LANDSAT 7 SCAN LINE CORRECTOR-OFF GAP-FILLED PRODUCT DEVELOPMENT

James Storey, Principal Systems Engineer*
Pasquale Scaramuzza, Senior Scientist*
Gail Schmidt, Software Project Lead*
Science Applications International Corporation (SAIC)
U.S. Geological Survey (USGS)
Earth Resources Observation and Science (EROS) Center
Sioux Falls, SD 57198
James.C.Storey.1@gsfc.nasa.gov
pscar@usgs.gov
gschmidt@usgs.gov
* Work performed under USGS contract 03CRCN0001

Julia Barsi, Calibration Analyst
Science Systems and Applications, Inc.
Code 614.4
NASA Goddard Space Flight Center
Greenbelt, MD 20771
julia.barsi@gsfc.nasa.gov

ABSTRACT

The Landsat 7 Enhanced Thematic Mapper Plus (ETM+) scan line corrector (SLC) failed on May 31, 2003, causing the scanning pattern to exhibit wedge-shaped scan-to-scan gaps. The ETM+ has continued to acquire data with the SLC powered off, leading to images that are missing approximately 22 percent of the normal scene area. To improve the utility of the SLC-off data, the U.S. Geological Survey (USGS) developed new products that use the data from multiple ETM+ scenes to provide complete ground coverage. These gap-filled products were developed and deployed in two phases. The gaps in the Phase I products are filled with data from imagery collected previously with a functional SLC (SLC-on). A single SLC-on scene provides complete coverage of the scan gaps, making the gap-filling procedure straightforward. Several radiometric adjustment techniques for matching the SLC-on fill scene to the SLC-off primary scene were evaluated for performance, processing speed, and ease of implementation. A simple local histogram matching method was adopted as a result of this evaluation. The Phase II products use data from multiple SLC-off scenes to fill the scan gaps with more recent data. Because the locations of the scan gaps are different for each SLC-off scene, the gap-filling process must account for scan gap interactions. The Phase II product development included a more comprehensive study of candidate radiometric adjustment techniques. This study showed that the histogram matching method used in Phase I, with minor refinements, provided the best overall performance and was adopted for Phase II as well.

INTRODUCTION

On May 31, 2003 the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) experienced a failure in its scan line corrector (SLC) mechanism. When operating nominally, the SLC deflects the ETM+ line-of-sight in the along-track direction to ensure that the sensor's bi-directional cross-track scanning pattern provides continuous coverage of the full Landsat swath [Landsat Project Science Office, 1998]. Without this compensation, the resulting zigzag scanning pattern exhibits wedge-shaped scan-to-scan gaps, alternating with scan-to-scan overlap, of increasing magnitude away from nadir. Figure 1 shows an example of the impact of the SLC failure on ETM+ imagery.

The ETM+ has continued to acquire data with the SLC powered off, leading to images that are missing approximately 22 percent of the normal scene area but are otherwise of the same radiometric and geometric quality as images collected prior to the failure. The USGS Earth Resources Observation and Science (EROS) Center developed the infrastructure to implement a production capability for multi-scene (same path/row) gap-filled products in an effort to improve the usability of ETM+ data acquired after the SLC failure. These gap-filled products

were intended to benefit applications that are more sensitive to the presence of the data gaps than to the need for same-day coverage. Mapping applications are thus likely to be better served by these products than are agricultural applications that analyze within-growing season vegetation changes.

Figure 1. Landsat 7 ETM+ images of the San Francisco Bay area acquired before and after the SLC failure.

This paper describes the methodology used to develop the SLC-off gap-filled products, the processing flow for gap-filled product generation, the radiometric adjustment algorithms used to combine data from multiple scenes when filling the scan gap areas, and plans for future improvements to the initial gap-filled product generation methods.

GAP-FILLED PRODUCT DEVELOPMENT PROCESS

The ETM+ SLC was declared inoperable in September 2003, after all attempts to restore its proper functioning had failed, and routine ETM+ imaging operations were resumed with the SLC powered off. In parallel with the SLC recovery efforts during the summer of 2003, software development activities were underway to adapt the Landsat ground processing systems to accommodate ETM+ data acquired with the SLC powered off. Initial SLC-off products were made available to the public in November 2003. These products provided basic Level 1 radiometric and geometric correction for SLC-off data, including limited scan gap interpolation capabilities. A product ordering option that allows the user to define the width of gap interpolation to allow was subsequently added based upon feedback from data users.

The USGS Landsat Project and NASA Landsat Project Science Office jointly held a Landsat 7 SLC-off product improvement workshop in October 2003 [Beck, 2003]. The workshop identified methods to enhance the use of SLC-off data and new potential SLC-off data products. Scientists from the former Landsat 7 Science Team and technical representatives from USGS and NASA participated. One of the recommendations of this working group was that the USGS and NASA pursue the development of new composite data products that combine an SLC-off image with one or more SLC-on or SLC-off images to produce a current Landsat 7 scene that eliminates the scan gaps present in a single SLC-off scene.

Following this recommendation, a joint USGS/NASA study team was formed to develop these so called gap-filled products. The development activity was divided into two overlapping phases so that an initial demonstration product could be quickly deployed while more comprehensive algorithm testing and refinement proceeded in parallel. The focus of the Phase I activity was to develop a simple, easily implemented radiometric adjustment

technique that would allow the scan gaps in an SLC-off scene to be seamlessly filled using data from a gap-free SLC-on scene, acquired during the first four years of the mission. With the emphasis on rapid development, only straightforward, easily coded, or commercially available algorithms were considered in this phase. The algorithms considered will be described below. The Phase I development activity concluded with the successful release of SLC-off to SLC-on gap-filled products in May 2004.

While the Phase I development activity was still underway, a more formal radiometric adjustment trade study was initiated as part of the Phase II development effort. In addition to providing a more thorough examination of the candidate radiometric adjustment algorithms, the Phase II activity sought to address merging data from multiple SLC-off scenes. The interaction among the scan gaps in the SLC-off scenes makes this a more complex problem than the SLC-off to SLC-on gap-fill logic developed for Phase I. The Phase II development came to a successful conclusion with the release of multi-scene SLC-off gap-filled products in November 2004.

An independent Science Advisory Panel oversaw the progress of both product development phases. The Panel membership was drawn largely from the participants in the original product improvement workshop, including representatives from USGS, NASA, academia, and large user organizations such as the U.S. Department of Agriculture (USDA). The development group made regular technical presentations to the Science Advisory Panel, and the algorithms selected for eventual implementation were strongly influenced by Advisory Panel suggestions.

Numerous sample products were created during the development process to demonstrate the usability of the new gap-filled products. In particular, a set of "control" test products were generated from SLC-on data by simulating the presence of scan gaps and then applying the gap-fill algorithm. These simulated gap-filled products could then be compared directly to the original SLC-on data to provide a quantitative evaluation of the gap-fill algorithm performance. These products were also supplied to eight teams of application scientists from the USGS Earth Resources Observation and Science (EROS) Center who were asked to use both the original SLC-on and simulated gap-filled products in their particular applications and to compare the results. The application areas that were tested included burn severity mapping, crop type mapping, geologic mapping, land sustainability, burn area mapping, canopy cover and impervious surface mapping, land cover change, and land fire detection. These tests showed that, while the usability of the gap-filled products varied with the input scene characteristics and application requirements, results comparable to the original data could be achieved for some applications. The results of some of these studies have appeared in the literature [Howard and Lacasse, 2004]. These results were sufficiently encouraging to proceed with the development and deployment of gap-filled products.

PHASE I SLC-OFF TO SLC-ON GAP-FILLED PRODUCTS

The Phase I development included two major tasks: 1) selecting the initial radiometric adjustment algorithm used to match the SLC-on "fill" scene data to the SLC-off "primary" scene and 2) implementing software to apply the selected adjustment algorithm to provide replacement values for the pixels in the SLC-off scene scan gaps and create a gap-filled output product. The radiometric adjustment algorithm testing and selection process was carried out in two steps. First, three variants of the familiar histogram matching algorithm [Gonzalez and Wintz, 1977] were compared to identify the most effective implementation for this application. The histogram matching method was then compared to the regression tree algorithm [Quinlan, 1993], as implemented in the commercial CUBIST software product. In each comparison, the accuracy of each method was evaluated by applying the radiometric adjustment technique to fill (SLC-on) scene pixels that had corresponding primary (SLC-off) scene values to serve as a basis for comparison. In addition to considering the accuracy of the gap-fill pixel values, the criteria for selecting an algorithm for implementation included operational considerations such as expected processing burden and ease of algorithm implementation and integration into the Landsat ground processing system.

Radiometric Adjustment Algorithm Selection

Having simplicity and ease of implementation as primary requirements, the Phase I radiometric adjustment algorithm selection was immediately limited to two classes of algorithms: histogram matching and regression tree. These two approaches were chosen based upon the ready availability of commercial or prototype implementations for testing and the familiarity of the development team with these methods. Three variants of the histogram matching method, which adjusts the fill image pixels to match the image statistics of the primary image, were compared first:

1. Global linear—Compute a gain and bias adjustment using image statistics for all pixels in the image.
2. Global nonlinear—Compute piecewise adjustment using histogram information for all pixels in the image.

3. Local linear—Compute a gain and bias adjustment using image statistics from a local neighborhood around the scan gap pixel being filled.

The six ETM+ reflective multispectral bands were tested in these comparisons. The histogram matching adjustment was applied independently to each spectral band. The panchromatic and thermal bands were not used during the initial testing to avoid the complexities involved in dealing with images at different spatial resolutions.

The estimated errors for each of the three methods were calculated by applying the histogram matching technique to the entire SLC-on scene, not just the pixels that fell in the SLC-off scan gap, and computing the difference between the adjusted SLC-on data and the existing data in the SLC-off scene. The results for three test scenes, selected to exhibit a variety of image content (including clouds), are shown in Table 1. The Pxxx/Ryyy designations in the table are the Worldwide Reference System (WRS) path and row coordinates of the test sites.

Table 1. RMS difference between adjusted SLC-on pixel values and actual SLC-off pixel values in units of DN.

	Site	P036/R037			P018/R038			P039R028		
	Method	Global Linear	Global Nonlinear	Local Linear	Global Linear	Global Nonlinear	Local Linear	Global Linear	Global Nonlinear	Local Linear
Band	1	7.89	5.25	4.77	6.74	6.97	3.08	48.34	46.44	5.46
	2	6.65	5.21	4.23	6.79	6.83	3.45	51.21	47.76	5.53
	3	11.36	9.55	7.81	10.38	10.35	5.83	47.81	44.14	6.35
	4	6.10	4.37	3.86	8.54	8.84	5.35	38.07	38.41	6.06
	5	12.04	9.28	8.28	14.47	14.63	9.39	29.51	28.34	6.57
	7	10.55	8.99	7.36	13.36	13.38	8.60	29.33	28.04	6.11

For all three test scenes the local linear method did the best job of reconstructing the primary scene pixel values from the fill scene data. Sample windows from the output gap-filled images for the three test scenes, shown in Figure 2, demonstrate how the histogram matching methods react to various scene content. Note that all three methods have problems with clouds.

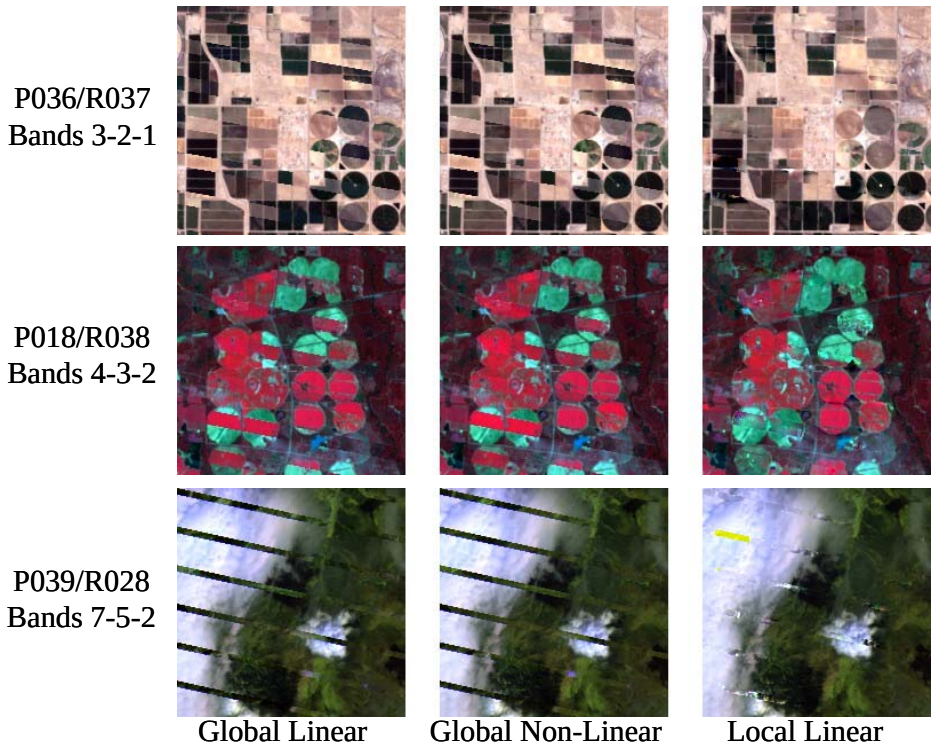

Figure 2. Gap-fill test images for three test sites generated using three variants of the histogram matching algorithm.

Having selected the local linear method as the best representative of the histogram matching class of algorithms, a second round of comparisons against the regression tree method was planned. The regression tree method uses the information from all six spectral bands of the fill scene to derive a set of rules to predict the corresponding value for each primary scene band. This method uses the pixels that have values in both input images to derive the rules, which are then applied to the primary scene, scan gap pixels to compute gap-fill values. Several hundred common pixels are generally required to "train" the algorithm (i.e., derive rules) so this method is somewhat less "local" than the histogram matching approach.

Difficulties in implementing and running the prototype regression tree test software limited the amount of testing that could be done with that method in the time available for Phase I development. Although the tests that were run yielded good results, the regression tree implementation available at the time was too immature to recommend including it in the Phase I operational deployment. The development team therefore recommended the local linear histogram algorithm as the preferred method for Phase I implementation to the Science Advisory Panel. The regression tree approach was recommended for inclusion in the Phase II algorithm trade study where, with some additional development, it had the potential to produce better results than the local linear histogram method.

Radiometric Adjustment Algorithm Description

The local linear histogram matching radiometric adjustment algorithm was implemented in the Phase I gap-filled product generation process. This algorithm requires co-registered image products for both the primary (SLC-off) and the fill (SLC-on) scene. The Phase I methodology iterates the following procedure over all gap pixels in the primary scene:

1. Extract an n x n pixel window about the gap pixel in both the primary and the fill data. The window width was defined as 19 for the 30-meter bands, 11 for the 60-meter thermal band, and 31 for the 15-meter pan band. These sizes were chosen to ensure that the window around a pixel in the center of a scan gap spans the gap. The maximum gap width is 14 pixels in the 30-meter bands, 7 pixels in the thermal band, and 28 pixels in the pan band.
2. Exclude from the common pixels any high or low saturated data. This step masks out the gap pixels as well as any high-saturated clouds in both the primary and fill data sets. Excluding clouds has been shown to significantly improve the gain calculation. After exclusions are made, the common pixels are counted. If there are fewer than a minimum number of common pixels (currently defined as 2), then the fill data value is used to fill the gap pixel without modification.
3. Calculate a gain and bias using the mean and standard deviation of the common pixels:

$$gain = \frac{\sigma_P}{\sigma_F} \tag{1}$$

$$bias = \mu_P - \mu_F * gain \tag{2}$$

where:

$$\mu_P = \frac{1}{N}\sum_{i=1}^{N} primaryDN_i \quad \text{the primary image window mean} \tag{3}$$

$$\sigma_P = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(primaryDN_i - \mu_P\right)^2 \quad \text{the primary image standard deviation} \tag{4}$$

$$\mu_F = \frac{1}{N}\sum_{i=1}^{N} fillDN_i \quad \text{the fill image window mean} \tag{5}$$

$$\sigma_F = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(fillDN_i - \mu_F\right)^2 \quad \text{the fill image standard deviation} \tag{6}$$

No constraints were placed on the computed gain and bias values in Phase I gap-fill processing.

4. Calculate the new value for the fill pixel as a 16-bit number.

$$mergedDN = fillDN * gain + bias \tag{7}$$

Check that the filled value is within the unsigned 8-bit scale of 0 to 255. If it is less than 0, assign the pixel in the filled scene a value of 0. If it is greater than 255, then assign a value of 255. Otherwise, cast the filled value as an 8-bit number and assign the pixel filled scene the calculated value.

A scan gap mask that identifies the source of each pixel in the product accompanies each gap-filled product. For Phase I products, this gap mask simply identifies those pixels that were non-zero in the original SLC-off image. Any other pixels that are populated in the gap-filled product were derived from the SLC-on fill scene.

From May 2004 through October 2004, gap-filled products were made available from the USGS using this simple histogram matching algorithm. The initial Phase I algorithm allowed only a single SLC-on fill scene, selected from the ETM+ data acquired prior to the SLC failure. Following the release of the multi-scene SLC-off to SLC-off capability in November 2004, all gap-filled products, even those that use only a single SLC-on fill scene, are created using the Phase II processing.

PHASE II SLC-OFF TO SLC-OFF GAP-FILLED PRODUCTS

Work began on the Phase II multi-scene SLC-off gap-filled product development while Phase I was still underway. The first task of the Phase II activity was to perform a more thorough radiometric adjustment algorithm trade study to identify and test alternatives to the local linear histogram matching algorithm employed in Phase I. This trade study is described in more detail below. The Phase II development also addressed the additional complexity introduced by the geometric interactions among the scan gaps from multiple SLC-off scenes.

The random placement of the scan gaps from one acquisition to the next leads to variations in the amount of gap-fill and, conversely, common data that is available when combining the data from a pair of SLC-off scenes. If the primary and fill scene scan gaps do not overlap, then a complete gap-filled product can be created using only two SLC-off scenes. In this case, however, there will be only a few pixels of overlapping common data near the scene edges with which to train the radiometric matching algorithm, making it more difficult to achieve a good radiometric match. This effect had a significant influence on the design of the Phase II radiometric matching algorithm. On the other hand, if the primary and fill scene scan gaps fall in nearly the same place, the radiometric matching problem is simplified but relatively little gap-fill is provided so a larger number of fill scenes will be required.

The importance of this scan gap overlap geometry motivated the development of new SLC-off scene metadata that predicts the scan gap location. This information is used in the gap-filled product ordering process to predict the degree of gap coverage provided by the user-selected set of fill scenes. Although the gap location predictions are only estimates, they help users decide how many fill scenes to request to provide full scan gap coverage.

Radiometric Adjustment Algorithm Trade Study

The Phase II gap-fill algorithm trade study, conducted jointly by USGS EROS and the Biospheric Sciences Branch at NASA Goddard Space Flight Center (GSFC), began with a review of journal articles on image normalization techniques, including previously conducted comparative reviews [Yang and Lo, 2000; Yuan and Elvidge, 1996]. From this literature review, four algorithms were selected for prototype implementation and testing against a variety of scene types:

1. An iterative normalization technique developed at the Canada Centre for Remote Sensing [Du, 2001]
2. The radiometric rectification method developed at NASA GSFC [Hall, 1991]
3. The regression tree technique investigated in Phase I
4. The local linear histogram method implemented in Phase I

The iterative normalization method performs a linear regression using the common image pixels and iteratively prunes away the points furthest from the regression line until the correlation in the remaining data exceeds a specified threshold. The final regression is then applied to the fill scene data to provide gap-fill values. This method does a good job of compensating for effects such as atmospheric variability, solar geometry differences, and, to some extent, phenology, but the iterative pruning tends to reject pixels that exhibit real temporal change.

The radiometric rectification method transforms the image data to Kauth-Thomas (tasseled cap) space and identifies bright and dark invariant (nonvegetated) targets along the soil line. This transformation is applied to both images, and the corresponding invariant targets are then fit using a linear transformation. Since this method intentionally excludes vegetation from the derivation of the transformation, it tends to preserve scene-to-scene phenological differences. It was also found to be sensitive to cloud contamination.

SLC-on data containing simulated scan gaps were used to compute quantitative gap-fill accuracy for each of the four candidate methods and for the "control" gap-fill method of applying no radiometric adjustment. Figure 3 shows a scatter plot, derived from a representative test scene, that plots the gap-fill error in the local linear histogram test on the X axis and the corresponding gap-fill error using the other methods on the Y axis. A one-to-one line is shown for reference. Note that only the regression tree method is consistently at or below the one-to-one line representing the performance of the local linear histogram method.

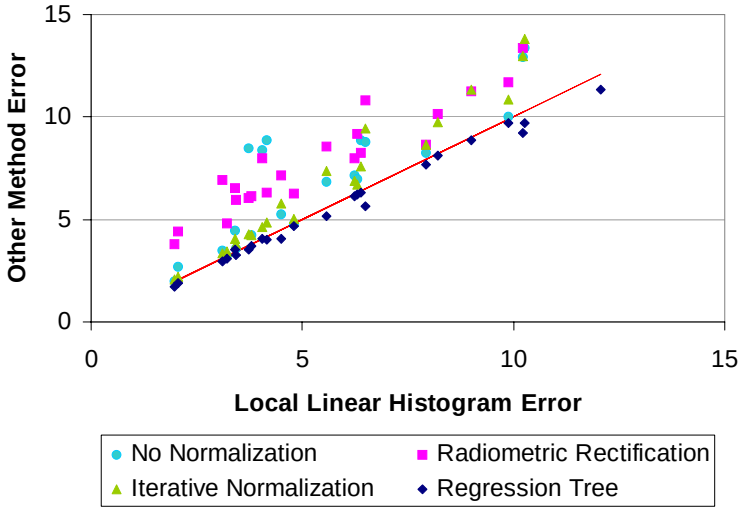

Figure 3. Gap-fill accuracy of candidate methods relative to the Phase I local linear histogram method.

Although the local linear histogram algorithm used for Phase I generally provided good performance, it proved to be sensitive to the scan gap overlap geometry. When the primary and fill scene scan gaps are adjacent, the use of a fixed local window size provides an insufficient number of common points to compute reliable radiometric adjustment coefficients. This can lead to "flare" artifacts at the edge of the scan gap. The iterative normalization algorithm, which uses a larger training window, is resistant to this effect. A comparison of these two methods in an image with adjacent scan gaps is shown in Figure 4.

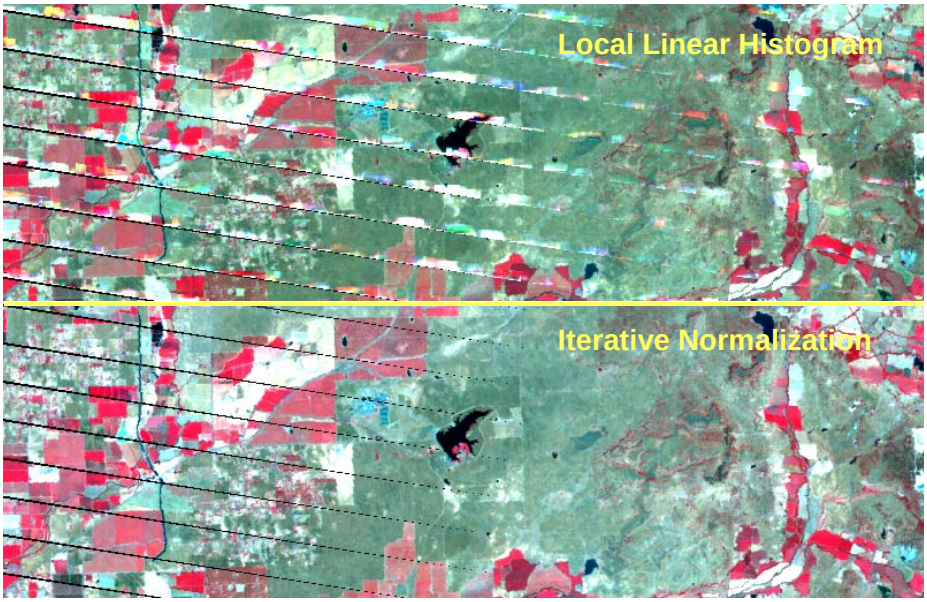

Figure 4. Adjacent scan gap flare artifacts in the local linear histogram method (top) are suppressed by the iterative normalization method (bottom).

Using a larger window for the local linear histogram improved but did not eliminate the “flare” artifact. Following the advice of the Science Advisory Panel, a hybrid approach was developed to combine the best features of the local linear histogram and iterative normalization methods. Like the iterative normalization approach, this

new adaptive local linear histogram method uses a linear regression to compute the radiometric adjustment gain and bias, although it does not perform the iterative outlier pruning. In order to maintain the local temporal change reversal ability of the original local linear algorithm, it uses a variable window size, adaptively selected to provide the smallest window that guarantees a minimum number of common training points. This new hybrid local linear histogram method and the regression tree method were the final two candidate algorithms considered in the trade study.

The quantitative and qualitative comparisons of gap-fill algorithm performance gave the regression tree approach a slight advantage. In addition to generally providing smaller gap-fill errors, the regression tree method was less sensitive to cloud contamination. Nevertheless, operational considerations once again tilted the algorithm selection to the (adaptive) local linear histogram method. The extra processing burden imposed by the regression tree algorithm and the relative ease of implementing the straightforward local linear histogram method as compared to integrating the complex, commercial software implementation of the regression tree outweighed its slight gap-fill performance advantage.

Image Selection and Registration

All of the scenes being integrated into a single gap-filled product must be geometrically registered prior to the radiometric adjustment and merging (gap-fill) step. The initial approach to the image registration problem was to treat the primary SLC-off scene as the geometric reference. Image chips were extracted from a systematically corrected Level 1 image (L1Gs) of the primary scene for use as ground control points in the Level 1 precision correction (L1Gp) processing of the SLC-on fill scene [Landsat Project Science Office, 1998]. Experience with the Phase I SLC-off to SLC-on image registration process demonstrated that the control point chip correlation process works better with fully populated, i.e., gapless, image chips. Because of this, the Phase I flow was redesigned to use the SLC-on fill image as the geometric reference and, hence, source of control point chips. This provides another motivation (in addition to guaranteeing complete gap-fill) for recommending that an SLC-on scene be included at the end of the fill scene list in the multi-scene SLC-off gap-fill process. The SLC-on scene can be used as the geometrical reference, providing the benefits of gap-free control point chips.

If an SLC-on scene is available, it is used as the reference for control point extraction as noted above. If no SLC-on scene is provided, then the primary SLC-off scene is used as the reference. The gap-filled product ordering interface was designed to make it easier for users to decide which scene to use as the reference in the work order setup process. The ordering interface provides three sets of scene ID parameters: a primary scene ID, a list of fill scene IDs, and a geometric reference scene ID. The primary scene ID is populated first when the user selects the initial SLC-off scene. SLC-off fill scenes are then selected and added to the fill scene list. Finally, the user has the option to select a "background" SLC-on scene. If an SLC-on scene is selected, its ID is added at the end of the fill scene list, and it is used as the geometric reference scene. If the user does not select an SLC-on scene, a cloud-free scene is automatically selected for use as a geometric reference, but this scene is not included in the gap-fill process.

Data processing proceeds with the geometric reference scene. The input unprocessed Level 0 data (L0Rp) are processed to a Band 8 (panchromatic) Level 1G systematic north-up Universal Transverse Mercator (UTM) image from which control point chips are extracted. These control points are used to process the primary scene and each of the fill scenes to Level 1G precision using the user-defined processing parameters. The control point correlation segment of the Level 1G precision process uses the mutual information (MI) correlation technique [Maes, 1997].

Phase II Gap-Fill Processing Algorithm

The adaptive local linear histogram matching method adopted for Phase II is a modified version of the Phase I algorithm. The most significant differences between the Phase I and II algorithms are the use of an adaptive, rather than fixed size, image window to ensure a minimum number of common training points and the computation of the radiometric adjustment gain and bias through linear regression (following the iterative normalization algorithm), rather than the image window mean and standard deviation statistics.

The Phase II method for creating a histogram-matched multi-scene SLC-off composite is to iterate the following over all gap pixels in the primary scene:

1. Extract an n x n window about the gap pixel in both the primary and fill data. The Phase I algorithm extracted an n x n window, with the intent of using the entire window. The Phase II algorithm uses the smallest possible chip within the n x n window (Step 2); however, in order to change the code as little as possible, the read logic is maintained from Phase I. The window dimension, n, has been defined within the code as 31 to ensure that the window will completely span the worst-case gap overlap geometry caused by adjacent scan gaps. Larger windows were found to yield essentially identical results while imposing a processing time performance penalty.

2. Find the smallest square of pixels between the two chips that contains at least a minimum number of common pixels. Common pixels are those that are valid data in both the primary and fill scenes. A mask is created for both the primary and fill chip, distinguishing between valid data and invalid data (low-saturated data such as gaps or data dropouts and high-saturated data such as clouds). Beginning with the smallest square (1 x 1 pixel), the number of valid common pixels between the primary mask and the fill mask is counted. If the minimum number of common pixels is not met, the square is expanded by one pixel on each side (i.e., 3 x 3 to 5 x 5 to 7 x 7…). The common pixels are counted again. Iterate until the minimum number of common pixels is met. If the minimum number of pixels is not met before reaching the maximum extent of the window, all available common pixels are used, regardless of the total.

The minimum number of common pixels has been defined within the code as 144, so that the valid data content will be equivalent to the original (Phase I) 17 x 17 fixed window with 50 percent valid data. A new mask is generated with only the common pixels of the chips denoted (for use in Step 3).

3. Using the common pixel mask to extract only the common pixels from the chip, regress the primary chip against the fill chip to find the transformation between the two scenes. This regression computes the least squares solution for the following over-determined system of linear equations:

$$\begin{bmatrix} 1 & fillDN_1 \\ 1 & fillDN_2 \\ & \vdots \ \ \vdots \\ 1 & fillDN_N \end{bmatrix} \begin{bmatrix} bias \\ gain \end{bmatrix} = \begin{bmatrix} primeDN_1 \\ primeDN_2 \\ \vdots \\ primeDN_N \end{bmatrix} \tag{8}$$

In these equations $fillDN_i$ denotes the fill scene pixel value for pixel i, and $primeDN_i$ denotes the corresponding primary scene pixel value for pixel i. Each of the N pixels in the image window that have valid values in both the primary and fill scenes contributes one observation to the least squares solution.

The least squares solution is given by:

$$\begin{bmatrix} bias \\ gain \end{bmatrix} = \begin{bmatrix} N & \sum_{i=1}^{N} fillDN_i \\ \sum_{i=1}^{N} fillDN_i & \sum_{i=1}^{N} (fillDN_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{N} primeDN_i \\ \sum_{i=1}^{N} (fillDN_i)(primeDN_i) \end{bmatrix} \tag{9}$$

Check for a “reasonable” gain to prevent outliers from having a strong effect on the transformation. If the calculated gain is greater than the maximum allowable gain (*Mgain*) or less than the inverse (1/*Mgain*) of the maximum allowable gain, then calculate gain and bias using the Phase I mean and standard deviation equations shown in Equations (1) and (2) above.

If this gain is also greater than the maximum allowable gain (or less than its inverse), then set:

$$\begin{aligned} gain &= 1 \\ bias &= \mu_P - \mu_F \end{aligned} \tag{10}$$

No reasonable gain can be calculated, so perform a bias-only adjustment to the fill scene. The maximum allowable gain is a processing parameter, currently set at 3.0, making the allowable gain range $1/3 < gain < 3$.

4. Calculate the new value for the fill pixel as an 8-bit number using Equation (7) above. Check that the filled value does not overflow the 8-bit scale. If it does, assign the pixel in the filled scene a value of 255. Otherwise, cast the filled value as an 8-bit number and assign the pixel-filled scene the calculated value.

Complete this procedure, shown as a process flow in Figure 5, for each pixel with a value of 0 in the primary scene. Note that this algorithm will also replace some normal “fill” data around the edges of the scene, where the fill scene coverage is shifted relative to the primary scene, thereby expanding the extent of the primary scene. If there are gaps remaining in the merged scene, treat the merged scene as the “primary,” proceed to the second fill scene in the list of available scenes, and repeat the process.

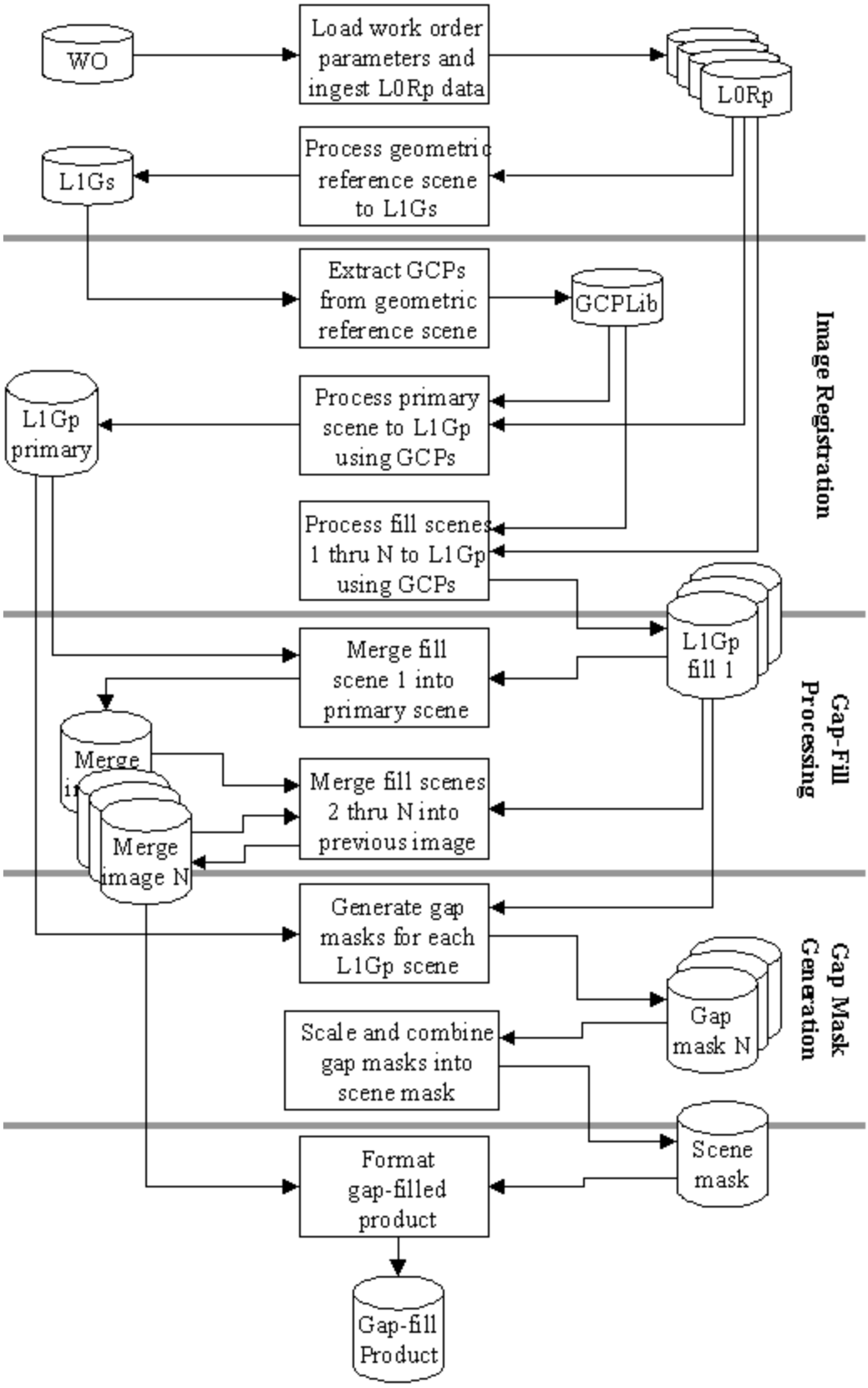

Figure 5. Multi-scene SLC-off gap-filled product processing flow.

Gap Mask Generation

A gap mask accompanies each gap-filled product to identify the source of each image pixel in the output product. For the Phase I SLC-off to SLC-on gap-filled products, it was sufficient to show those pixels taken from the SLC-off "primary" scene. All other non-fill pixels in the output product could be presumed to originate from the SLC-on "fill" scene. The conventional SLC-off gap mask was able to provide the necessary information since it identified the primary scene pixels with 1s and all other pixels with 0s. The gap masks require somewhat more sophisticated processing in the case of products created using multiple fill scenes.

For the Phase II multi-scene SLC-off gap-filled products the gap mask must be able to identify up to six source images (one primary image plus up to five fill images—limited by the number of fill image dates that can be listed in the gap-filled product metadata). Instead of containing only 0s and 1s, the Phase II masks will contain numbers ranging from 0 to 6, with the meanings shown in Table 2.

Table 2. Gap mask codes.

Value	Meaning
0	No Data
1	Primary Scene
2	Fill Scene #1
3	Fill Scene #2
4	Fill Scene #3
5	Fill Scene #4
6	Fill Scene #5

The method implemented to generate the gap mask is as follows:

1. Generate gap masks for each input scene using the normal SLC-off gap mask generation method (i.e., resample the image a second time with the output DN range clipped to 0 to 1).
2. Multiply the gap masks for the fill scenes by the appropriate scene code from Table 1. For example, the mask for Fill Scene 3 would be multiplied by 4.
3. Merge the fill scene masks with the primary scene mask, clipping off any imagery falling outside the primary scene frame, in the order in which the scenes were merged:
    a. Merge the Fill Scene 1 mask with the primary scene mask by replacing zeros in the primary mask with the value from the fill mask.
    b. Merge the Fill Scene 2 mask with the mask resulting from Step 3a by replacing zeros in the product mask with the value from the fill mask.
    c. Repeat Step 3b until all fill scene masks have been incorporated.
4. Repeat the above procedure for each spectral band and format the resulting image as a GeoTIFF product. This product should geometrically match the gap-filled image product.
5. Use the GNU zip (gzip) utility to compress the mask files.

This approach has the disadvantage of requiring a second resampling operation on each input image. This was deemed preferable to developing a new application to create gap masks directly from the product image during the SLC-off product development.

Known Algorithm Limitations

The adaptive local linear histogram adjustment algorithm can yield poor results if the scenes being combined exhibit radical differences in target radiance due, for example, to the presence of clouds, snow, or sun glint. Although attempts were made to include a cloud detection and masking capability in the Phase II gap-fill process, the problems associated with identifying thin clouds and cloud edges as well as cloud shadows proved too difficult to overcome in the available time.

Clouds present in the primary SLC-off scene tend to pull nearby fill scene data to high bias levels so that the fill scene data appear "cloud-like" even if the scene is, in fact, cloud free. This creates a superficially reasonable looking gap-filled image but discards any benefit to be gained from the cloud-free fill scene data in the primary scene cloud areas. Fill scene clouds also cause problems for the algorithm. In addition to inserting cloudy "wedges" into otherwise cloud-free primary scene imagery, where fill scene clouds overlap valid primary scene data, the gain and bias computations are corrupted for nearby pixels. When these bright fill scene pixels are compared to the

corresponding darker primary scene pixels, the resulting gain and bias adjustments will tend to reduce the fill scene image DN value. This leads to dark areas in the fill regions in the neighborhood of fill scene clouds.

Similar effects can be expected wherever large changes in scene content and radiance occur, such as changes in snow cover or specular reflection off water due to changes in solar illumination geometry.

Although the gap-fill algorithm can perform well when "adjusting" for temporal changes in homogenous regions such as agricultural fields, this algorithm has difficulty when the size of the regions exhibiting this change becomes too small. Fill data are typically better matched in the large agricultural fields typical of the American Midwest and West than in the smaller fields of the Northeast and, presumably, much of Europe and Asia. In general, features smaller than the local window size (nominally 19 x 19 pixels) are difficult to model if significant change has occurred.

FUTURE GAP-FILLED PRODUCT IMPROVEMENTS

The first Phase II multi-scene SLC-off gap-filled products were released to the public in November 2004. Since then, much experience, both good and bad, has been gained with these products under varying scene content and gap geometry conditions. Work continues to improve these products based on this experience. One area of active study is developing improvements to the gap-fill radiometric adjustment algorithm. A second study area involves cloud and cloud shadow identification and, if possible, filling cloud and cloud shadow areas

Refinements to the local linear histogram method are being developed to address shortcomings identified in problem scenes. Much of this work seeks to provide more graceful handling of cases of real scene-to-scene temporal change. Work also continues on the regression tree algorithm, particularly in the context of cloud filling, where its ability to yield good results over larger areas would be beneficial.

Single and multi-scene cloud identification techniques are also being developed. At a minimum, excluding clouds from the radiometric adjustment process should improve the performance of the gap-fill algorithm. Ideally, primary scene cloud and cloud shadow areas could be identified and filled with cloud-free fill scene data, using the techniques developed for scan gap-fill, to provide cloud- and scan gap-free imagery. Early experiments using the Landsat 7 single scene automated cloud cover assessment (ACCA) algorithm [Irish, 2000] to identify and exclude clouds produced disappointing results. The ACCA method's inability to detect thin clouds, cloud edges, and cloud shadows undermined most of the benefits of cloud masking. Using the single scene ACCA approach in conjunction with cross-scene comparisons for detecting clouds and shadows has shown more promise. A study is currently underway to determine whether cloud detection and replacement is feasible, with results expected in late 2005.

REFERENCES

Beck, R. (2003). Scan Line Corrector-off Products Available. Landsat Project News October/November 2003. http://landsat.usgs.gov/project_news/October_November_2003.php

Du, Y., J. Cihlar, J. Beaubien, and R. Latifovic (2001). Radiometric normalization, compositing, and quality control for satellite high resolution image mosaics over large areas. *IEEE Transactions on Geoscience and Remote Sensing, 39*(3):623–634.

Gonzalez, R. C., and P. A. Wintz (1987). *Digital Image Processing.* Addison-Wesley Publishing Company, Reading, MA, pp. 152–158.

Hall, F. G., D. E. Strebel, J. E. Nickeson, and S. J. Goetz (1991). Radiometric rectification - Toward a common radiometric response among multidate, multisensor images. *Remote Sensing of Environment, 35*(1):11–27.

Howard, S. M. and J. M. Lacasse (2004). An evaluation of gap-filled Landsat SLC-off imagery for wildland fire burn severity mapping. *Photogrammetric Engineering and Remote Sensing, 70*(8):877–880.

Irish, R. (2000), Landsat 7 Automatic Cloud Cover Assessment. In: *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery.* SPIE Vol. 4049, pp. 348–355.

Landsat Project Science Office (1998). *Landsat 7 Science Data Users Handbook.* Updated January 7, 2005. http;//ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html

Maes, F., A. Collignon, D. Vandermeulen, et al. (1997). Multimodality image registration by maximization of mutual information, *IEEE Transactions on Medical Imaging, 16*(2):187–198.

Quinlan, J. R. (1993). Combining instance-based and model-based learning. In: *Proceedings of the Tenth International Conference on Machine Learning*. Morgan Kaufmann, San Mateo, CA.

Yang, X., and C. P. Lo (2000). Relative radiometric normalization performance for change detection from multi-date satellite images. *Photogrammetric Engineering and Remote Sensing, 66*(8):967–980.

Yuan, D., and C. D. Elvidge (1996). Comparison of relative radiometric normalization techniques. *ISPRS Journal of Photogrammetry and Remote Sensing, 51*(3):117–126.