Landsat 7 SLC-Off Gap-filling for Interim Data Continuity in Northern Regions using a Robust Radiometric Normalisation Technique

Jonathan Orazietti[1] Ian Olthof[2] Robert Fraser[2]

[1] Noetix Research Inc, Natural Resources Canada, 588 Booth St, Ottawa ON, K1A 0Y7
Tel: (613) 947-1326 Fax: (613) 947-1385, e-mail: Jonathan.Orazietti@CCRS.NRCan.gc.ca
[2] Canada Centre for Remote Sensing, Natural Resources Canada, Ottawa ON

ABSTRACT

Much of the current climate change research in Canada involves monitoring northern environments, which require a continuous source of data for change detection. Unfortunately, data continuity of Landsat 7 imagery has been interrupted due to the recent failure of the Scan Line Corrector (SLC), which compensates for the forward motion of the satellite. Although Landsat 7 data are still being acquired, each scene has gaps of no-data values due to the SLC failure. We have created a methodology to produce a consistent source of data for medium-resolution northern biomass study and change detection until a suitable replacement for Landsat 7 is launched. The relatively short growing season and low sun angle in the north provide few opportunities for acquisition of usable earth observation data, and coupled with the failure of the SLC, further reduces the amount of usable data per scene. Fortunately, the large overlap between adjacent orbits in northern Landsat scenes provides a potentially large sample to use for gap infilling. We propose normalization of scenes using overlap regions, semi-invariant targets and a robust regression technique called Thiel-Sen prior to infilling gaps in SLC-Off imagery using the normalised adjacent scene. The new technique enhances the Level 1G (L1G) products provided by the USGS and is modification to gap-filling procedures in the data flow of the Gap-Fill Algorithm currently in use by the USGS. The resulting imagery is much more suitable for biomass modelling using vegetation indices such as the Reduced Simple Ratio (RSR) and Normalised Difference Vegetation Index (NDVI). The current methodology shows great potential for maintaining data continuity of Landsat 7 imagery for northern environmental monitoring procedures requiring the use of vegetation indices.

Keywords: Landsat 7, SLC-Off, Vegetation Indices, preprocessing.

1 INTRODUCTION

Much of the current climate change research in Canada involves monitoring northern environments, which require a continuous and consistent source of data for biophysical modelling and change detection. The data continuity of Landsat 7 imagery has been episodic since the recent failure of the Scan Line Corrector (SLC), which compensates for the forward motion of the satellite. The SLC failure causes gaps of no-data values in the imagery that are more pronounced along the edges. Image gap infilling is one option to aid in the creation of a consistent source of data during the transition to other sensors such as SPOT 5 or until the launch of another Landsat-type sensor. Another option to maintain data continuity is to use imagery from the Landsat 5 Sensor. Unfortunately, Landsat 5 has no redundancy left in its prime components and the fuel will be depleted in 2008. Within the present methodology, data from adjacent Landsat 7 scenes are used to infill the gaps of no-data values in the scene of interest.

Image normalisation techniques vary from those using simple regression, as in reference [1] to complex and robust approaches as in reference [2]. Reference [3] compared existing methods to two new methods for normalisation of Landsat imagery over tree line transition and northern areas. Both new methods used Theil-Sen regression [4], [5], to normalise Landsat scenes to either adjacent scenes using overlap regions, or separately to coarse resolution (1km) SPOT VEGETATION (SPOT VGT) composite data. In comparing the two new methods, the authors concluded that although normalisation to coarse resolution data avoided error propagation that can occur when overlap regions are used for normalizing large mosaics consisting of many individual scenes, Theil-Sen normalisation using overlap regions may be an appropriate option when the goal is to normalize adjacent scenes to

Presented at the 26th Canadian Symposium on Remote Sensing in Wolfville, Nova Scotia, June 14-16, 2005

each other. To assess infill data consistency, the current study compares fill imagery from an adjacent Landsat 7 image normalised using the Theil-Sen and least squares (OLS) regression equations to the original data using simulated SLC-off imagery. This allows the comparison of the normalised imagery from adjacent scenes to the removed original imagery. After gap infilling, data consistency is assessed to determine the potential to use the resultant data for biophysical modelling through the use of vegetation indices such as the Reduced Simple Ratio (RSR) and Normalised Difference Vegetation Index (NDVI).

1.1 Study Area

The study area covers a section of the Dempster Highway in the Yukon Territory in Canada. The centres of the two scenes used in the study are 136d07’17.00W, 66d56’23.44N for the scene from 2000/08/31, path 63, row 13 (63/13) and 137d34’14.20W, 66d56’43.10N for the scene 1999/08/04, path 64, row 13 (64/13). The vegetation in the study area ranges from open coniferous forest through shrub and herb dominated tundra to bare rock in the mountains. The scenes were selected to represent the variation in vegetation that covers most of the transition zone in the north.

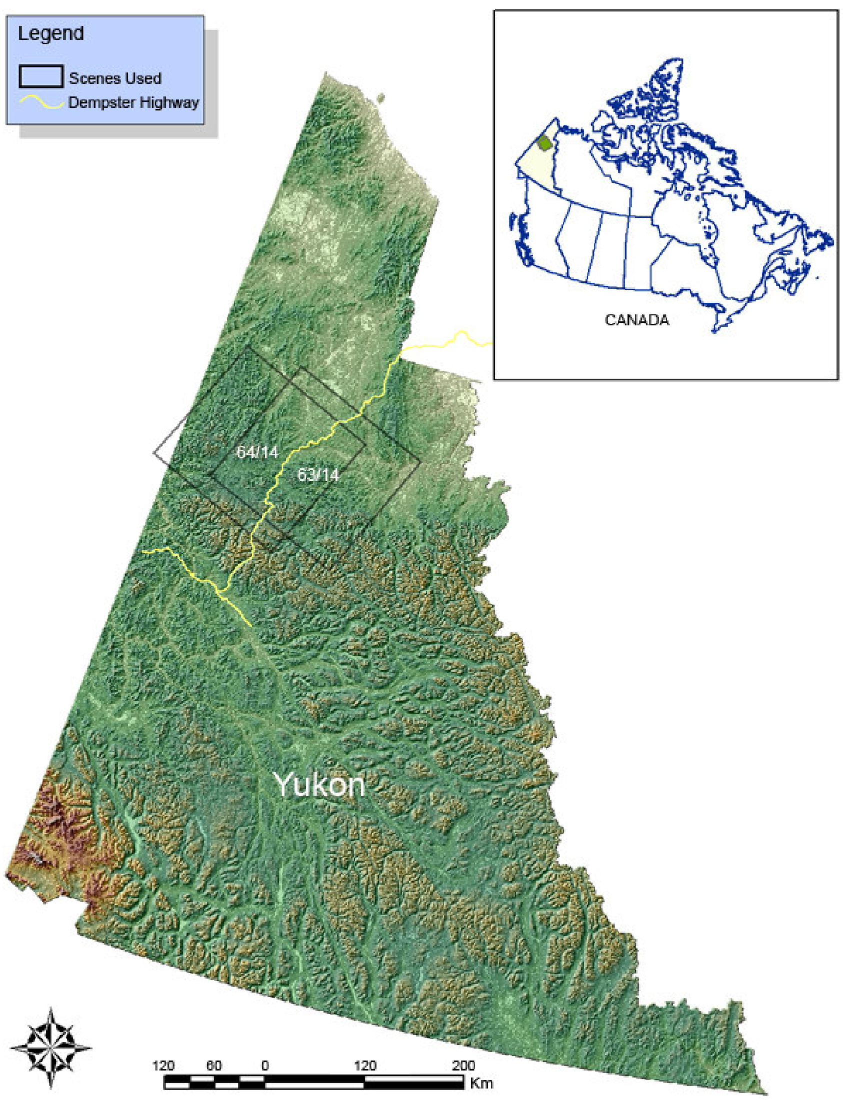

Figure 1. Scenes used for SLC-off comparison

2 METHODS

2.1 Simulated SLC-off imagery

More than one additional scene may be necessary to create a gap-filled data product. For the present study, we only looked at the gap filling procedure for one half of the scene, as the methodology remains consistent regardless of the number of input scenes. If the gap-filled product covers more than the area of four Landsat scenes, normalisation to SPOT VGT is an option because not all ground features in the gap filled area will be present in the Landsat scene containing the normalisation targets [3].

SLC-off data were simulated using pre-SLC-off data from circa 2000 because consistency between the original and fill imagery was being measured. If true SLC-off imagery were used assessing the quality of the product would be quite difficult. The current method allows comparison between the original data and the data used to fill the simulated SLC-off data gaps, therby giving a measure to the quality of the gap-filled product.

To facilitate the creation of the simulated products, two sets of adjacent Landsat scenes were collected. One set of complete scenes in the north was used to simulate and assess gap-infilling, while the other set of SLC-off images was used to establish stripes resulting from SLC failure. Striping and striping overlap between adjacent scenes determines sampling for inter-stripe radiometric normalization and assessment. Northern scenes consisted of Landsat ETM+ bands 3,4 and 5 (Red, NIR and SWIR) from path/row 63/13 and 64/13 from August 4th 1999 and August 31 2000 respectively. Both scenes are quite clear of haze and are close enough together in the season to represent SLC-off scenes in consecutive orbits. Previous studies have shown that the variation between Landsat scenes acquired a year apart but on a similar date seasonally can be smaller than images acquired in the same season, but months apart [6]. We therefore chose scenes acquired a year apart and removed temporally variant targets such as fire regeneration. If successful, the methodology will be easily transferable to recent Landsat scenes that are acquired on subsequent passes.

Adjacent SLC-off Landsat scenes from Québec were used to properly estimate the scan line effect within scenes and between adjacent images. Band 4 (NIR) of the two Quebec Landsat ETM+ scenes acquired 25 days apart in September 2003 were calculated to stripe and no stripe values of 0 and 255 respectively and registered image-to-image to the set of northern Landsat scenes. The bounding coordinates of the resultant images matched the northern scenes accurately representing the extent and overlap of SLC-off stripes on the images. The stripe layers were transferred to the northern scenes as masks to simulate the SLC-off error. The original northern images could then be separated into overlapping areas without stripes to create a sample for normalisation and areas of striping for an assessment of consistency between the original imagery and the areas to be infilled with adjacent imagery.

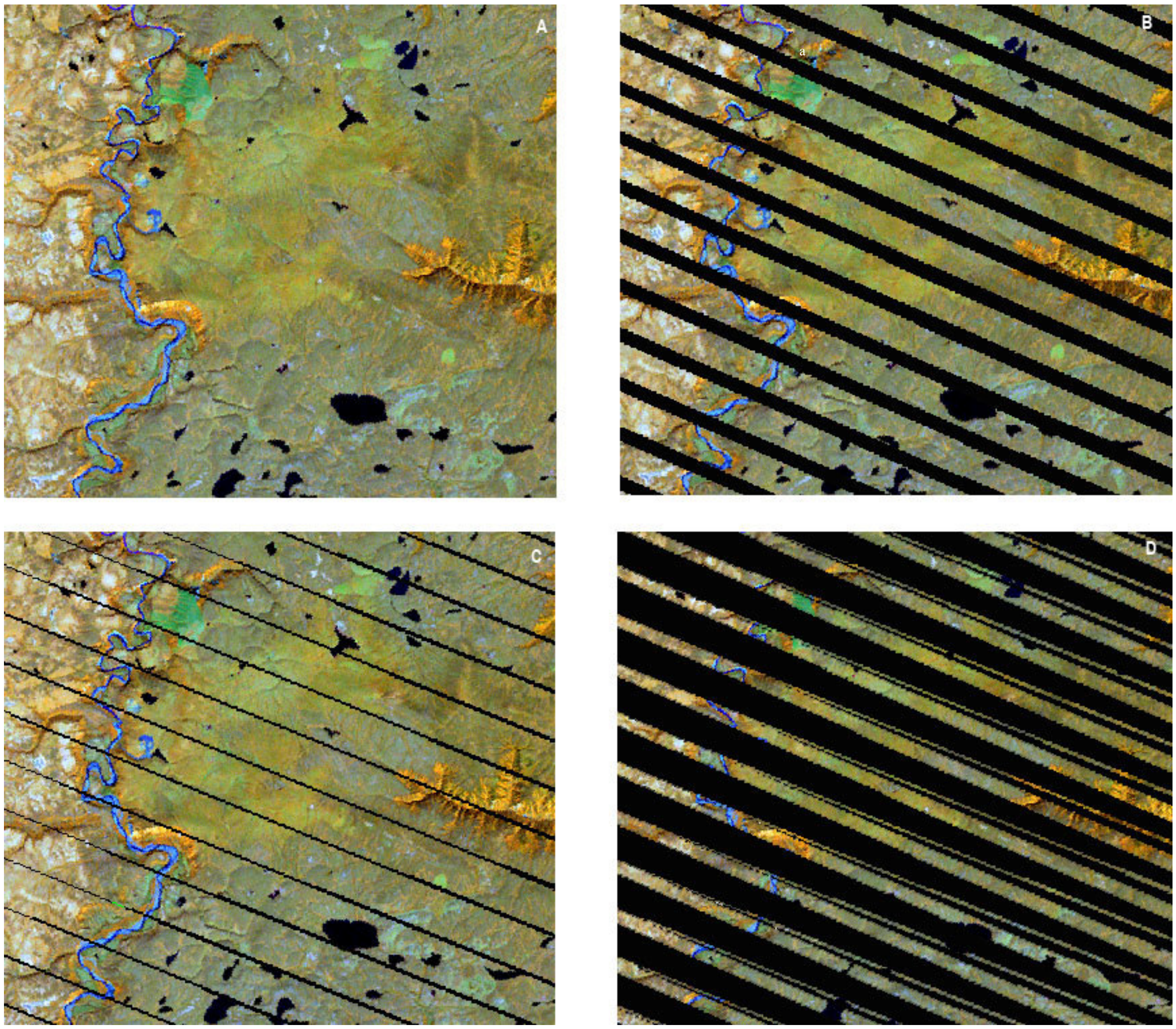

Figure 2. A) Original scene for comparison, B) striping in original image with striped areas used for consistency comparison, C) striping in adjacent image, D) overlap imagery for use in regression algorithms

2.2 Normalisation

Because the two scenes were acquired one year and 26 days apart, areas of fires and snow were excluded from the sample used in the normalisation because of their low probability of occurring on consecutive passes of the sensor in a single season. The only relevant effects remaining in the imagery after normalization were those caused by atmosphere, minor cloud and phenological variation due to the short growing season. The scenes were inspected visually and were deemed to be representative of the features and conditions met in images from the same season. As well, only the data that were available in both the simulated SLC-off scenes were utilized in the normalisation to mimic actual scene conditions.

A random set of 1000 samples based on 5X5 average filtered images was used to create the coefficients separately for three spectral bands to reduce the effect of any georeferencing error. The Theil-Sen [4] regression, as introduced to the remote sensing community by [5] and utilised by [3] for large-area Landsat mosaics, was used for the normalisation of the fill image to the primary image. Theil-Sen is usefull as a robust regression technique for normalisation in the north because it is insensitive to ~30% of outliers [5]. Often the outliers can be in the form of phenological differences due to short growing season, sporadic cloud cover, local variance in surface moisture or subtle atmospheric caused by low sun angles at the time of image acquisition. A full discussion on Theil-Sen regression is available in [5] and its use in signature extension for northern mapping [7] as well as in the creation of a landcover map of Northern Canada in [8] thereby negating the need for further examination.

As a gap-filling procedure, Theil-Sen regression for use in normalisation is compared to the ordinary least squares regression (OLS) method used by [9] for normalisation within their standard gap-fill algorithm. The authors of [5] suggest that Theil-Sen is a potential replacement for the OLS regression in remote sensing applications due to its simplicity in computation, robustness to outliers, and its need for limited a priori information regarding measurement errors. The result of the normalisation based on OLS was therefore assessed along with the results of the Theil-Sen based normalisation procedure.

2.3 Vegetation Indices

The NDVI,

$$NDVI = (NIR - Red)/(NIR + Red), \quad (1)$$

is useful due to its ability to normalise the difference between maximum absorption within the red wavelengths and peak reflection in the NIR, and is used in many studies of northern vegetation and it's environment [10], [11], [12], [13], [14], [15], [16].

Unfortunately, it has been shown to be insensitive to subtle variations in sparsely vegetated regions of the north [17], [18], [19]. The RSR is calculated as

$$RSR = SR*\{1-[(SWIR-SWIRmin)/(SWIRmax-SWIRmin)]\}, \quad (2)$$

and was explored as an alternative index to compensate for the shortcomings of NDVI because the RSR has the potential to be applicable across coniferous and deciduous species, thus reducing the need for a thematic map, and is more highly correlated to biophysical measurements such as LAI due to the incorporation of the SWIR band to compensate for differences in canopy closure [20].

Since we did not have values for maximum and minimum SWIR values to represent completely open and completely closed canopies, we calculated them from the imagery as in [20]. The reflectance for the minimum SWIR was 10% and the reflectance for maximum SWIR was 20% These results were acceptable as there were no fully closed canopies in the study area, but much more open canopies than those in [20].

A detailed discussion of the vegetation indices used as well as others available is included in [19] for northern study areas and [20] for forest mapping and will therefore not be discussed further. Both indices were calculated for the gap filled areas of the images and the results were compared to the calculated indices for the original imagery.

2.4 Quality Assessment

After areas of cloud, shadow and fire were removed a random sample of 1500 points from the gap-filled sections of the imagery was compared to the original data that originally populated the gaps in terms of band reflectance, NDVI and RSR. The scatter plot and histogram statistics were then calculated to gain an understanding of both normalisation processes. The DNs were graphed in scatter plots with the original data on the x-axis and the normalised data on the y as well as the results of NDVI and RSR in separate scatter plots and the correlation coefficients were calculated. The dynamic ranges of the data sets were calculated from the histograms of gap-filled areas of the imagery as well as the original imagery to support the findings of the gains and offsets observed in the scatter plots.

3 RESULTS

The results of the Theil-Sen and OLS regression normalisation techniques were compared to the original data as well as the fill data without any normalisation applied. Because the correlation coefficients do not change with normalisation, the improvements can be measured using the modeling efficiency statistic (EF) which accounts for bias in the data. EF is calculated as

$$EF = 1 - \sum (y_i - \hat{y})^2 / \sum (y_i - \bar{y}_i)^2 \quad (3)$$

EF is a dimensionless statistic that is a measure of model performance that directly relates model predictions to observed data by comparing the observed results to the predicted [21].

EF is a similar measure to R^2 in that they both are interpreted as the proportion of variation explained by the fitted regression line [22]. They differ in that R^2 can remain high even when the data contain potential bias because the resultant data are not being compared to the predicted result. Since EF compares the actual result to the predicted (assuming the predicted result is accurate) any bias in the data will become evident due to a lower EF value. Even though the bands were fairly well correlated to begin with (R2 of 0.63 for Red and 0.84 for NIR and SWIR) the normalisation using Theil-Sen regression gave improved results for normalisation in two of the three bands (Red and SWIR) while OLS reduced the quality of the data in all three cases. In the Red band the EF measure was 0.63 for Theil-Sen, -0.80 for OLS and –1.24 for the data without normalisation. In the NIR, both Theil-Sen and OLS reduced the quality of the data with EF of 0.78 and 0.31 respectively while the uncorrected data had an EF of 0.84. Finally, the SWIR band had an EF of 0.78 for Theil-Sen and 0.35 for OLS while the uncorrected data had an EF of 0.72.

Figure 3 shows the scatter plots for the Red band. Although both regression and the uncorrected data have R2 values of 0.63 the EF values give an indication as to the quality of the data after normalisation. Perfectly normalised data would have a 1 to 1 relationship and EF value of 1 while the EF has a theoretical lower bound of negative infinity although, as stated by Mayer and Butler (1993), any model or equation with a negative EF value cannot be recommended for use.

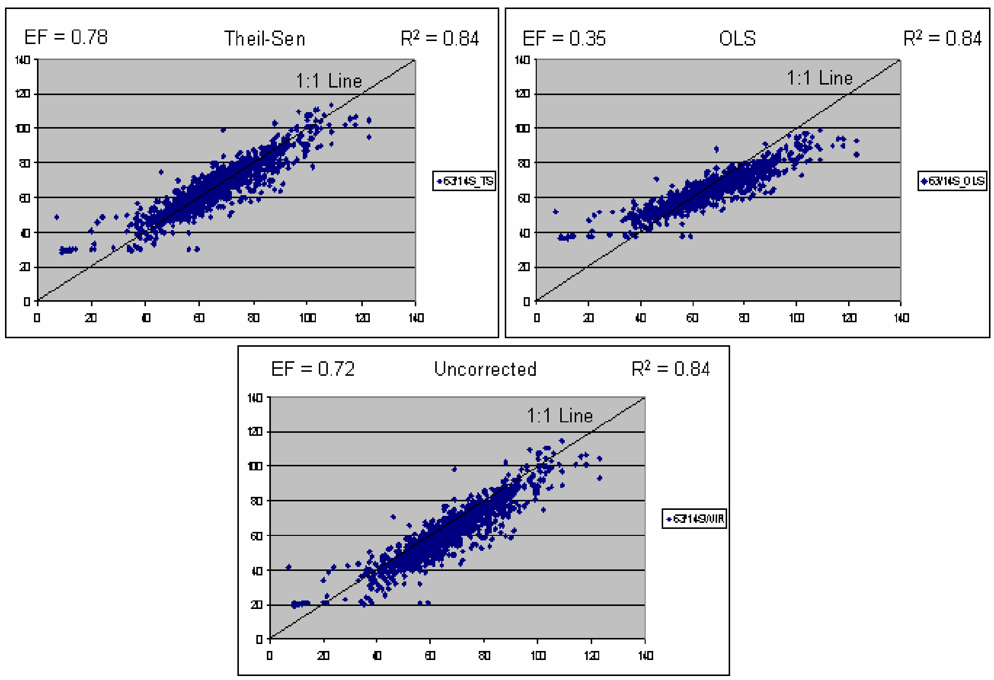

Figure 3. Scatter plots for Theil-Sen, OLS and Uncorrected data in the SWIR band.

Normalisation using Theil-Sen regression would give a more consistent result across all bands. Although the NIR was reduced in quality, an EF value of 0.78 is quite acceptable. The consistency of the bands becomes very important when the data are used in vegetation indices (VIs) and VI modeling through time. The improvement in both the Red and SWIR bands offsets the slightly poorer result in the NIR band.

When the resultant VIs are compared, Theil-Sen shows great potential for maintaining consistency, as the R^2 for the RSR improves to 0.71 with a EF of 0.53 over uncorrected data (R^2 of 0.63, EF of –0.75) while the OLS does not change the correlation and has an EF of 0.30. With NDVI both the Theil-Sen and OLS greatly improve the correlation with R^2 0.76 and 0.79 respectively when compared to the uncorrected data (R^2 of 0.56) but the EF shows that only the bands normalised using the Theil-Sen regression give an appropriate NDVI as the EF is 0.47 while OLS has a EF of –0.93 when compared to the uncorrected data (EF of –1.76).

Table 1. Correlation coefficients and Modelling Efficiency for Theil-Sen and OLS regression normalisation results of individual bands

	R^2	TS EF	OLS EF	Uncorrected EF
Red	0.63	0.63	-0.80	-1.24
NIR	0.84	0.78	0.31	0.84
SWIR	0.84	0.78	0.35	0.72

Table 2. Correlation coefficients and Modelling Efficiency for Theil-Sen and OLS regression normalisation results of vegetation indices

	TS R^2	TS EF	OLS R^2	OLS EF	Uncorrected R^2	Uncorrected EF
RSR	0.71	0.53	0.63	0.30	0.63	-0.75
NDVI	0.76	0.47	0.79	-0.93	0.56	-1.76

4 DISCUSSION

Although the quality of the NIR band was reduced, Theil-Sen performed better overall than OLS because of the improvement in the modeling of the Red and SWIR bands. The uncorrected imagery was well correlated in both the NIR and SWIR bands, but not nearly as well correlated in the red band. Theil-Sen outperformed the OLS in the two indices tested due to the consistency in the data. Although the NDVI points created with OLS have a higher R^2 than the NDVI created using Theil-Sen normalised data, the EF shows the data created using OLS regression to be unusable for NDVI.

When the data between adjacent images are well calibrated to begin with, the OLS regression can still be suitable for normalisation but problems will be encountered when outliers are present in the data. Theil-Sen on the other hand, excels when the fill image is less correlated (as in the red wavelengths) due to insensitivity to ~30% of outliers. Theil-Sen consistently maintained the dynamic range of the data where OLS reduced the dynamic range of the fill data. Without normalisation the user would be unable to use the data from the original imagery to infill the gaps because the reflectance in the Red band is too low to give consistent results in any of the indices. Due to the open nature and low biomass of the study area, the results for the indices should fall below 2.5 for the RSR and 0.6 for NDVI, yet the results for the non-normalised data have maximum values of 6, and 0.85 for RSR and NDVI respectively with the mean values for the resultant indices being well above the maximum RSR and NDVI values for the vegetation in the study area of 2.5 and 0.6 respectively.

5 CONCLUSION

As striping of Landsat 7 data will need to be corrected until a suitable replacement for the sensor is in place, the data will have to be gap-filled before it can be used for any biophysical modelling using measurements such as VIs. Erroneous data in the individual satellite bands can be amplified in the VIs. A robust regression technique is therefore needed to create the gap-filled products using adjacent satellite scenes for any biophysical modelling. The Thiel-Sen regression seems to be a more appropriate technique for pre-processing and gap-filling procedures than OLS as its insensitivity to outliers creates a better fit to the original image data. Although the R^2 values are similar between Theil-Sen and OLS normalised data, EF shows the shortcomings of OLS as a regression technique. Given that Theil-Sen was most effective in normalising the data across all bands, when compared to the OLS and original data, it is a valuable tool in the normalisation of data for gap-filling procedures in the north where the sparse nature of the vegetation causes atmosphere and moisture to have a significant effect on the digital signal. The resultant VIs are more consistent when the adjacent data are normalised using the Theil-Sen regression than the OLS or simple infilling using uncorrected data.

ACKNOWLEDGEMENTS

The authors would like to thank Darren Pouliot for his comments and introduction of EF.

REFERENCES

[1] JENSEN, J. R., 1996, Introductory Image Processing: A remote sensing perspective, New Jersey: Prentice Hall.

[2] DU, Y., CIHLAR, J., LATIFOVIC, R., 2001, Radiometric normalisation, compositing, and quality control for satellite high resolution image mosaics over large areas, *IEEE Transactions on Geoscience and Remote Sensing,* 39, 623-634

[3] OLTHOF, I., POULIOT, D., FERNANDES, R., LATIFOVIC, R., 2005b, Landsat-7 ETM+ radiometric normalisation comparison for northern mapping applications, *Remote Sensing of Environment,* 95, 388-398

[4] KENDALL, M. G., STUART, A. S., 1967, Advanced Theory of Statistics, vol.2, London: Charles Griffin and Company

[5] FERNANDES, R., LEBLANC, S., 2005, Parametric (modified least squares) and non-parametric (Theil-Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors, *Remote Sensing of Environment,* 95, 303-316

[6] WULDER, M.A., FRANKLIN, S.E., WHITE, J.C. 2004, Sensitivity of hyperclustering and labelling land cover classes to Landsat image acquisition date, *International Journal of Remote Sensing,* 25 5337-5344.

[7] OLTHOF, I., BUTSON, C., FRASER, R., 2005a, Signature extension through space for northern landcover classification: A comparison of radiometric correction methods, *Remote Sensing of Environment,* 95, 290-302

[8] OLTHOF, I., BUTSON, C., FERNANDES, R., FRASER, R., LATIFOVIC, R., ORAZIETTI, J., 2005c, Landsat ETM+ mosaic of northern Canada, *Canadian Journal of Remote Sensing* (submitted)

[9] NASA, 2004, SLC-off Gap-Filled Products, Gap-Fill Algorithm Methodology, Version 2, NASA

[10] HOPE, A.S., KIMBALL, J.S., STOW, D.A. 1993, The relationship between tussock tundra spectral reflectance properties and biomass and vegetation composition, *International Journal of Remote Sensing,* 14, 1861-1874

[11] STOW, D.A., BURNS, B.H., HOPE, A.S., 1993, Spectral, spatial and temporal characteristics of Arctic tundra reflectance, *International Journal of Remote Sensing,* 14, 2445-2462

[12] MCMICHAEL, C.E., HOPE, A.S. STOW, D.A., FLEMING, J.B., 1997, The relation between active layer depth and a spectral vegetation index in arctic tundra landscapes of the North Slope of Alaska, *International Journal of Remote Sensing,* 18, 2371-2382

[13] GOULD, W.A., EDLUND, S., ZOLTAI, S., RAYNOLDS, M., WALKER, D.A., MAIER, H., 2002, Canadian Arctic vegetation mapping, *International Journal of Remote Sensing,* 23, 4597-4609

[14] GOULD, W.A., RAYNOLDS, M., WALKER, D.A., 2003, Vegetation, plant biomass, and net primary productivity patterns in the Canadian Arctic, *Journal of Geophysical Research,* 108, 8-1–8-14

[15] JIA, G.J., EPSTEIN, H., 2003, Greening of Arctic Alaska, 1981-2001, *Geophysical Research Letters,* 30, 3-1-3-4

[16] WALKER, D.A., EPSTEIN, H.E., JIA, G.J., BALASER, A., COPASS, C., EDWARDS, E.J., GOULD, W.A., HOLLINGSWORTH, J., KNUDSON, J., MAIER, H.A., MOODY, A., RAYNOLDS, M.K., 2003, Phytomass, LAI, and NDVI in northern Alaska: Relationships to summer warmth, soil pH, plant function types, and extrapolation to the circumpolar Arctic, *Journal of Geophysical Research,* 108, 10-1-1015

[17] HUETE, A. R., TUCKER, C. J., 1991, Investigation of soil influences in AVHRR red and near infrared vegetation index imagery, *International Journal of Remote Sensing,* 12, 1223-1242

[18] REES, W. G., GOLUBEVA, E.I., WILLIAMS, M., 1998, Are vegetation indices useful in the arctic? *Polar Record,* 34, 333-336

[19] LAIDLER, G. J., TREITZ, P., 2003, Biophysical remote sensing of arctic environments, *Progress in Physical Geography,* 27, 44-68

[20] BROWN, L., CHEN, J., LEBLANC, S., CIHLAR, J., 2000, A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: An image and model analysis, *Remote Sensing of Environment,* 71, 16-25

[21] LOAGUE, K., GREEN, R.E., 1991, Statistical and graphical methods for evaluating solute transportation models: Overview and Application, *Journal Contaminant Hydrology,* 7, 51-73

[22] MAYER, D.G., BUTLER, D.G., 1993, Statistical validation, *Ecological Modelling,* 68, 21-32