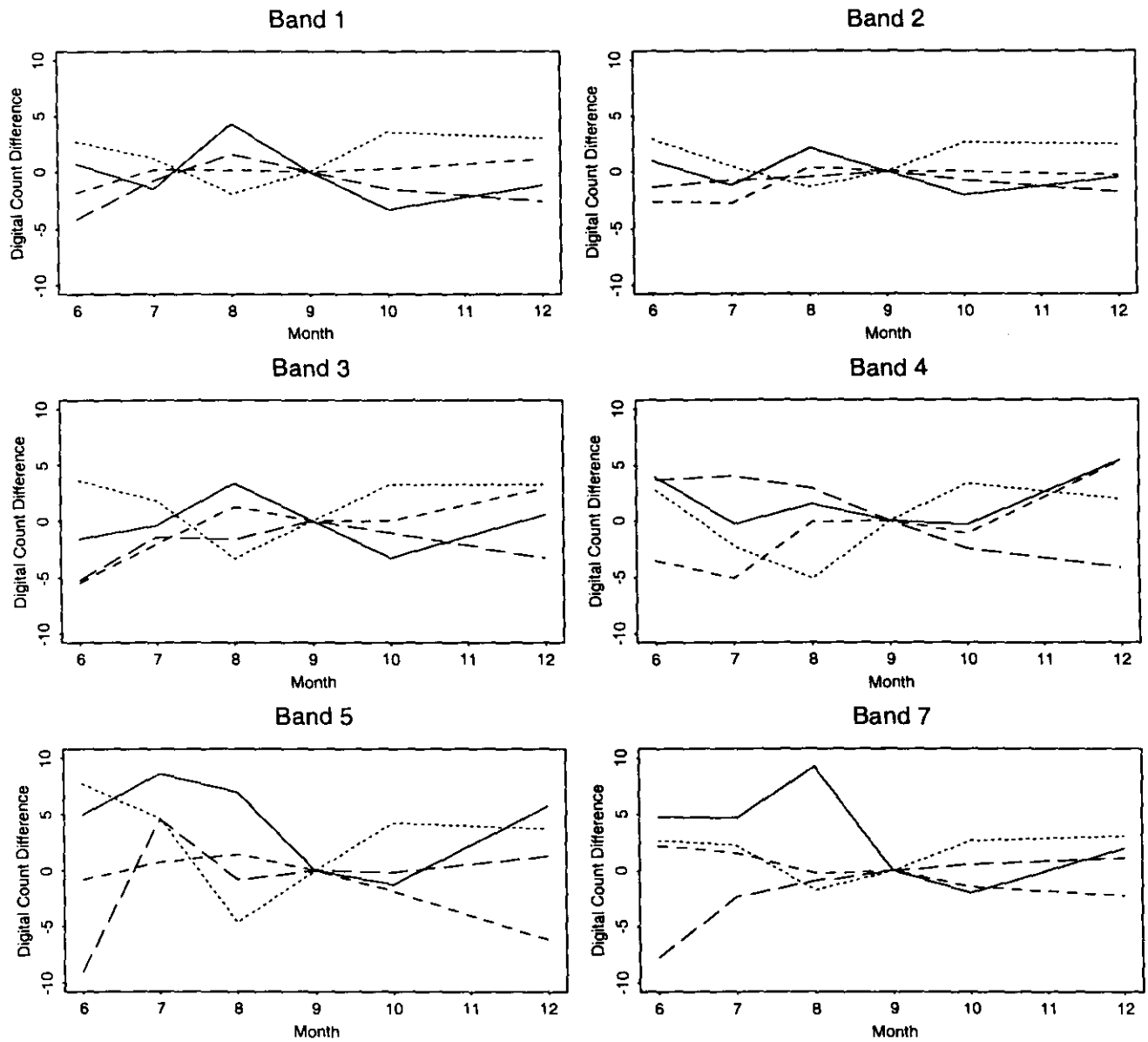


Fig. 5. Calibrated intensity values for four targets (bare soil, gravel airstrip, rock, and water) from June to December. The x axis indicates the month and the y axis is the calibrated intensity for each date subtracted from the intensity in September in units of digital counts from the September image. Solid line: bare soil (catchment area); dotted line: gravel airstrip; short dashed line: water; long dashed line: rock.

times the standard deviation within homogeneous training sites of some of the major cover types in the scene. Ninety percent of the pixels in the site would be expected to have intensity values within two standard deviations of the site mean, giving a measure of the variation within a training site. The data in the table show that the differences in digital counts of the magnitude caused by the calibration procedure are of the same order or less than the variation within training sites. Hence, the 'calibration variation' would have little effect on the results of subsequent processing of the images and the calibrations are considered to be repeatable. An increased calibration "error" in TM Bands 5 and 7, relative to Bands 1–4, is also observed by Hall et al. (1991) and Schott et al. (1988).

As a final comparison of the calibration of the image sequence, each image was calibrated directly to the September reference image using the same invariant targets as in the previous example (one-step path). Four additional (independent) invariant targets were selected (bare soil, a gravel airstrip, a rocky outcrop, and a lake) and the calibrated intensity values compared through time. Fig. 5 shows the sequence of intensity values for the four targets. The x axis indicates the month, and the y axis is the calibrated intensity for each date subtracted from the intensity in September in units of digital counts from the September image. The intensities are presented as a difference from the September digital count in each band so that all the targets can be shown in a single plot. This axis can also be considered to be the calibration error. The variation in the intensity values from June to December is similar to that shown in the repeatability analyses.

5. Conclusion

An approach to scene-to-scene calibration using manually selected invariant targets has been presented. It is robust to changes in up to half of the selected invariant targets, reducing the need for detailed manual comparison of each image pair when more than two images of the same region are to be calibrated to each other. A number of invariant targets are used covering the entire brightness range of the images, so that the assumption of the linearity of the relationship between the image pairs can be assessed during the calibration.

This approach to calibration is being used operationally as part of our data preprocessing in land condition monitoring projects using Landsat TM data from 1987 up to the present. The calibrations allow us to calculate condition indices with associated thresholds that can be applied from year to year without modification, and to use temporal trends in spectral curves as indicators of changing condition (Palmer, Furby, & Wallace, 1994; Wallace, Furby, & Campbell, 1996; Wheaton, Palmer, Wallace, & McFarlane, 1995). Our analyses have shown that the calibrations are acceptably repeatable, at least within a scene.

Our analyses, and our experience of applications on many images, show that TM Band 5 is the most difficult band to calibrate. There is always greater scatter about the calibration line and our observation is that this scatter often increases in cloudy images. This is also observed to a lesser extent in TM Band 7. These are the bands for which the calibration is least repeatable. Certainly, Band 5 has the greatest dynamic range and hence, more subtle variations in reflectance are able to be detected than in other bands. This makes it harder to find targets that are invariant, or change less than the responsivity of the sensor. It is also in Band 5 that data white-outs occur most frequently, limiting the choice of bright invariant targets. We also speculate that variations in atmospheric moisture across the image may cause slight variations in gains across the image. Bands 5 and 7 would be more sensitive to these variations than the visible bands.

By making use of the overlap regions, this calibration procedure can be applied to adjoining images that can be mosaiced to cover a larger region. This has been done successfully for two- and three-image sequences in the southwest agricultural area in Western Australia. When a larger number of images are considered, covering several paths and rows, our analyses have shown that the calibrations across scenes vary greatly and depend on the order in which the calibrations are performed. The variation is sometimes three times more than for the within-scenes comparisons. Other calibration procedures are being investigated, some based on atmospheric models and another on simultaneously optimising the calibration coefficients over a grid of images, to attempt to produce better between-scenes calibrations.

The robust regression calibration procedure has also been applied to Landsat Multispectral Scanner (MSS) scenes with good results. Because of the larger pixel size, 80×60 m instead of 30 m, the targets used for the calibration need to be much larger. The robust calibration procedure can be applied to any system where the sensor and on-ground processing is linear, or where known nonlinearities can be removed from the data. It may not be suitable for very wide field-of-view sensors, such as NOAA Advanced Very High Resolution Radiometer (AVHRR), where big changes in view angle between scenes can cause nonlinear variations in the intensity values.

As mentioned in the Introduction, our ultimate goal is to be able to reliably calibrate image data to standard reflectance units for single images and sequences of images. If we have a reference image that we can calibrate to reflectances, for example, by coincident ground reflectance measurements, the robust procedure allows calibration of a sequence of images to that reference image and hence, to reflectance values as proposed by Hall et al. (1991).

References

Campbell, N. A., Lopuhaä, H. P., & Rousseeuw, P. J. (1998). On the calculation of a robust S-estimator of a covariance matrix. *Statistics in Medicine, 17*, 2685–2695.

Casselles, V., & Garcia, M. J. L. (1989). An alternative simple approach to estimate atmospheric correction in multitemporal studies. *International Journal of Remote Sensing, 10*, 1127–1134.

Chavez, P. S. Jr. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of the Environment, 24*, 459–479.

Chavez, P. S. Jr., & MacKinnon, D. J. (1994). Automatic detection of vegetation changes in the southwestern United States using remotely sensed images. *Photogrammetric Engineering and Remote Sensing, 60*, 571–583.

Dave, J. V. (1978). Extensive data sets of diffuse radiation in realistic atmospheric models with aerosols and common absorbing gases. *Solar Energy, 21*, 361–369.

Fraser, R. S., Ferrare, R. A., Kaufman, Y. J., & Mattoo, S. (1989). Algorithm for atmospheric corrections of aircraft and satellite imagery. *NASA Technical Memorandum, 100751* (106 pp.).

Hall, F. G., Strebel, D. E., Nickeson, J. E., & Goetz, S. J. (1991). Radiometric rectification: toward a common radiometric response among multidate, multisensor images. *Remote Sensing of the Environment, 35*, 11–27.

Kaufman, Y. J. (1988). Atmospheric effect on spectral signature. *IEEE Geoscience in Remote Sensing, 26*, 441–451.

Kniezys, F. X., Shettle, E. P., Gallery, W. O., Chetwynd, J. H., Abreu, L. W., Selby, J. E. A., Clough, S. A., & Fenn, R. W. (1983). *Atmospheric transmittance/radiance: computer code LOWTRAN-6 AFGL-TR-83-0187*. Hanscom AFB, MA: Air Force Geophysics Lab.

Kniezys, F. X., Shettle, E. P., Gallery, W. O., Chetwynd, J. H., Abreu, L. W., Selby, J. E. A., Clough, S. A., & Fenn, R. W. (1988). *Atmospheric transmittance/radiance: computer code LOWTRAN-7 AFGL-TR-88-0177*. Hanscom AFB, MA: Air Force Geophysics Lab.

Moran, M. S., Jackson, R. D., Slater, P. N., & Teillet, P. M. (1992). Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output. *Remote Sensing of the Environment, 41*, 169–184.

Palmer, M. J., Furby, S. L., & Wallace, J. F. (1994). Mapping wind erosion in the South Stirlings. *CSIRO Division of Mathematics and Statistics (Report to LWRRDC Project CMD1)*.

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. New York: Wiley.

Rousseeuw, P. J., & Yohai, V. (1984). Robust regression by means of S-estimators. In: J. Franke, W. Hardle, & R. D. Martin (Eds.), *Robust and nonlinear time series analysis* (pp. 256–272). Springer Verlag, Berlin (Lecture Notes in Statistics No 26).

Schott, J. R., Salvaggio, C., & Volchok, W. J. (1988). Radiometric scene normalisation using pseudoinvariant features. *Remote Sensing of the Environment, 26*, 1–16.

Slater, P. N., Biggar, S. F., Holm, R. G., Jackson, R. D., Mao, Y., Moran, M. S., Palmer, J. M., & Yuan, B. (1987). Reflectance- and radiance-based methods for the inflight absolute calibration of multispectral sensors. *Remote Sensing of the Environment, 22*, 11–37.

Smith, M. O., Ustin, S. L., Adams, J. B., & Gillespie, A. R. (1990). Vegetation in deserts: I. A regional measure of abundance from multispectral images. *Remote Sensing of the Environment, 31*, 1–26.

Tanre, D., Deroo, C., Duhaut, P., Herman, M., & Morcrette, J. J. (1990). Description of computer code to simulate the satellite signal in the solar spectrum: the 5S code. *International Journal of Remote Sensing, 2*, 659–668.

Teillet, P. M., & Fedosejevs, G. (1995). On the dark target approach to atmospheric correction of remotely sensed data. *Canadian Journal of Remote Sensing, 21* (4), 374–387.

Vermote, E., Tanre, D., Deuze, J. L., Herman, M., & Morcrette, J. J. (1994). *Second simulation of the satellite signal in the solar spectrum (6S). 6S User guide version 0.*

Wallace, J. F., Furby, S. L., & Campbell, N. A. (1996). Methods for monitoring the condition of native vegetation. In: *Proceedings of the 8th Australian Remote Sensing Conference, Australasian Remote Sensing and Photogrammetry Association of Australia, Canberra, Australia (CDrom).*

Wheaton, G. A., Palmer, M. J., Wallace, J. F., & McFarlane, D. J. (1995). Waterlogging in the Upper Great Southern 1987–1993. *CSIRO Division of Mathematics and Statistics (Report to LWRRDC Project CMD1).*