Alexandridis, T.K., Cherif, I., Kalogeropoulos, C., Monachou, S., Eskridge, K., and Silleos, N., 2013. Rapid error assessment for quantitative estimations from Landsat 7 gap-filled images. Remote Sensing Letters, 4(9): 920-928.

Rapid error assessment for quantitative estimations from Landsat 7 gap-filled images

Thomas K. Alexandridis[1]*, Ines Cherif[1], Christos Kalogeropoulos[1], Styliani Monachou[1], Kent Eskridge[2] and Nikolaos Silleos[1]

[1] *Lab of Remote Sensing and GIS, School of Agriculture, Aristotle University of Thessaloniki, Greece*

[2] *Department of Statistics, University of Nebraska, Lincoln, USA*

*Corresponding author: Thomas Alexandridis (thalex@agro.auth.gr)

Address: Lab of Remote Sensing and GIS, School of Agriculture, Aristotle University of Thessaloniki, University Campus, University Box 259, 54124 Thessaloniki, Greece

Abstract

The failure of the Scan Line Corrector (SLC) of the Landsat ETM+ instrument in 2003 had resulted missing values for 22% of each scene. As the remaining of the pixels were of high quality, several procedures had been developed to fill the gaps and increase the usability of the SLC-off images. In this paper, a methodology is presented to assess the error when estimating quantitative parameters from gap-filled Landsat 7 images. The error from the gap-filling procedure was estimated by using an external reference image. The methodology was applied in a Mediterranean river basin using two types of gap-filling methods and the error was estimated for Leaf Area Index (LAI), actual evapotranspiration (ETa) and soil moisture in the rootzone (SMrz), three remotely sensed products which are commonly used in hydrological studies. The results suggest that the interpolation method had lower errors in all examined products. The proposed methodology is an imperative step that each user of gap filled products could use to estimate the associated error before using the maps.

1. Introduction

On May 31, 2003 the Scan Line Corrector (SLC) in the ETM+ instrument on board the Landsat 7 satellite failed. The purpose of the SLC was to compensate for the forward motion (along-track) of the spacecraft so that the resulting scans would be aligned parallel to each other. Without the effects of the SLC, the instrument images the Earth in a "zig-zag" fashion, resulting in some areas that are imaged twice and others that are not imaged at all. The net effect is that approximately 22% of the data in a Landsat 7 scene is missing when acquired without a functional SLC (http://landsat7.usgs.gov/).

Several techniques have been suggested to fill in the missing Landsat values either by the use of other Landsat images (compositing method) either by the use of interpolation methods (Howard and Lacasse 2004). The USGS Landsat Project and NASA Landsat Project Science Office jointly produced gap-filled products using two versions of the compositing method (USGS *et al.* 2003). Other efforts to fill in gaps with compositing techniques include monthly (Roy *et al.* 2010) to multi-year aggregations (Potapov *et al.* 2011). However, the performance of this method is low if the scenes being combined exhibit high temporal variability and radical differences in target radiance due, for example, to the presence of clouds or land cover changes that have occurred between the times of images acquisition. Moreover, they assume that another Landsat image acquired at a convenient time (e.g. anniversary date) would be available.

Several interpolation techniques have been developed to overcome these issues, such as Inverse Distance Weighted, Neighborhood Similar Pixel Interpolator, co-kriging and kriging geostatistical methods, and spectral interpolation within a multi-scale segmentation model (Chen *et al.* 2011, Maxwell *et al.* 2007, Pringle *et al.* 2009, Zhang *et al.* 2007). However, they are computationally intensive, and may produce visual artifacts as well as smoothing effects, especially for small objects (Pringle *et al.* 2009, Zhang *et al.* 2007). Other studies made by the USGS showed that the histogram matching method in phase I provided the best overall performance and was adopted for phase II as well (Storey *et al.* 2005).

Recognizing the importance of error assessment of the newly proposed gap-fill techniques, several methods for error estimation have been used. Visual assessment of the results seems to be the most used technique, together with mean absolute differences of DN (Roy *et al.* 2010, Zhang *et al.* 2007), linear regressions of estimated with in-situ measurements (Kovalskyy *et al.* 2012), and post classification accuracy assessments (Bedard *et al.* 2008, Boloorani *et al.* 2008, Trigg *et al.* 2006). Several sources of reference data have been used, such as Landsat 5 TM images acquired a day apart from the filled-in Landsat 7 image (Bedard *et al.* 2008, Zhang *et al.* 2007), the original parts of Landsat 5 TM images in which gaps similar to SLC-off were simulated (Boloorani *et al.* 2008, Maxwell *et al.* 2007), and in-situ measured NDVI from flux tower (Kovalskyy *et al.* 2012).

However, the above mentioned error assessment techniques focus on the raw pixel values (DNs of the satellite images), or on qualitative results, such as areal estimates of land cover classes identified from image classification. The error is more complicated to assess when estimating quantitative parameters such as Leaf Area Index (LAI), especially when the parameters fluctuate daily, such as soil moisture content (SM), and in cases where in-situ measurements are difficult to obtain, such as actual evapotranspiration (ETa). The aim of this work is to present a methodology for assessing the error when estimating quantitative parameters from gap-filled Landsat 7 images.

The methodology was applied for estimations of LAI, ETa and SM for a river basin in Greece/Bulgaria within the frame of an FP7-EU project (Merging hydrological models and Earth observation data for reliable information on water - MyWater). In the project's workflow, gap-filled Landsat 7 images were used after the cease of operations of Landsat 5, and the accuracy of the parameters estimation was important before they were used in hydrological applications.

2. Proposed methodology

In order to assess the accuracy of using Landsat 7 gap-filled images for the estimation of a quantitative parameter P (e.g. ETa, LAI or SM), the estimates within the gaps must be compared to the equivalent Landsat 7 SLC-on values. Since such values are missing in the L7 SLC-off image a reference contemporary map is used. For instance the reference map could be a map estimated using Landsat 5 input data.

The error due to gap filling when estimating the parameter P is quantified using two metrics: the mean (*E*) and variance (*Var*) of the difference between the estimates of P using the Landsat 7 gap-filled and Landsat 7 SLC-on image. The mean difference and error variance are two quantities that measure respectively the bias and the precision of the P estimates. Equations (1) and (2) are used to derive the mean difference and error variance and clarify the need to introduce a reference image to account for the missing values in the Landsat 7 SLC-off image.

$$E(P(\mathrm{L7GF}) - \mathrm{P}(\mathrm{L7})) = E(\ \mathrm{P}(\mathrm{L7GF}) - \mathrm{P}(\mathrm{ref})) - E(P(\mathrm{L7}) \text{-} P(\mathrm{ref})) \quad (1)$$

$$\begin{aligned} Var(P(\mathrm{L7GF}) - \mathrm{P}(\mathrm{L7})) &= Var(\ \mathrm{P}(\mathrm{L7GF}) - \mathrm{P}(\mathrm{ref}) - P(\mathrm{L7}) + P(\mathrm{ref})) \\ &= Var(P(\mathrm{L7GF}) - P(\mathrm{ref})) + Var(\mathrm{P}(\mathrm{L7}) - P(\mathrm{ref})) \\ &\quad -2 * Cov((P(\mathrm{L7GF}) - P(\mathrm{ref}));(\mathrm{P}(\mathrm{L7}) - P(\mathrm{ref}))) \end{aligned} \quad (2)$$

where: L7GF is the Landsat 7 gap-filled image, L7 is the SLC-on image, ref is the reference image (e.g. contemporary or anniversary image).

The equations are applied at the location of the L7 SLC-off gaps. As shown above the introduction of the reference image in the equations does not affect the final results. The effect of the reference image appears when *E*(P(L7) - P(ref)) and *Var*(P(L7) - P(ref)) are approximated by their value outside the gaps as they can not be computed within the gaps. The term *Cov*((P(L7GF) - P(ref)); (P(L7) - P(ref)) is calculated in simulated gaps on the original Landsat 7 SLC-off image, as no common gap and non-gap pixels can exist in the image.

3. Application in a Mediterranean river basin

3.1. Selection of the images

The proposed methodology was applied in Nestos river basin (Greece and Bulgaria, WRS2 183/031). The following images were used: a Landsat 7 SLC-off image acquired on 23/08/2011, a Landsat 5 image acquired on 24/07/2009 as fill-in image and a Landsat 5 image acquired on 15/08/2011 as reference image. All images where cloud free, and the reference image was acquired only 8 days apart from the SLC-off image, which is the best case scenario.

3.2. Gap-filling procedures

Two methods for gap-filling were applied on the Landsat 7 SLC-off images. The first was the compositing method of Phase 1 methodology suggested by USGS (Roy *et al.* 2010), which is based on mosaiking data from another image to fill in the gaps. This method requires another image which is spectrally similar and acquired as close to the date of the original image acquisition (anniversary images). A histogram matching is applied to the fill-in image in order to minimize any spectral difference. The success of applying this method lies on the existence of an appropriate fill-in image.

The second method was the Inverse Distance Weighted (IDW) interpolation approach (GDAL 2012), which is based on interpolating the missing values at the gaps using the existing values in the vicinity of a 17x17 pixel moving window. This method does not require another image, but its performance is low in areas with detailed spatial patterns.

3.3. LAI estimation

LAI was estimated by applying first the dark object subtraction atmospheric correction and afterwards the cosine topographic correction on the Landsat images (Civco 1989). The normalized difference vegetation index (NDVI) was calculated and finally a LAI model was used (Topaloglou *et al.* 2013). The model for estimating LAI was based on the look up table (LUT) method of MODIS Land Team, in which a list of NDVI values that correspond to certain LAI values on different land cover types has been proposed after an investigation with ground truth data and Landsat 5 TM images all over the world (Knyazikhin *et al.* 1999).

3.4. ETa estimation

The Landsat images were processed using the ITA-MyWater algorithms, a continuation of the ITA-Water (Alexandridis *et al.* 2009). These algorithms for evapotranspiration estimation are based on the application of the Surface Energy Balance Algorithm for Land (SEBAL) that derives ETa through the computation of the three Energy Balance (EB) terms Rn, G_0 and H (Bastiaanssen *et al.* 1998). At the evaporating surface the EB equation can be written as: $Rn - G - \lambda ET - H = 0$, where Rn is the net radiation, H is the sensible heat flux, G is the soil heat flux, and λET is the latent heat flux. λET is calculated as a residual after estimating the remaining terms. All Landsat bands were used within ITA-MyWater together with spatially distributed meteorological data from weather forecast models. The main outputs of the processing chain were the evaporative fraction (EF) and actual evapotranspiration (ETa) maps with a spatial resolution of 30m. EF represents the ratio between the latent heat flux (λET) and the available energy at the land surface.

3.5. SMrz estimation

Within ITA-MyWater algorithms, soil moisture in the root zone (SMrz) is estimated using a global empirical method based on Landsat input data (Ahmad and Bastiaanssen 2003). The relative soil moisture is linked by an exponential relationship to the EF. The soil moisture is then retrieved by multiplication with the saturated water content map that brings information about soil porosity in the root zone. The output is a root zone soil moisture map with a spatial resolution of 30m.

4. Results

The study area was covered by 18% from gaps. This is a good case scenario, as it is below the mean Landsat scene average (22%). In figure 1, the location of the study area within the Landsat 7 scene is displayed, together with the spatial distribution of gaps.

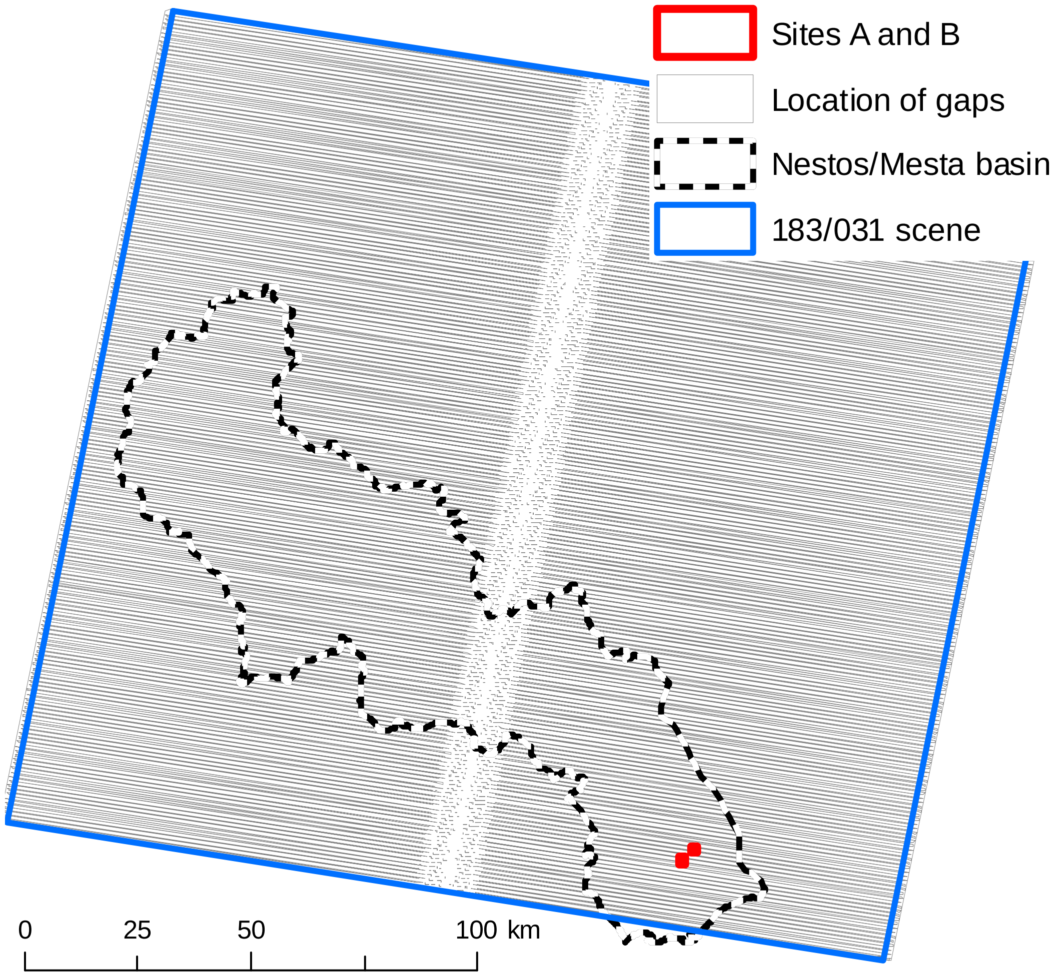

Figure 1. Study area location related to the Landsat 7 scene and the distribution of gaps.

The variance of the error was estimated with equation (2), but standard deviation values are given here for the coherence of units. The standard deviation of the error of estimating ETa, SMrz and LAI in the study area using the Landsat 7 gap-filled image with the compositing method is 0.761 mm/day, 0.040 cm^3/cm^3 and 0.616 m^2/m^2, respectively (Table 1). The bias is low, 0.144 mm/day for ETa, 0.012 cm^3/cm^3 for SMrz, and -0.264 m^2/m^2 for LAI. When using the interpolation method for gap-filling the Landsat 7 image, the standard deviation of the error of estimating ETa, SMrz and LAI in the study area is 0.292 mm/day, 0.015 cm^3/cm^3 and 0.567 m^2/m^2, respectively. The bias is equally low, -0.064 mm/day for ETa and 0.005 cm^3/cm^3 for SMrz, with the exception of the somewhat larger value of -0.399 m^2/m^2 for LAI.

Table 1: Standard deviation and mean of the differences in parameters estimation using the gap-filled Landsat 7 for two gap-filling methods.

	ETa [mm/day]	SMrz [cm^3/cm^3]	LAI [m^2/m^2]
-------------------- Compositing Method -----------------			
σ (P(L7GF)-P(L7))	0.761	0.040	0.616
E (P(L7GF)-P(L7))	0.144	0.012	-0.264

	--------------------- Interpolation Method ------------------		
σ (P(L7GF)-P(L7))	0.292	0.015	0.567
E (P(L7GF)-P(L7))	-0.064	0.005	-0.399

Where σ is the standard deviation, E is the mean, and P is the parameter (ETa, SMrz, or LAI).

The above mentioned results quantify the error induced for the entire maps. Furthermore, the study of the spatial distribution of errors can be used to show specific errors in several cases (Maxwell and Craig 2008, Zhang *et al.* 2007). As an example, the spatial differences between the filled-in and the reference image for SMrz estimation are shown in Figure 2. Two specific sites have been selected for that purpose, at the fringe of the irrigation network of Nestos floodplain: Site A with irrigated herbaceous crops and Site B with rainfed herbaceous crops. In the case of Site A there are evident differences within the gaps when using the compositing fill-in technique (Figure 2a), which are minimized when using the interpolation fill-in technique (Figure 2b). These differences are attributed to the difference of soil moisture between the dates of acquisition of the two images. Some parcels in Site A had been irrigated just before the acquisition of the fill-in image leading to higher SMrz values in the gaps. The detail of spatial patterns within the gaps due to irrigation is nevertheless conserved. The low difference in Site A in Figure 2b is because of the interpolation of the parameter values that stretched the SMrz values on the width of the gap. The parameter estimation is acceptable even though the spatial detail due to the patchwork of agricultural land is lost. In Site B, where rain is the main factor affecting SMrz, the fill-in results from the two methods are similar, as no rain event is recorded on any of the acquisition dates.

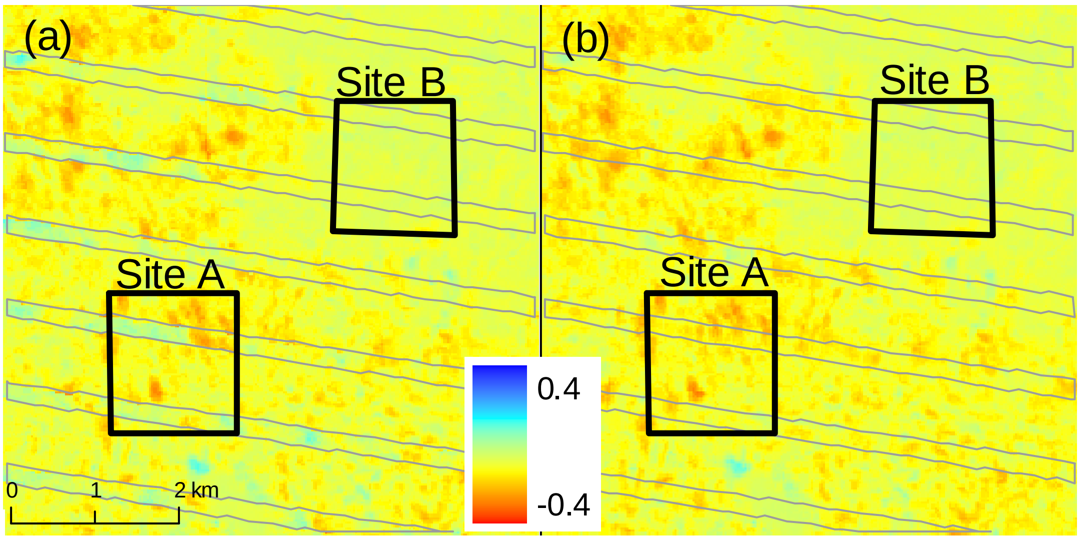

Figure 2. Spatial distribution of errors in SMrz estimation using (a) the compositing method, and (b) the interpolation method. Gray polygons symbolize the extents of the filled gaps.

5. Discussion and conclusion

The suggested methodology was used successfully to estimate the error when using gap-filled Landsat 7 images to estimate LAI, ETa, and SMrz in a river basin. This result is useful for understanding the usability of the products from gap-filled images in downstream applications. Thus, users of these products are urged to estimate the induced error from any gap-filling procedure using the suggested easy to use methodology.

A positive (resp. negative) value of the mean difference corresponds to an overestimation (resp. underestimation) in average of the parameter and reflects (i) the difference in vegetation or water conditions during the acquisition date of the fill-in and SLC-off images in the case of the compositing method, and (ii) the higher variability in the spatial distribution of the parameters in the SLC-on image than in the gap-filled image when applying the interpolation method.

It is evident that the interpolation gap-filling method performed better than the compositing one in the study area. For all products, the compositing method had higher error, which could be due to the different development of vegetation and water conditions between the SLC-off and the fill-in image. In such cases, the interpolation method is reported to perform better (Pringle *et al.* 2009, Zhang *et al.* 2007). Clearly, the large temporal difference between the SLC-off and the fill-in image did not cause a prohibiting large error, as both images were acquired in the same season (hot, dry Mediterranean summer conditions).

In the case of the interpolation method, the error standard deviation values show that 95% of the differences in LAI, ETa, and SMrz estimates are within +/- 1.135, 0.583, and 0.029, respectively (+/- 2 σ). These error bounds are lower (in the case of ETa and SMrz) and relatively close (in the case of LAI) to the reported error of the equivalent estimation methodology. Bastiaanssen et al. (2005) reported 85% accuracy for daily ETa estimation using Landsat images in wet conditions and at field scale. For the Nestos river basin, daily ETa for irrigated areas is typically 6 to 8 mm/day during the peak of the growing season, the error is thus around 1 mm/day. Ahmad and Bastiaanssen (2003) obtained in irrigated plains in central Pakistan and Mexico an average error of 0.035 cm^3/cm^3 when estimating soil moisture in the root zone. For LAI, an RMSE of 0.8 m^2/m^2 has been estimated in the same test site (Topaloglou *et al.* 2013). Therefore it is safe to use the LAI, ETa, and SMrz products in quantitative estimations in this site and in downstream hydrological applications.

The error assessment methodology is based on the fundamental statistical concepts of bias and precision which are critical in any comparison of competing methods of estimating a quantity. It is necessary to consider both characteristics since a method that is unbiased may not be precise and vice versa. For example with LAI in this case, the compositing method is most precise (smallest σ) but has the largest bias.

As shown, the proposed methodology can be applied to both the compositing and interpolation methods for gap-filling. It could be used to choose the most appropriate gap-fill method, which may depend on the image conditions (e.g. % cloud cover), location of area of interest according to the gap distribution zones, date difference with the fill-in image, and spatial pattern of the area of interest (Chen *et al.* 2011, Zhang *et al.* 2007). Thus, it's an imperative step before further use of the gap-filled results.

With the loss of ALOS/AVNIR-2, the recent loss of Landsat 5 TM, the malfunction of MODIS/ASTER, and the recent end of operations of SPOT 4, there are very few options left for high resolution (10-30m) monitoring of vegetation and water parameters. Thus, until the newly launched Landsat 8 LDCM (February 2013) is operational and SENTINEL-2 (planned for 2014) is successfully in orbit, the use of Landsat 7 ETM+ gap-filled images may be the optimum source of data for operational monitoring. Specifically for hydrological studies, the Landsat 7 imagery is essential due to the fact that it is the only satellite in operation equipped with high resolution thermal infrared sensor.

6. References

AHMAD, M.U.D., and BASTIAANSSEN, W.G.M., 2003. Retrieving soil moisture storage in the unsaturated zone using satellite imagery and bi-annual phreatic surface fluctuations. *Irrigation and Drainage Systems,* **14**, 141-161.

ALEXANDRIDIS, T.K., CHERIF, I., CHEMIN, Y., SILLEOS, G.N., STAVRINOS, E., and ZALIDIS, G.C., 2009. Integrated Methodology for Estimating Water Use in Mediterranean Agricultural Areas. *Remote Sensing,* **1**, 445-465.

BASTIAANSSEN, W.G.M., MENENTI, M., FEDDES, R.A., and HOLTSLAG, A.A.M., 1998. A remote sensing surface energy balance algorithm for land (SEBAL): 1. Formulation. *Journal of Hydrology,* **212-213**, 198-212.

BASTIAANSSEN, W.G.M., NOORDMAN, E.J.M., PELGRUM, H., DAVIDS, G., THORESON, B.P., and ALLEN, R.G., 2005. SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. *Journal of Irrigation and Drainage Engineering,* **131**, 85-93.

BEDARD, F., REICHERT, G., DOBBINS, R., and TREPANIER, I., 2008. Evaluation of segment-based gap-filled Landsat ETM+ SLC-off satellite data for land cover classification in southern Saskatchewan, Canada. *International Journal of Remote Sensing,* **29**, 2041-2054.

BOLOORANI, A.D., ERASMI, S., and KAPPAS, M., 2008. Multi-source remotely sensed data combination: Projection transformation gap-fill procedure. *Sensors,* **8**, 4429-4440.

CHEN, J., ZHU, X., VOGELMANN, J.E., GAO, F., and JIN, S., 2011. A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. *Remote Sensing of Environment,* **115**, 1053-1064.

CIVCO, D.L., 1989. Topographic normalization of Landsat Thematic Mapper digital imagery. *Photogrammetric Engineering & Remote Sensing,* **55**, 1303-1309.

GDAL, 2012. gdal_fillnodata.py - Fill raster regions by interpolation from edges, http://www.gdal.org/gdal_fillnodata.html. (Last accessed on 18/12/2012).

HOWARD, S.M., and LACASSE, J.M., 2004. An evaluation of gap-filled landsat SLC-off imagery for wildland fire burn severity mapping. *Photogrammetric Engineering and Remote Sensing,* **70**, 877-880.

KNYAZIKHIN, Y., MYNENI, R.B., TIAN, Y., WANG, Y., and ZHANG, Y., 1999. Estimation of vegetation canopy leaf area index and fraction of photosynthetically active radiation absorbed by vegetation from remotely sensed multi-angle and multi-spectral data. In, *International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 1872-1874)

KOVALSKYY, V., ROY, D., ZHANG, X.Y., and JU, J., 2012. The suitability of multi-temporal web-enabled Landsat data NDVI for phenological monitoring - A comparison with flux tower and MODIS NDVI. *Remote Sensing Letters,* **3**, 325-334.

MAXWELL, S.K., and CRAIG, M.E., 2008. Use of landsat ETM+ SLC-off segment-based gap-filled imagery for crop type mapping. *Geocarto International,* **23**, 169-179.

MAXWELL, S.K., SCHMIDT, G.L., and STOREY, J.C., 2007. A multi-scale segmentation approach to filling gaps in Landsat ETM+ SLC-off images. *International Journal of Remote Sensing,* **28**, 5339-5356.

POTAPOV, P., TURUBANOVA, S., and HANSEN, M.C., 2011. Regional-scale boreal forest cover and change mapping using Landsat data composites for European Russia. *Remote Sensing of Environment,* **115**, 548-561.

PRINGLE, M.J., SCHMIDT, M., and MUIR, J.S., 2009. Geostatistical interpolation of SLC-off Landsat ETM+ images. *ISPRS Journal of Photogrammetry and Remote Sensing,* **64**, 654-664.

ROY, D.P., JU, J., KLINE, K., SCARAMUZZA, P.L., KOVALSKYY, V., HANSEN, M., LOVELAND, T.R., VERMOTE, E., and ZHANG, C., 2010. Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sensing of Environment,* **114**, 35-49.

STOREY, J., SCARAMUZZA, P., and SCHMIDT, G., 2005. Landsat 7 scan line corrector-off gap-filled product development. In, *Pecora 16 'Global Priorities in Land Remote Sensing'.* . Sioux Falls, South Dakota

TOPALOGLOU, C., MONACHOU, S., STRATI, S., ALEXANDRIDIS, T., SILLEOS, N., and MISOPOLINOS, N., 2013. Modelling LAI based on land cover map and NDVI using SPOT and TM data in two Mediterranean sites. In, *First International Conference on Remote Sensing and Geoinformation 2013*. Paphos, Cyprus

TRIGG, S.N., CURRAN, L.M., and MCDONALD, A.K., 2006. Utility of Landsat 7 satellite data for continued monitoring of forest cover change in protected areas in Southeast Asia. *Singapore Journal of Tropical Geography,* **27**, 49-66.

USGS, NASA, and LANDSAT 7 SCIENCE TEAM, 2003. Preliminary Assessment of the Value of Landsat 7 ETM+ Data following Scan Line Corrector Malfunction, http://landsat.usgs.gov/documents/SLC_off_Scientific_Usability.pdf. (Last accessed on 18/12/2012).

ZHANG, C., LI, W., and TRAVIS, D., 2007. Gaps-fill of SLC-off Landsat ETM+ satellite image using a geostatistical approach. *International Journal of Remote Sensing,* **28**, 5103-5122.