Relative Radiometric Normalization
Performance for Change Detection from Multi-Date Satellite Images

Xiaojun Yang and C.P. Lo

Abstract
Relative radiometric normalization (RRN) minimizes radiometric differences among images caused by inconsistencies of acquisition conditions rather than changes in surface reflectance. Five methods of RRN have been applied to 1973, 1983, and 1988 Landsat MSS images of the Atlanta area for evaluating their performance in relation to change detection. These methods include pseudoinvariant features (PIF), radiometric control set (RCS), image regression (IR), no-change set determined from scattergrams (NC), and histogram matching (HM), all requiring the use of a reference-subject image pair. They were compared in terms of their capability to improve visual image quality and statistical robustness. The way in which different RRN methods affect the results of information extraction in change detection was explored. It was found that RRN methods which employed a large sample size to relate targets of subject images to the reference image exhibited a better overall performance, but tended to reduce the dynamic range and coefficient of variation of the images, thus undermining the accuracy of image classification. It was also found that visually and statistically robust RRN methods tended to substantially reduce the magnitude of spectral differences which can be linked to meaningful changes in landscapes. Finally, factors affecting the performance of relative radiometric normalization were identified, which include land-use/land-cover distribution, water-land proportion, topographic relief, similarity between the subject and reference images, and sample size.

Introduction
Spectral data acquired by satellite sensors are influenced by a number of factors, such as atmospheric absorption and scattering, sensor-target-illumination geometry, sensor calibration, and image data processing procedures, which tend to change through time (Teillet, 1986). Targets in multi-date scenes are extremely variable and have been nearly impossible to compare in an automated mode (Kim and Elman, 1990). In order to detect genuine landscape changes as revealed by changes in surface reflectance from multi-date satellite images, it is necessary to carry out radiometric correction. Two approaches to radiometric correction are possible: absolute and relative (Lo and Yang, 1998). The absolute approach requires the use of ground measurements at the time of data acquisition for atmospheric correction and sensor calibration. This is not only costly but also impractical when archival satellite image data are used for change analysis (Hall et al., 1991). The relative approach to radiometric correction, known as relative radiometric normalization (RRN), is preferred because no in situ atmospheric data at the time of satellite overpasses are required. This method involves normalizing or rectifying the intensities or digital numbers (DN) of multi-date images band-by-band to a reference image selected by the analyst. The normalized images would appear as if they were acquired with the same sensor under similar atmospheric and illumination conditions to those of the reference image.

In connection with the NASA-funded project ATLANTA (ATLanta Land-use AAnalysis: Temperature and Air-quality) which necessitates accurately mapping changes in the land-use/land-cover of the Atlanta Metropolitan Region, Georgia, for the past 25 years using Landsat MSS data, the RRN approach is the reasonable choice. Different RRN methods developed for radiometric correction of multi-date satellite images were evaluated to identify the right one for this project. The only published work of a systematic comparison was carried out by Yuan and Elvidge (1996), who applied seven empirical RRN methods to two Landsat MSS images of the Washington, D.C. area. Their comparisons were made visually, using a measure of agreement based on standard error statistics.

Despite the excellent effort of Yuan and Elvidge, the performance of the different RRN methods deserves a more thorough analysis for the following three reasons. First, it is noted that the Landsat MSS images of the Washington, D.C. area used by Yuan and Elvidge contain a high proportion of clear water and urban area, which should favor methods relying on these ground targets in determining the transformation coefficients. In contrast, the Landsat MSS data for the Atlanta region contain a much smaller fraction of water and urban area but a larger proportion of forest and cropland distributed over a foothill and mountainous area. Therefore, these RRN methods may perform differently when applied to images with more diverse geographical features. Understanding the variation of RRN performance with respect to different scene conditions is important for reinforcing the absolute and comparative utility of these methods. Second, the effects of alternative radiometric normalization methods upon the outcomes of change detection are unknown. To bridge this gap, possible impacts of different RRN methods on image classification and spectral change detection will be examined. Finally, it is believed that there are factors that may affect the performance of radiometric normalization other than the RRN methods themselves. By comparing the performance variations of RRN methods as applied to two reference-subject image pairs with different levels of similarity, this
study attempts to uncover these factors and suggest means to handle them from an operational perspective.

Relative Radiometric Normalization (RRN) Methods

In this paper, only those RRN methods using a reference-subject image pair were evaluated. These methods may be further subdivided into three groups: statistical adjustments, histogram matching, and linear regression normalization.

The statistical adjustments approach includes several methods which are based on the linear adjustment of two images to resemble each other in terms of their dynamic range (minimum and maximum DN values), statistical mean and standard deviation, or other possible statistical variables. Two methods—the minimum-maximum normalization and the mean-standard deviation normalization—have been evaluated by Yuan and Elvidge (1996). Their results showed that these methods did not perform well.

Histogram matching is a commonly used radiometric enhancement technique fully integrated in many image processing software packages (Richards, 1986). The transformation is a non-linear matching of histograms of two images so that apparent distributions of brightness values in the two images correspond as closely as possible. The subject image's histogram must be equalized first to obtain an intermediate histogram, which is then modified to match the shape of the reference image's histogram. In essence, histogram match is a process of determining a lookup table that will convert the histogram of one image to resemble that of another. It is, therefore, a useful technique for matching image data of the same scene acquired at different dates with slightly different sun angles or atmospheric effects.

Linear regression normalization includes diverse methods developed over the past decade. One fundamental premise behind these methods is that the radiance reaching an airborne or satellite sensor in a given spectral channel can be expressed as a linear function of reflectivity (Schott et al., 1988). For many sensors, the digital numbers (DN) in each band are a simple linear function of the radiance reaching the sensor. Thus, the atmospheric and calibration differences between scenes are linearly related (Schott et al., 1988; Cassesse and Garcia, 1989; Hall et al., 1991). Accordingly, a linear equation can be used to perform the normalization: i.e.,

\[ S'_k = m_k S_k + b_k \]  

where \( S_k \) is the digital number (DN) of band \( k \) in image \( S \) on date 1, \( S'_k \) is the normalized digital number of band \( k \) on date 1, \( m_k \) is the slope or gain, and \( b_k \) is the intercept or offset. Both \( m_k \) and \( b_k \) are computed through a linear (least-squares) regression performed on two radiometric control sets sampled from the reference and subject images, respectively, or on the whole scene of the image pair. The four major empirical methods used to perform RRN in this way are summarized graphically in Figure 1.

Radiometric normalization using image regression (IR) simply involves relating each pixel of the subject image with that in the reference image band by band to produce a linear equation either through a least-squares regression (Jensen, 1983; Singh, 1989) or a robust regression (Olsson, 1993). The regression technique accounts for the differences in the mean and variance between radiance values for different dates. The calculation of transformation coefficients through a least-squares regression is illustrated in Figure 1A. This method uses all the pixels in the reference-subject image pair.

Radiometric normalization using pseudoinvariant features (PIF) was developed by Schott et al. (1988) and Salvaggio (1993). Pseudoinvariant features are those with nearly invariant reflectivity from one image scene to another. According to Schott et al. (1988), these are typically man-made objects whose reflectance is independent of seasonal or biological cycles. Differences in the brightness distributions of these invariant elements are assumed to be a linear function. Schott et al. (1988) isolated the urban features from the satellite images using an infrared-to-red ratio image, which is quite effective in differentiating urban and water from vegetation. Once the urban features pixels are isolated, linear regression equations are developed to relate the subject images to the reference image band by band (Figure 1B). This method was used by Henebry and Su (1993) to rectify radiometrically eleven TM images of tallgrass prairie spanning a period from 1987 to 1988.

Radiometric normalization using the radiometric control sets (RCS) was developed by Hall et al. (1991). The underlying assumption is that an image always contains at least some pixels that have the same average surface reflectance among images acquired at different dates. This relationship can be best examined by comparing the band-to-band scattergrams in which the pixel values along the diagonal line should have little or no variation through the time period represented by the dates of the two images. These objects serve as a so-called radiometric control set (RCS). Instead of taking a single sample set of bright pixels representing the urban built-up areas, this method uses the two non-vegetated extremes of the Kauth-Thomas (KT) greenness-brightness scattergram which is constructed using the first two bands of a Tasseled Cap transformation. The KT scattergram isolates a dark control set, theoretically corresponding to pixels of deep water (e.g., reservoirs) and a bright control set, representing similar elements of pseudoinvariant features defined by Schott et al. (1988). The transformation coefficients are calculated from the individual band means (in raw digital counts) of the radiometric control sets in each image (Figure 1C). This method performed well for the visible and near-infrared bands of a pair of Landsat 5 TM images, adjusting surface reflectance for the effects of relative atmospheric differences to within 1 percent (Hall et al., 1991). Franklin and Giles (1995) used this method to normalize four different dates of Landsat MSS images for classification purposes in the Wood Buffalo National Park Remote Sensing Project.

Radiometric normalization through no-change set determined from scattergram (NC) was developed by Evdige et al. (1995). The no-change set is obtained from a region identified as no change in the scattergram between a subject image and the reference image band by band. A no-change pixel set contains pixels occupying the core of the water and land data clusters observable in the scattergrams of the near-infrared bands. The centers of the water and land data clusters are located with reference to the local maxima for the near-infrared band scattergrams. The water data cluster center is located at low digital number values near the lower left corner of the scattergrams close to the origin while the center of the land data cluster is located near the center of the scattergram (Figure 1D). The no-change region for a particular band is determined by linking the centers of water and land for that band and expanding perpendicularly to the water land line to some digital numbers on either side of the line. Pixels falling within the no-change region will be used to compute regression lines for all bands. This method works best if the majority of the pixels of the subject scene at one date have the same land cover and vegetation growth stage as the reference scene at another date (Evdige et al., 1995).

Data and Methodology

This study specifically focuses on the RRN methods which have a solid theoretical background and the potential to be operational. Therefore, only five methods, including the histogram matching method (Richards, 1986) and the four linear regression normalization methods mentioned above, will be evaluated.
Figure 1. A graphical summary of four major empirical methods of relative radiometric normalization by developing linear regression equations. (A) Image regression (IR) method with the solutions based on least-squares transformation (Jensen, 1983; Yuan and Elvidge, 1996). (B) Pseudoinvariant feature (PIF) method (Schott et al., 1988). (C) Radiometric control set (RCS) method (Hall et al., 1991). (D) No-change (NC) set determined from scattergram method (Elvidge et al., 1995). HPW = half perpendicular width. HVW = half vertical width.
Plate 1. Two image pairs: 1988 (reference)—1973 (subject) and 1988 (reference)—1983 (subject). The original images and those normalized by different methods are shown. They were displayed as a standard false-color composite (RGB-B421) after applying an identical linear stretch. The image dimension is approximately 140 km by 150 km.
appreciated both visually (Plate 1) and statistically (Figure 2). For the 1973–1988 image pair, the 1973 image has a much duller appearance, particularly for the forest lands which should be red or reddish in a standard false color composite (Plate 1). It has a much narrower dynamic range which can be observed from the scattergrams (Figure 2). Given these observations, the two image pairs can be further distinguished as an image pair with larger differences in terms of scene characteristics (the 1973–1988 image pair) and an image pair with smaller differences (the 1983–1988 image pair). They are, therefore, ideal for use in evaluating the performance variations, the two image pairs can be further distinguished as an image pair with larger differences in terms of scene characteristics (the 1973–1988 image pair) and an image pair with smaller differences (the 1983–1988 image pair). These methods were used for the four experimental procedures that produce the computation of a linear regression equation.

Methods
The experimental procedures include data preprocessing, selection of sample sets, calculation of transformation coefficients, transformation operation, and performance assessment. These procedures were used for the four RRN methods requiring the computation of a linear regression equation.

(1) Preprocessing. The georeferencing strategy used here was to geometrically correct one image, i.e., the 1988 MSS image. Then the corrected 1988 scene was used as the reference image for image-to-image registration. The resultant planimetric accuracy was excellent (±0.5 pixel size).

(2) Selecting Sample Sets. With the exception of the image regression (IR) method which makes use of the whole image scene, the other three methods employ very different strategies to select sample sets. The theoretical background of these strategies has already been explained.

Pseudoinvariant features (PIFs) were segmented by intersecting the mask from the infrared-to-red ratio image (Band 4/Band 2) with the mask from the infrared image (Band 4). These masks were isolated using different threshold values obtained by interactive observations (Table 1). The intersect operation corresponds to an “and” logical operation: i.e.,

$$PIF = \left\{ \frac{\text{Band 4}}{\text{Band 2} \leq t_1} \right\} \text{and} \left\{ \frac{\text{Band 4} \geq t_2}{} \right\}$$

where $t_1$ and $t_2$ are the threshold values. The resultant PIF sample sets are mainly the scene elements covering downtown Atlanta, the Hartsfield International Airport, and large commercial corridors along the Interstate Highways (I75 and I85). The initial radiometric control sets (RCS) were segmented by intersecting the mask from the brightness band with the mask from the greenness band. The two bands were computed using the following equations (Crist and Kauth, 1986):

$$\text{Brightness} = 0.3320 \times \text{Band 1} + 0.6030 \times \text{Band 2} + 0.6750 \times \text{Band 3} + 0.2820 \times \text{Band 4}$$

Figure 2. Scene scattergrams of 1988 (reference image) versus 1973/1983 (subject image) digital number values (DN) band by band. Note that the regions having the densest numbers of pixels are displayed in white, flanked by zones of gray and black, corresponding to declining numbers of pixels. Areas outside or intercepting the data clusters, which have no pixels present, are white.

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING

August 2000 971
The masks were isolated using different threshold values obtained by interactive observations. The equations for segmenting the initial radiometric control sets (RCSs) are:

\[
\text{Greenness} = (-0.2830) \times (\text{Band 1}) + (-0.6600) \times (\text{Band 2}) + 0.5770 \times (\text{Band 3}) + 0.3880 \times (\text{Band 4})
\]

(4)

These masks were isolated using different threshold values obtained by interactive observations. The equations for segmenting the initial radiometric control sets (RCSs) are:

**Initial dark sets**

\[
\text{Initial dark sets} = \{(\text{greenness} \leq t_1) \text{ and } (\text{brightness} \leq t_2)\}
\]

(5)

**Initial bright sets**

\[
\text{Initial bright sets} = \{(\text{greenness} \leq t_1) \text{ and } (\text{brightness} \geq t_2)\}
\]

(6)

where \(t_1\) and \(t_2\) are the threshold values (Table 2). The final dark set was defined by selecting the pixels from the lower half of the initial set because the upper half could contain dark pixels contaminated by sun glint or sediments. The final bright set consists of scene elements covering downtown Atlanta and the Hartsfield International Airport.

The no-change (NC) sets were defined using the following equation (Evidge et al., 1995):

\[
\text{NC} = \{(R_1 - m_{30}S_1 - b_{30}) \leq \text{HVW}_3\} \text{ and } \{(R_4 - m_{40}S_4 - b_{40}) \leq \text{HVW}_4\}
\]

(7)

where \(m_{30}, b_{30}, m_{40}, b_{40}\) are initial estimates for \(m_3, b_3, m_4,\) and \(b_4\) through locating the centers of water-and land-surface data clusters from Band 3 and Band 4 scattergrams; and \(\text{HVW}_3\) and \(\text{HVW}_4\) are the corresponding half vertical widths of the no-change regions in the scattergrams. The initial estimates for \(m_{30}, b_{30}, m_{40},\) and \(b_{40}\) were computed using the equations given by Evidge et al. (1995, p. 1259). The \(\text{HVW}\) and \(\text{HPW}\) have a simple relationship:

\[
\text{HVW} = \sqrt{1 + m^2} \times \text{HPW}
\]

(8)

where \(m\) is the slope of the initially estimated axis for a given band, \(\text{HPW}\) is the half perpendicular width which is assigned as ten digital numbers.

**Results and Discussions**

**Visual Closeness and Statistic Robustness**

Comparing the visual appearance of normalized products is the most straightforward way to judge the overall performance of these methods. In doing so, both the radiometrically normalized images and the reference image are displayed side by side on the monitor screen, and the visual closeness of each normalized image to the reference image is determined qualitatively.

The root-mean-square error (RMSE) is used to measure the statistical agreement of a normalized image with the reference image, as follows:

\[
\text{RMSE}_k = \sqrt{\frac{1}{|\text{scene}|} \sum_{\text{scene}} (S_k - R_k)^2}
\]

(9)

where \(S_k\) is the radiometrically normalized digital number of band \(k\) in the image \(S\) (subject image) on date 1 and \(R_k\) is the digital number of band \(k\) in image \(R\) (reference image) on date 2, and \(|\text{scene}|\) is the total number of pixels of the scene. Thus, the digital numbers of pixels of the radiometrically normalized image are compared with those of the reference image of the same band. If the difference between these numbers is small, the RMSE will be small, implying that the subject image is radiometrically more similar to the reference image.

To facilitate the visual comparison, a high-quality hardcopy in color was produced (Plate 1). Images on these prints were displayed as a standard false color composite after applying an identical linear stretch to each image. For the 1983-1988 image pair, it was found that the images produced by histogram matching (HM) and image regression (IR) were more alike, and gave better results than those generated by other methods. They were followed by images generated by no-change set (NC) determined from the scattergram method, radiometric control set (RCS) method, and pseudoinvariant feature (PIF) method, in this order. The images produced by the HM, IR, NC, and RCS methods exhibited much improved visual closeness to the reference image than that of the 1983 raw image. Only the image produced by the PIF method gave the worst

| Table 1. The PIF Sample Sets with the Threshold Values Determined Through Interactive Observations |
|---|---|---|---|
| Image | \(t_1\) | \(t_2\) | PIF sample size |
| 1973 | 1.00 | 25 | 173,280 |
| 1983 | 0.50 | 30 | 170,275 |
| 1988 | 1.20 | 30 | 171,500 |

*The threshold value (70) was used to exclude the cloud pixels in the 1983 image. Any pixels with a brightness value larger than 163 were considered to be cloud pixels, and should be isolated. Note that both the brightness and greenness were converted into an 8-bit unsigned data format.

| Table 2. The Radiometric Control Sets (RCS) with the Threshold Values Determined through Interactive Observations |
|---|---|---|---|---|---|---|
| Image | \(t_1\) | \(t_2\) | Initial Size | Final Size | \(t_1\) | \(t_2\) | Initial Size | Final Size |
| 1973 | 77 | 41 | 47,052 | 25,992 | 100 | 100 | 44,764 | 23,823 |
| 1983 | 77 | 52 | 61,250 | 37,056 | 77 | 118-163* | 11,025 | 7,044 |
| 1988 | 93 | 27 | 64,925 | 39,813 | 114 | 106 | 20,825 | 11,025 |

*The upper limit of brightness (163) was used to exclude the cloud pixels in the 1983 image. Any pixels with a brightness value larger than 163 were considered to be cloud pixels, and should be isolated.
result because its overall visual appearance was very different from that of the reference image. As for the 1973–1988 image pair, it seemed that none of the five methods had improved substantially the overall closeness of their resultant products to the reference image when compared to the raw image. However, a close examination revealed the following descending order of improvement in visual quality: the HM, IR, PIF, NC, and RCS methods.

Obviously, radiometric normalization of the 1983 image (subject) to the 1988 image (reference) is better than that between 1973 and 1988. This performance variation may stem largely from the scene variations in the two image pairs. Variations in vegetation growth season and land use/land cover are more drastic in the 1973–1988 image pair than those in the 1983–1988 image pair. Also, there is a difference in the quality of the sensors between Landsat-1 and Landsat-5 for the 1973 and 1988 MSS images (see Thome et al., 1997). Clearly, radiometric normalization performs better if the subject and reference images exhibit similar characteristics in vegetation growth season (affected by temperature and rainfall), the spatial distribution of land use/land cover, and similarities in sensor characteristics.

The RMSes computed for each band for the five methods of radiometric correction are shown in Table 5 and Figure 3, which also show the number of pixels used by each method to determine the transformation coefficients for the linear regression equations. For comparison purpose, RMSes of the raw subject image data without any radiometric normalization were also computed for all the four bands.

If RMSes are used to evaluate the different methods of radiometric normalization, it is clear that the IR, HM, and NC methods are better than the RCS and PIF methods. For the 1983–1988 image pair, the IR method is the best with the smallest average RMSE, followed by the NC, HM, RCS, and PIF methods. For the 1973–1988 image pair, the IR method is also the best, followed by the HM, NC, PIF, and RCS methods. All the methods have reduced to varying degrees the radiometric difference between the subject image and the reference image after radiometric normalization, as indicated by the fact that the average RMSEs of each normalized image are much smaller than those for the raw image data. It is also clear that radiometric normalization of the 1983 image to the 1988 image is statistically better than that from the 1973 image to the 1988 image as shown by the fact that the average RMSEs of the normalized images of 1983 are generally smaller than those of 1973. In this respect, the statistical comparison further confirms the visual one.

The statistical comparison also helps to pinpoint the much finer differences among the different methods. For example, by comparing the RMSes band by band, it was revealed that each method performed best for particular spectral bands. The visible wavebands, i.e., Band 1 (0.5 to 0.6 μm green) and Band 2 (0.6 to 0.7 μm red), can be better normalized radiometrically than the two near-infrared bands (Band 3: 0.7 to 0.8 μm, and Band 4: 0.8 to 1.1 μm). Such an observation was also made by Yuan and Elvidge (1996). The reason behind such a variation is not clear. However, this poorer statistical agreement for the two infrared bands may be related to larger variations in vegetation growing conditions (such as season and moisture) between the two images, which are better detected in the infrared portion of the electromagnetic spectrum.

Another important point to note from Table 5 is that the sample size of targets has affected the results of radiometric normalization. The image regression (IR) method and histogram matching (HM) make use of all pixels in the image scene. The no-change (NC) set method also employs over 50 percent of the pixels in the scene. All of these three methods are better overall performers in the evaluation. Both the radiometric control set (RCS) and the pseudoinvariant feature (PIF) methods use a small percent of pixels in the image scene, and their performance suffers because of this feature (Table 5).

### Impacts of RRN Methods on Image Frequency Distribution

Radiometric normalization alters spectral signatures, contrasts between (feature) categories, or variances and covariances of

---

**TABLE 3. The No-Change (NC) Sets with Parameters Defined**

<table>
<thead>
<tr>
<th>Image Pair</th>
<th>Band 3</th>
<th>Band 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1973–1988</td>
<td>$m_{30}$</td>
<td>$b_{30}$</td>
</tr>
</tbody>
</table>

* or 67.24% of the total scene pixels. ** or 60.43% of the total scene pixels.

---

**TABLE 4. Radiometric Normalized Coefficients (Slope-m and Intercept-b)**

<table>
<thead>
<tr>
<th>Method</th>
<th>Band 1</th>
<th>Band 2</th>
<th>Band 3</th>
<th>Band 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m_1$</td>
<td>$b_1$</td>
<td>$m_2$</td>
<td>$b_2$</td>
</tr>
<tr>
<td>PIF</td>
<td>1.275</td>
<td>-28.495</td>
<td>1.119</td>
<td>-12.501</td>
</tr>
<tr>
<td>RCS</td>
<td>1.219</td>
<td>-30.301</td>
<td>1.135</td>
<td>-15.307</td>
</tr>
<tr>
<td>IR</td>
<td>0.442</td>
<td>-1.877</td>
<td>0.425</td>
<td>2.776</td>
</tr>
<tr>
<td>NC</td>
<td>0.379</td>
<td>-0.334</td>
<td>0.272</td>
<td>4.610</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Band 1</th>
<th>Band 2</th>
<th>Band 3</th>
<th>Band 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m_1$</td>
<td>$b_1$</td>
<td>$m_2$</td>
<td>$b_2$</td>
</tr>
<tr>
<td>PIF</td>
<td>1.598</td>
<td>-34.937</td>
<td>1.228</td>
<td>-83.740</td>
</tr>
<tr>
<td>RCS</td>
<td>1.114</td>
<td>-15.375</td>
<td>0.646</td>
<td>-14.925</td>
</tr>
<tr>
<td>IR</td>
<td>0.551</td>
<td>-2.803</td>
<td>0.246</td>
<td>0.400</td>
</tr>
<tr>
<td>NC</td>
<td>0.652</td>
<td>-5.400</td>
<td>0.290</td>
<td>-1.929</td>
</tr>
</tbody>
</table>

Note: PIF—Pseudoinvariant Feature Method; RCS—Radiometric Control Set Method; IR—Image Regression Method; and NC—No-change set determined from Scattergrams method.
### Table 5. Statistics of Sampling Sets and RMS (Root-Mean-Square) Errors

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>%</td>
<td>Number</td>
</tr>
<tr>
<td>Raw</td>
<td>6669836</td>
<td>100</td>
<td>6669836</td>
</tr>
<tr>
<td>PIF</td>
<td>173280</td>
<td>2.60</td>
<td>171500</td>
</tr>
<tr>
<td>RCS</td>
<td>23826(B)</td>
<td>0.36</td>
<td>11025(B)</td>
</tr>
<tr>
<td>IR</td>
<td>6669836</td>
<td>100</td>
<td>6669836</td>
</tr>
<tr>
<td>HM</td>
<td>6669836</td>
<td>100</td>
<td>6669836</td>
</tr>
</tbody>
</table>

Notes: For the Radiometric Control Set (RCS) method, we provide the average numbers of the sample sets; D—the dark control set, B—the bright control set, and T—the total number combining the two sets.

spectral bands, which may further affect the results of image classification and change detection. The way in which different RNR methods alter statistical measures of image data is a fundamental concern for remote sensing analysts in selecting the best method for specific applications. To explore this relationship, the frequency distribution of each normalized image was examined because it can affect the separability of spectral classes in automatic classification. Images with disperse distributions are easier to be clustered than those with compact distributions although this relationship may be complicated by the existence of other statistical phenomena (Arbia et al., 1996). There are a number of different measures which can be used to evaluate the spread of dispersion of a distribution, such as range, quartile, mean deviation, standard deviation, variance, and coefficient of variation (Burt and Barber, 1996). Based on the understanding that most (raw and normalized) images in this study have shown a single-modal distribution of frequency, two measures, namely dynamic range and coefficient of variation, were chosen for comparing the dispersion of each image band by band generated by different RNR methods.

**Dynamic range**, or the difference between the maximum and the minimum digital numbers, is an important feature related to the information content. Data with larger range often contain much finer details that are very helpful in feature identification and information extraction. The alteration of range from a larger one to a smaller one risks a loss of information.

**Coefficient of variation (CV) or the relative variability** of a frequency distribution is measured by the ratio of the standard deviation (σ) to the mean (μ): i.e.,

\[
CV = \frac{\sigma}{\mu}
\]  

Although standard deviation and variance can also be used to judge the dispersion, it is improper to rely on them in comparing multiple distributions with different means (Burt and Barber, 1996), for which coefficient of variation is a better measure. Obviously, the larger the coefficient of variation, the more dispersed is the distribution.

The **dynamic range and coefficient of variation** computed for each band for different RNR methods are shown in Figure 4, which also shows the averages of these measures for four bands as well as the measures for the NDVI images. For comparison purpose, these same measures for the raw images were also computed.

For the 1973 scene, the image produced by the HM method exhibits the largest dynamic range as a whole, followed by the PIF, RCS, NC, and IR methods (Figure 4E). The 1973 NDVI image computed from the normalized image by the PIF method has the largest dynamic range, followed by the RCS, HM, NC, and IR methods (Figure 4K). The 1983 NDVI image computed from the image normalized by the PIF method is also the best in terms of its dynamic ranges, followed by the RCS, HM, IR, and NC methods (Figure 4L). It should be noted that, in both the 1973 and 1983 scenes, the images normalized by the PIF and RCS methods as well as their NDVI images have consistently larger dynamic range than their raw images (Figures 4E, 4F, 4K, and 4L). In terms of dynamic range, the outstanding status of the PIF and RCS methods over all methods tested is consistent for the 1983 images band by band but variable for the 1973 images on a band basis.

As for the coefficient of variation (CV), the PIF and RCF methods are consistently the best among all the five methods for both the normalized images and the NDVI images as a whole in two scenes (Figures 4E, 4F, 4K, and 4L). A very important point to note from these comparisons is that those RNR methods which perform best in the visual and statistical comparisons (such as the IR and NC methods) tend to reduce the dynamic ranges and coefficient of variations of the raw images. These methods must be used with caution because any drop in dynamic range and coefficient of variation will reduce the dispersion of frequency distribution of an image.
Figure 3. Comparisons of Root-Mean-Square Error (RMSE) for the raw images and their normalized products by different RRN methods. The vertical axis is RMSE. Raw, raw image; NC, no-change set determined from scattergram method; RCS, radiometric control set method; HM, histogram matching; PIF, pseudoinvariant feature method; and IR, image regression method with the solutions based on least-squares transformation.
Figure 4. Comparisons of dynamic range and coefficient of variation (CV) for the raw data and their products normalized by different methods. Note that in all except F and L, the CV is multiplied by 100. Raw, raw image; NC, no-change set determined from scattergram method; RCS, radiometric control set method; HM, histogram matching; PIF, pseudoinvariant feature method; and IR, image regression method with the solutions based on least-squares transformation.
band, which might adversely affect the separability of spectral classes in image interpretation and classification.

Impacts of RRN Methods on Magnitude of Change

Radiometric normalization minimizes the impacts caused by acquisition conditions (such as sensor variations, atmospheric effects, and viewing geometry) other than the effects due to changes in vegetation growth seasons and land use/land cover. In other words, regardless of the relative normalization methods employed, the normalized images should still exhibit certain levels of differences that are related to surface reflectance change. In this sense, an RRN method that dramatically reduces the numeric difference between two images may result in suppressing the magnitude of change related to vegetation growth seasons and land use/land cover. In actual applications, ignoring this aspect may produce incorrect results for change detection. Thus, how different RRN methods will affect the change magnitude needs to be investigated. To explore this relationship, a simple change detection algorithm, i.e., image subtraction, was applied to the data: i.e.,

\[ RS_k = R_k - S_k \]  

where \( RS_k \) is the digital number of band \( k \) in the output image, \( R_k \) is the digital number of band \( k \) in the reference image, and \( S_k \) is the digital number of band \( k \) in a normalized image or subject image. Then, the global means of these output images were compared band by band to determine the magnitude of change. Clearly, a smaller mean indicates a lower magnitude of change between two images.

The mean of digital number (DN) for the images produced by a simple subtraction of a normalized image from the 1988 image is shown in Figure 5. For comparison purposes, the DN mean for the outputs produced by the subtraction of a subject image from the reference image was also computed. In Figures 5E and 5F, the average of the means of four bands were used as an absolute value.

Based on Figure 5, it is clear that the difference in digital number between the 1988 image and the images produced by the RRN methods with good visual and statistical characteristics tends to be smaller (Figures 5E and 5K). This difference is inversely proportional to the RMSE as discussed before. However, these two comparisons are quite different in nature. The RMSE is used as a measure of goodness of fit in judging the statistical robustness for a specific method. Image differencing is a standard algorithm used in change detection analysis (Jensen, 1996).

These comparisons revealed that the HM, IR, and NC methods tend to suppress a larger part of change between the two images. This trend is extremely polarized for the 1988-1983 image pair in which a 96 to 99 percent change in digital number has been suppressed by these three methods. In contrast, the RCS and PIF methods have suppressed a 30 to 44 percent change for the 1988-1973 image pair and a 62 to 66 percent for the 1988-1983 pair, leaving more room for change-detection analysis related to phenomena such as land-cover changes.

Thus, these RRN methods which tend to substantially suppress the magnitude of spectral changes between two images must be used with discretion in change analysis.

Operational Perspective

Operational perspective is another important concern in selecting an RRN method. Both computation intensity and ease of implementation of each method can be used to assess this aspect. Computation intensity should be measured in terms of the amount of processing time needed. However, these methods were generally implemented in an interactive mode; thus, it was impossible to measure the time in an absolute way. Also, with advances in computation power, the issue of computation intensity is of limited meaning and, thus, was not considered here. The assessment of operational perspective focused on the ease of implementation for each RRN method.

As can be seen from the previous sections, the major bottleneck to implementing a specific method is its procedure in segmenting sample sets. Because both the histogram matching (HM) and image regression (IR) methods do not require this procedure, they are more easily implemented. Indeed, many image processing software packages have the HM method as a built-in function as a radiometric enhancement enhancement technique. The other three methods (PIF, RCS, and NC) are generally more difficult to implement. The PIF method is easier to be executed than the RCS method because the PIF method only requires a set of segmented bright pixels while the RCS method needs two sets of samples. The NC method is the most difficult to implement because it involves a much more complicated procedure for selecting the no-change set.

Factors Affecting Radiometric Normalization

The results of the RRN performance evaluation undertaken in this research suggest the following considerations in the choice of an RRN method for change analysis: reference image, sample set(s), sample size, and scene characteristics.

Reference Image

In RRN, the role of the reference image is critical because all of the other original (subject) images are to be adjusted to it. A reference image should have better visual quality than other images in the same data set as well as a larger dynamic range so that the information can be maintained as much as possible (Lo and Yang, 1998).

In selecting a reference image, priority should be given to those which have good ground reference data available within a close time span (Jensen et al., 1995). These reference data can be large-scale aerial photographs or ground observations and measurements to help in the identification of radiometric control sets used by different RRN methods. For a large amount of image data covering a long time span, it is preferable to select the scene closest to the middle of the time sequence instead of the most recent scene as the reference image. In doing so, the difference in land use/land cover between the subject image and the reference image is minimized, thus maximizing the similarity between the two images.

Sample Sets

Sample sets are also called radiometric normalization target sets. They contain the actual scene elements through which the transformation coefficients are derived. There are three types of sample sets used by different methods: pseudoinvariant feature (PIF) sets, (dark-bright) radiometric control sets (RCS), and no-change (NC) sets. The majority of the RRN methods employ a classic thresholding technique through a histogram or a two-dimensional scatterplot to segment these sample sets. While this technique has certain strengths, it also has a number of weaknesses. First, this technique is generally not easy to implement because it involves many complicated procedures. It also requires intensive computation. These are not quite desirable from an operational point of view. Second, this technique only accounts for the targets' statistical behaviors, but ignores other important criteria, such as the targets' dimension, physical properties, and environmental settings, which are crucial for ensuring purer scene elements to be segmented.

According to Eckhardt et al. (1990), the radiometric normalization targets should be at approximately the same elevation as the other features within the scene so that the atmospheric condition estimated from these targets should be identical to that over other parts of the scene. Additionally, these targets should be in a relatively flat area so that incremental
Figure 5. Comparisons of DN means for the outputs after applying a simple change detection algorithm, namely, image subtraction. In E and K, the average of means of four bands in absolute value was used. Raw, raw image; NC, no-change set determined from scattergram method; RCS, radiometric control set method; HM, histogram matching; PIF, pseudoinvariant feature method; and IR, image regression method with the solutions based on least-squares transformation.
changes in sun angle from date to date will have the same proportional increase or decrease in direct beam sunlight for all normalization targets. For defining the dark radiometric control targets, the dimension of the water body may be important. Targets with small dimensions may suffer from background contamination because the apparent reflectance of water targets increases considerably as a result of a signal component of the total solar flux reflected by the targets and transmitted to the sensor by atmospheric scattering (Hill and Sturm, 1991). For defining the pseudo-invariant features (PIF)/bright radiometric control targets, one has to consider not only the statistical behavior of the samples but also their physical properties, such as the surface moisture and surface reflectance characteristics. Major changes in surface moisture will change the overall reflectivity of the features that were assumed to be invariant (Schott et al., 1988). Features with a non-Lambertian surface, such as a city’s downtown, may also cause problem in defining the PIF samples. To compensate for these issues, it is recommended that the statistically based thresholding method be used jointly with skill-based visual observations. The latter can be used to select those apparently suitable targets object by object through an on-screen digitizer. In doing so, the selection procedures can be substantially streamlined while excellent results are being maintained.

The no-change (NC) set was merely defined through a statistical approach (Evdige et al., 1995). Locating the water center may be tedious when a scene does not have enough water targets. This may affect the estimation of the initial transformation coefficients and, hence, the sample characteristics. Defining the width of the no-change region is also a tricky issue, because it affects not only sample size but also sample characteristics.

Sample Size
The size of the sample set has a significant impact on the results of radiometric normalization. A larger sample size often gives a better overall performance of an RRN method, both visually and statistically. However, it does not mean that the radiometrically normalized products can be used to yield more accurate change-detection results. In addition, a larger sample set may have more statistical outliers caused by noise in the database, which may substantially degrade the performance of radiometric normalization. For these considerations, the RRN methods of image regression (IR) and histogram matching (HM) should be used with special care. In particular, the HM method may alter the relative distribution of the land use/land cover because of the non-linear nature of its transformation. This will affect the subsequent image classification and change detection.

Scene Characteristics
Scene characteristics that may have a profound impact on radiometric normalization include land-use/land-cover distribution, proportion of water and land, and topographic relief. Land-use/land-cover distribution determines the nature of the sample set used to derive the transformation coefficients. The proportion of water and land may affect the effectiveness of different RRN methods. A higher proportion of these features means that ideal sample sets with large size and excellent quality can be selected using appropriate techniques. It should favor those methods relying on these spectral reference targets, such as the radiometric control set (RCS) and the no-change (NC) set methods. The research conducted by Yuan and Evdige (1996) used a pair of Washington, D.C. Landsat MSS scenes which contained a very high proportion of clear water and urban area (35 to 50 percent of the total image area). Their evaluation indicated that the NC and the RCS methods were better than all the other methods including the image regression (IR) method, pseudoinvariant feature (PIF) method, and several other lesser methods such as haze correction, minimum-maximum normalization, and mean-standard deviation normalization. When applied to the Landsat MSS scene of Atlanta in our evaluation, the NC and the RCS methods did not give the best results. The Atlanta scene has few water features (less than 2 percent), the majority of which are rivers and reservoirs. The urban use occupies 15 to 25 percent of the total image area, of which the core urban area is only 2 to 5 percent. Radiometric normalization generally performs better for a scene covering flat areas than for hilly areas because scene variations due to topographic relief (shadows) are more complicated, which often prevent an effective implementation of an RRN method. On the whole, radiometric normalization performs well for the subject and reference images that are similar in terms of scene characteristics.

Conclusions
The results of the performance evaluation of five different RRN methods, as applied to the multi-date Landsat MSS data of the Atlanta region, have shown that all five methods can reduce radiometric differences between the subject image and the reference image to varying degrees. Those RRN methods using a large number of sample pixels in deriving the transformation coefficients for the linear regression equations (such as the IR and NC methods) were the best overall performers. But they tend to shrink dynamic range and coefficient of variation of the raw images, reducing the dispersion of frequency distribution of an image band, which could affect adversely the separability of spectral classes in image interpretation and classification. In contrast, both the PIF and RCS methods increased the two measures consistently for both the normalized images and the NDVI images as a whole in two scenes. Those visually and statistically robust RRN methods (such as the HM, IR, and NC methods) tended to substantially reduce the magnitude of spectral change which might be related to meaningful changes in landscapes. In contrast, the RCS and PIF methods cut down only a moderate degree of spectral change between the two scenes, thus favoring better change detection.

Radiometric normalization of the 1983–1988 image pair is better than that of the 1973–1988 image pair, leading to the observation that the choice of RRN method should be based on the nature of the images, notably land-use/land-cover distribution, water-land proportion, topographical relief, and similarity between the subject and reference images. Another important consideration is the ease with which each method can be implemented in a conventional image processing platform.

Finally, to ensure the proper implementation of an RRN method in a practical application, some operational guidelines for selecting the reference image and defining the sample sets are given in this paper. A good reference image should display the best visual quality with the largest dynamic range. Preferably, ground data, atmospheric condition data, and sensor calibration data for the image should be available. The reference image should not be too far away in time from the subject image. For defining the sample sets, one should consider not only the targets’ statistical behaviors, but also their dimension, physical properties, and environmental settings, to ensure that only pure scene elements are segmented.

Acknowledgment
The research reported in this paper has been supported by a NASA EOS Interdisciplinary Science (IDS) research grant for Project ATLANTA.

References


---

Certification Seals & Stamps

SEND COMPLETED FORM WITH YOUR PAYMENT TO:
ASPRS Certification Seals & Stamps, 540 Grosvenor Lane, Suite 210, Bethesda, MD 20814-260

NAME: ____________________________ PHONE: ____________________________

CERTIFICATION #: ____________________________ EXPIRATION DATE: ____________________________

ADDRESS: ____________________________ CITY: ____________________________ STATE: ________ POSTAL CODE: ________ COUNTRY: ________

CREDIT CARD ACCOUNT NUMBER: ____________________________ EXPIRES: ____________________________

SIGNATURE: ____________________________ DATE: ____________________________

THE ACADEMY OF PHOTOGRAMMETRY & REMOTE SENSING

Photogrammetric Engineering & Remote Sensing