Floristic composition and across-track reflectance gradient in Landsat images over Amazonian forests

Javier Muro[a], Jasper Van doninck[a], Hanna Tuomisto[a], Mark A. Higgins[b], Gabriel M. Moulatlet[a], Kalle Ruokolainen[a,*]

[a] *Amazon Research Team, Department of Biology, University of Turku, FI-20014 Turku, Finland*
[b] *Department of Global Ecology, Carnegie Institution for Science, Stanford, CA 94305, USA*

Abstract

Remotely sensed image interpretation or classification of tropical forests can be severely hampered by the effects of the bidirectional reflection distribution function (BRDF). Even for narrow swath sensors like Landsat TM/ETM+, the influence of reflectance anisotropy can be sufficiently strong to introduce a cross-track reflectance gradient. If the BRDF could be assumed to be linear for the limited swath of Landsat, it would be possible to remove this gradient during image preprocessing using a simple empirical method. However, the existence of natural gradients in reflectance caused by spatial variation in floristic composition of the forest can restrict the applicability of such simple corrections. Here we use floristic information over Peruvian and Brazilian Amazonia acquired through field surveys, complemented with information from geological maps, to investigate the interaction of real floristic gradients and the effect of reflectance anisotropy on the observed reflectances in Landsat data. In addition, we test the assumption of linearity of the BRDF for a limited swath width, and whether different primary non-inundated forest types are characterized by different magnitudes of the directional reflectance gradient. Our results show that a linear function is adequate to empirically correct for view angle effects, and that the magnitude of the across-track reflectance gradient is independent of floristic

*Corresponding author
Email address: kalle.ruokolainen@utu.fi (Kalle Ruokolainen)

composition in the non-inundated forests we studied. This makes a routine correction of view angle effects possible. However, floristic variation complicates the issue, because different forest types have different mean reflectances. This must be taken into account when deriving the correction function in order to avoid eliminating natural gradients.

Keywords: Amazonia, BRDF, Landsat, ferns and lycophytes, floristic composition, Melastomataceae, radiometric correction, tropical forests

---

1. Introduction

Amazonian rain forests, especially the non-inundated ones, were traditionally considered rather homogeneous in terms of species composition (Encarnación, 1985; Pires and Prance, 1985; Salo et al., 1986). More recently, however, several authors have shown that plant species composition in these forests is significantly related to physical and chemical characteristics of soils at different spatial scales, from local to regional (Tuomisto et al., 1995, 2002, 2003a,b,c; Ruokolainen et al., 1997; Phillips et al., 2003; Costa et al., 2005; Duque et al., 2005; Bohlman et al., 2008). These findings have increased the need for a more detailed identification and mapping of spatial variation in plant species composition to support rainforest conservation and land use planning efforts. Interpretation of high resolution imagery, such as Landsat, is crucial for such mapping to be possible over large areas.

Remote sensing of tropical rain forests is associated with several challenges, one of the most obvious being the persistent cloud cover in these high-rainfall regions. Another challenge is that reflectance differences among forest types are often subtle. This makes it necessary to utilize the 8-bit satellite imagery (such as Landsat TM/ETM+) up to its radiometric limits, which in turn makes the interpretations more prone to errors due to radiometric distortions (Toivonen et al., 2006; Higgins et al., 2011). The importance of a careful radiometric and atmospheric correction is therefore evident.

In recent years, especially since the opening of the Landsat archive (Wulder et al., 2012), considerable effort has been put to the operational development and distribution of atmospherically corrected surface reflectance products (Roy et al., 2010; Ju et al., 2012). Because of the narrow field of view of Landsat (7.5 ° each side of nadir), the bidirectional reflectance distribution function (BRDF) effects are often ignored in the atmospheric correction, and a Lambertian behaviour of the surface is assumed (Masek et al., 2006; Ju et al., 2012; Roy et al., 2014). However, studies over Amazonian forests have demonstrated that directional scattering, which gives rise to an across-track gradient in digital number, can nullify the spectral separability of different forest types (Toivonen et al., 2006), or suggest the existence of vegetation patterns where none exist (Ruokolainen and Tuomisto, 1998).

Several authors have suggested methods for BRDF correction of Landsat imagery. Roy et al. (2008) developed a multi-temporal data fusion for Landsat and MODIS using the MODIS BRDF/Albedo land surface characterization product (Schaaf et al., 2002) to correct directional effects in Landsat. Given the vast areas covered by closed canopy, this method can be expected to be applicable over Amazonian forests in spite of the spatial resolution gap between both sensors. A more severe restriction is the persistent cloud cover in tropical areas. Generation of MODIS BRDF parameters assumes the availability of several cloud-free observations during a 16-day period. Even in the case of a cloudfree Landsat acquisition, the required number of MODIS observations during the adjacent days may not be reached. Furthermore, no MODIS BRDF parameters are available for the period before the start of MODIS operation in 2000. A different approach was suggested by Flood et al. (2013) for Landsat TM/ETM+ imagery over Eastern Australia. This method exploits the availability of multi-temporal acquisitions over this region to calculate BRDF parameters at the Landsat spatial resolution. Consequently, it is not applicable in areas with only occasional cloud-free conditions.

As an alternative to multi-sensor or multi-temporal approaches, empirical scene-based normalizations have been suggested by some authors (Toivonen et al., 2006; Hansen et al., 2008). These derive the relationship between the position of a pixel in the image and its digital number or reflectance through linear regression, and consecutively use this relationship to normalize the across-track gradient. This method relies on a number of assumptions. Firstly, it is assumed that the BRDF, and the brightness variation it gives rise to, is linear over a uniform land surface in Landsat's narrow field of view. Secondly, it is assumed that there are no east-west gradients in the imagery other than those caused by directional reflectance effects. Hansen et al. (2008) aimed to ensure this by deriving the regression coefficients only for forested pixels (as inferred from a MODIS forest mask), rather than for all pixels in the scene. However, not all forests are identical, and gradual or abrupt changes in floristic composition may exist and affect reflectance properties even within forested areas, and this could invalidate the empirical correction.

In this study, we investigate the reliability of empirical across-track reflectance gradient corrections for two regions in Peruvian and Brazilian Amazonia. First, we use in situ data on floristic composition to test if adequate corrections of Landsat TM/ETM+ imagery is obtained using simple linear models, and if variation in species composition of the forest confounds the correction. Then we test an earlier suggestion (Toivonen et al., 2006) that the radiometric gradient has different magnitudes over different kinds of forest in Amazonia. If different corrections are needed for different kinds of forest, the demand for ground truth data would increase and routine radiometric correction be severely hampered. We use a large image dataset combined with geological data (which acts as proxy for forest species composition) to quantify how much error is introduced if compositional variation in the forest is ignored. This will help in assessing whether or not these relatively simple, empirical angular normalization techniques are appropriate for a given application.

2. Study area and data sets

2.1. Study area

This study was conducted in tropical rain forests of the Amazonian lowlands in northern Peru and western Brazil (Fig. 1). The area is covered largely by undisturbed primary *terra firme* (non-inundated) forest under closed canopy. Areas that are sporadically, seasonally or permanently inundated occur along rivers. Climate is tropical, humid, and almost aseasonal. In the city of Iquitos, which is situated near the Peruvian study area, the mean monthly temperature is 25–27 °C throughout the year and annual precipitation is approximately 3100 mm (Marengo, 1998). The city of Eirunepé, near the Brazilian site on the Juruá River, experiences a mean monthly temperature of approximately 25 °C and an annual precipitation of 2195 mm (*http://inmet.gov.br/portal/*). Elevation ranges from 100 to 250 m above sea level in most of the study area, with a few precipitously hilly areas exceeding 400 m.

Two geological formations are exposed at the surface over most of the study area. These are known in Peru as the Pebas Formation and the Nauta Formation, and in Brazil as the Solimões Formation and the Içá Formation, respectively. For clarity, only the Peruvian names will be used here. The Pebas Formation consists of poorly weathered, relatively cation-rich (by Amazonian standards) clay sediments that were deposited under low-energy semi-marine or lacustrine conditions of the Pebas Embayment. The Nauta formation consists of more weathered, cation-poor sediments with coarser texture that were deposited on top of the Pebas Formation under higher-energy deltaic to fluvial conditions (Räsänen et al., 1995; Rebata-H. et al., 2006). The cation concentration in the soils derived from the Pebas Formation is about one order of magnitude higher than that in the Nauta Formation. This difference in soils is reflected in the species composition of the primary *terra firme* forest, and gives rise to a plant species turnover of 80–90 % across the geological boundary (Higgins et al., 2011). A third distinct unit of the *terra firme* landscape is formed by river terraces of

sedimentary deposits from Andean origin dated from mid/late Pleistocene. In terms of sedimentation environment and soil nutrient content, the terraces are similar to the Nauta formation. The main differences between the two is that terraces are practically flat in topography (Irion and Kalliola, 2009).

The digitized geological maps used in this study were obtained from the Instituto Geológico Minero y Metalúrgico (INGEMMET, *http://www.ingemmet.gob.pe*) and the Geological Survey of Brazil (CPRM, *http://geobank.sa.cprm.gov.br*). The boundaries between the Pebas and Nauta Formations in these maps were modified according to the discontinuity identified in Landsat imagery and Shuttle Radar Topography Mission (SRTM) elevation data (Higgins et al., 2011).

2.2. Floristic data

The surface reflectance of dense forests, as observed by satellite sensors operating in the optical domain, is mostly determined by properties of the forest canopy, including tree species composition. Because collection of field data on canopy species composition in these high-diversity forests is difficult and time consuming, approaches have been developed to use floristic variation in more easily observable understory plant groups as indicators of floristic variation in canopy trees (Ruokolainen et al., 1997, 2007). Here we use a floristic dataset of two such indicator groups, pteridophytes (ferns and lycophytes) and the Melastomataceae. Both groups have been found to reproduce well the floristic patterns observed in trees (Tuomisto et al., 1995; Ruokolainen et al., 1997, 2007). The data have been inventoried using a standard procedure (Tuomisto et al., 2003a), where presence-absence of different species was recorded along 500 m long and 5 m wide transects. Two sets of such transects are used here: 105 transects in Peru, inventoried in 2005–2006 and 45 transects in Brazil, inventoried in 2012. Both sets were purposefully placed to sample the Pebas/Nauta boundary, and eleven of the Brazilian transects sample terraces of the Juruá and Tarauacá rivers (Fig. 1). The Peruvian dataset is described in more detail in (Higgins et al., 2011) and (Higgins et al., 2012).

2.3. Landsat data

Two sets of Landsat TM/ETM+ images are used here. The first consists of two image mosaics, one covering the Peruvian and the other covering the Brazilian field inventory sites (Fig. 1). The five selected and visually inspected cloud-free Level 1T images used to produce these mosaics (Table 1) were obtained through the USGS EarthExplorer website (*http://earthexplorer.usgs.gov/*). For the Brazilian study area, the two images were acquired during the same overpass, so they could be easily combined into a seamless mosaic. For the Peruvian site, two images were obtained during the same overpass, but the third was acquired 17 days, and seven years, later and was normalized with respect to the other two images using histogram matching in the area of overlap. The seasonal 17 day gap between the acquisitions is sufficiently small to ensure that vegetation throughout the mosaic is at the same phenological state. Due to the persistent cloud cover, there were no images available from these adjacent paths with a gap smaller than seven years, with such a small seasonal difference. Also, there is a time gap between the acquisition of the images and the field data collection for both sites. However, since the study sites are in undisturbed forests, it can be expected that no changes in floristic composition occurred during this period.

Table 1: Landsat L1T images used to construct the mosaics

Area	Path	Row	Date	Sensor	Azimuth	Solar elevation
Peru	7	62	21/08/1999	ETM+	62.6001968	56.7000504
Peru	7	63	21/08/1999	ETM+	60.8536224	55.8320427
Peru	8	62	08/09/2006	TM	73.3150337	60.4587121
Brazil	3	64	04/08/2006	TM	52.5639618	51.7608786
Brazil	3	65	04/08/2006	TM	51.3322926	50.6973778

The second image dataset comprises 45 predominantly cloud-free Landsat TM/ETM+ images acquired between 1987 and 2011 throughout the year. Sev-

enteen, 14 and 14 acquisitions were used for the scenes with path/row combination 003/064, 007/062 and 008/062, respectively. These three scenes cover a boundary of the Pebas and Nauta formations in Brazil and northern Peru. Contrary to the two mosaics described above, for which the original Landsat images were used, we here downloaded the Landsat Surface Reflectance Climate Data Record (CDR) images through EarthExplorer portal.

These are generated using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS, version 2.1.0) algorithm which includes conversion of radiance to reflectance values and application of the MODIS atmospheric correction (Masek et al., 2006). As a result, reflectance differences resulting from varying atmospheric conditions between images acquired during different seasons are minimised. The CDR images are accompanied by a number of automatically generated mask bands. Pixels that were flagged in any of the bands as cloud, cloud shadow, cloud buffer, or water were excluded from further analyses. Seasonally inundated forests were also masked based on the available geological maps.

3. Methods

3.1. Quantification of floristic gradient

The field inventory resulted, for each 5 m by 500 m transect, in a listing of the presence or absence of species of ferns, lycophytes and Melastomataceae. Typically plant species occurrences are correlated with each other so that one can reduce the dimensionality of the data table by finding eigenvectors that summarize the intercorrelated descriptors (species). A simple principal components analysis is, however, not an ecologically suitable method because both ecological theory and empirical data show that species abundances have unimodal responses to environmental variables (Legendre and Legendre, 1998). Therefore we first calculated an ecologically meaningful species compositional distance (one-complement of Jaccard index of similarity (Jaccard, 1912)) between each pair of transects separately for the Peruvian and Brazilian transects. Then

we used non-metric multidimensional scaling analysis (NMDS) (Legendre and Legendre, 1998) for finding the best possible unidimensional representation of the multidimensional variation in each of the two data matrices. NMDS is an iterative method that aims to map objects (transects) into a specified number of dimensions such that the rank order of the distances as measured from the ordination solution are as similar as possible to those in the original dissimilarity matrix. The NMDS scores obtained for the Peruvian dataset have previously been observed to closely correlate with soil cation concentration (Higgins et al., 2011, 2012).

3.2. Testing the linearity and forest-type independence of BRDF

In order to test both the linearity of the BRDF for the limited swath width of Landsat and the forest-type independence of the empirically observed reflectance gradient, we fitted a cubic trend surface (Borcard et al., 1992) to the floristic data and remotely sensed data of the sampling sites:

$$DN = a + b\Gamma + c\nu + d\Gamma\nu + e\Gamma^2 + f\nu^2 + g\Gamma\nu^2 + h\Gamma^2\nu + i\Gamma^3 + j\nu^3 \ , \qquad (1)$$

where DN is the Landsat digital number corresponding to the sampling site, Γ is the column number corresponding to the site, and ν is the NMDS-score of the site (all unitless).

In order to obtain a representative DN for each 5 m by 500 m field sampling transect, we first made a multiband segmentation on each Landsat mosaic (Fig. 2). The segmentation was performed in eCognition v8.8, with parameters Scale=20, Shape=0.4, and Compactness=0.4 (all unitless). A representative DN of each band was then calculated the as a weighted average of the DN means of segments that the transect traversed. The weight of each segment was the proportional length of the transect overlying the segment. Through this procedure we acknowledge that local edaphic variation can control plant species composition at spatial resolutions smaller than our 500-m-long sampling transects (Tuomisto et al., 1995; Poulsen et al., 2006), and that edaphically induced

floristic variation is detectable in DN values (Tuomisto et al., 1995, 2003a; Higgins et al., 2011; Thessler et al., 2005). Segmentation also removes undesired noise that can result from small-scale topographic features, while keeping the directional component unaffected. Column number Γ was calculated as the perpendicular distance from the centroid of the segment to the western edge of the Landsat mosaic.

In the first step, only the two first terms of Eq. 1 were used to build the model. This model captures the proportion in DN variance that is linearly explained by Γ. In the next step, the term in ν was added to the model to capture the proportion of DN variance that is explained by column number and NMDS-score together but without any interaction between the two. Finally, the remaining terms were added step-wise into the model, starting from the ones with the largest adjusted coefficient of determination ($\bar{R}^2$) and proceeding to those with smaller $\bar{R}^2$ until the next one to be added was not statistically significant ($p = 0.05$). The assumption of linearity can be accepted if the addition of higher order polynomial terms containing the column number Γ does not increase the adjusted coefficient of determination. On the other hand, if adding first or higher order terms containing NMDS score ν increases $\bar{R}^2$, then floristic variation within the rain forest can be concluded to influence the empirical estimation of the across-track reflectance gradient. This analysis was performed separately for the Peruvian and Brazilian study areas and for each spectral band.

3.3. Comparison of reflectance gradient for different geological formations

The magnitude of the reflectance gradient is influenced by the optical properties of the surface. Different species of the forest canopy may have different canopy structures, leaf angles and various leaf and bark characteristics that affect reflectance. Therefore, floristically distinct forests may present different reflectance anisotropy and magnitudes of the reflectance gradient. To test whether this is the case, the difference in the magnitude of the gradient be-

tween *terra firme* forests on the Pebas and Nauta formations was analysed for the 45 Landsat surface reflectance images. No field data was available to verify the existence of floristic difference between forests on the Pebas and Nauta formations in these images, so the interpretations of the results are based on two assumptions: 1) the limit between the two formations is correctly shown in the available geological maps, and 2) the floristic patterns follow the limit between the geological formations in the same way as has been documented for other areas. To reduce the uncertainty related to the first point, each image was visually screened and transitional areas between the two geological formations and between non-inundated and inundated areas were manually masked and excluded from the analysis. For each image, and for each spectral band, the slope b (unitless) of the reflectance gradient was derived as follows:

$$\rho = a + b\Gamma \ , \tag{2}$$

where ρ is the surface reflectance of the pixel. In each of the 45 Landsat images, 20,000 randomly selected pixels were placed over forested areas, half on the Pebas Formation and half on the Nauta Formation. Additionally, the reflectance gradient was calculated for each image without stratification into Pebas and Nauta, but with masking of other formations such as terraces or floodplain forests.

Empirical reflectance gradients over tropical forests have been found to vary throughout the year (Hansen et al., 2008). These variations may be caused by changes in sun-sensor-geometry, seasonally varying atmospheric conditions or vegetation phenology. Amazonian tropical forests are characterized by moderate phenological cycles (Silva et al., 2013), driven by seasonal changes in temperature and precipitation. It is not known to what extent Amazonian vegetation phenology influences directional scattering for the range of Landsat view angles, and this is beyond the scope of this research. However, it can be assumed that, within the limited extent of a single Landsat scene, spatial variability of temperature and rainfall is low. Phenological cycles of different types of *terra firme* will therefore be similar. We derive and compare the empirical reflectance gradient

for the Pebas and Nauta formation for each of the 45 Landsat images separately. The influence of varying sun-sensor geometry, seasonally varying atmospheric conditions and forest phenology will be similar for the two formations in each of the individual 45 comparisons. Therefore it is meaningful to analyse if the reflectance gradient consistently differs between the two formations.

4. Results

4.1. Floristic gradient

Fig. 3 shows scatterplots of the unidimensional NMDS scores versus column number for the Peruvian and Brazilian sites. The NMDS scores of the sites on the Pebas Formation are generally higher than those of sites on the Nauta Formation. In addition, in both image mosaics the NMDS scores decrease from west to east within each geological formation, and also within the terrace surface in Brazil. This indicates that there is an east-west trend in floristic composition, which is associated with a similar spatial trend in soil cation concentration (Higgins et al., 2011, 2012; Tuomisto). There is also a general geologically controlled trend in Amazonian rain forest area of decreasing soil fertility from west to east (Räsänen et al., 1995; Hoorn et al., 2010).

How digital numbers corresponding to each transect relate to column number and NMDS score is shown in Fig. 4 for the Peruvian transects , and Fig. 5 for the Brazilian. The value of the third variable (NMDS score for the Γ-DN plot, column number for the ν-DN plot) is indicated by colours. These figures show that erroneous conclusions may be drawn when only one explanatory variable is taken into account. E.g., a strong possible correlation between NMDS score and digital number could be suggested when anisotropy is ignored. In reality, however, the increase in NMDS score with increasing column number is only limited —if present at all— for a given small range of column numbers. The decrease of digital number with increasing column number, on the other hand, is the result of an interplay of effects of anisotropy and floristic composition.

4.2. Reflectance gradient modelling

Adjusted coefficients of determination of the iterative application of Eq. 1 are given in Table 2. Regressions with pixel value as the dependent variable and the first order term of column number as the only explanatory variable gave consistently high $\bar{R}^2$ values. Adding the first order term of the NMDS score did not improve the $\bar{R}^2$ for visible light bands in either study area, but for the infrared bands it did. Then $\bar{R}^2$ increased most for band 4 in Peru (0.102) and for band 7 in Brazil (0.051); for in the other bands, increase in $\bar{R}^2$ was detectable but rather small (0.017–0.022). Adding higher order terms of the NMDS score increased $\bar{R}^2$ little or not at all, except for band 4 in Brazil, where the increase was 0.154.

4.3. Pebas/Nauta difference of the reflectance gradient

An empirical reflectance gradient for the forests on the two geological formations (Pebas and Nauta) was obtained by regressing surface reflectance against column number (Eq. 2) using 10,000 random points for each geological formation. This was done separately for each of the 45 Landsat images and each spectral band. Mean slopes and standard deviations of these gradients are given in Table 3, together with the differences between the mean slopes on the Pebas and Nauta formations. A paired t-test was performed to check if significant differences between the slopes on the two formations exist. Because the Shapiro-Wilks test for normality indicated that 10 out of 24 Pebas Formation datasets and 4 out of 24 Nauta Formation datasets were not normally distributed (at p=0.05), the non-parametric Wilcoxon signed-rank test was also used. Results of the t-test and the Wilcoxon test were almost identical, however. Both revealed significant differences (at p=0.05) between the mean slopes in the infrared bands for scenes 003/064 and 008/062. In scene 003/064, the reflectance gradient is stronger (more negative) for the Pebas than for the Nauta Formation which gives a negative difference, but in scene 008/062 the situation was reversed and the difference was positive. As a result, differences for all data

pooled are much smaller, and not significant for either test at p=0.05.

In Table 4, the slopes obtained for the Pebas and Nauta formation are combined to calculate a stratified reflectance gradient. This is simply done by, for each image, taking the average of the slopes as measured over the Pebas and Nauta formations separately. The stratified slopes are then compared to the unstratified slopes, which were obtained by linear regression of digital number versus column number for pixels over both formations combined. A significant difference between the stratified and unstratified slopes can be observed for the green bands and all infrared bands, for each of the image scenes. The sign of this difference is opposite for the Brazilian scene 003/064 with respect to the two Peruvian scenes. Differences between stratified and unstratified slopes are linked to the relative spatial pattern of the two geological formations and the associated forest cover. In the Brazilian scene, the Nauta Formation (with forests of lower reflectance in the infrared) dominates the eastern side of the image and the Pebas Formation (with forests of higher reflectance) the western side. In the Peruvian scenes, the situation is the opposite. In addition, the geological interface is oriented in a northwest-southeast direction in the Peruvian site but north-south (almost perpendicular to the Landsat scan direction) in the Brazilian site. This results in a much larger difference between the stratified and unstratified slopes in the Brazilian scenes.

5. Discussion

Our analyses showed that the single most important predictor of the digital numbers corresponding to in situ sampling sites in *terra firme* forests was the east-west position of the site in a Landsat mosaic. The largest part of the variability in DN was explainable as a linear function of column number (Γ). This was the case for all spectral bands and in both study areas (Brazil and Peru). The increase in $\bar{R}^2$ when higher-order polynomial terms of column number were added to the model was small or statistically insignificant. This result suggests

that a linear relationship between column number —which represents sensor view angle— and image digital number —which represents surface reflectance— can be used to correct for bidirectional reflection distribution function (BRDF) in Landsat images over Amazonian *terra firme* forests. The same principle may be applicable to other high resolution imagery with a similar swath width.

It should be noted that the topography in our study area is relatively flat. At the local scale, the slopes of the hills can be steep, but as the hills themselves rarely exceed 50 m in height, their effect on surface reflectance is smaller than in areas of truly accented topography. There is a discontinuity in this respect between the Pebas and Nauta Formation, as the former typically has flatter topography than the latter. With increasing topographic relief, the range of possible sun-terrain-sensor configurations will increase. At some point it will then become necessary to apply more complex topographic/BRDF corrections, and these may be non-linear (Li et al., 2012; Flood et al., 2013).

We found that digital numbers of the blue, green and red Landsat bands were unrelated to floristic composition, as represented by the NMDS scores (ν). This is in accordance with the findings of Higgins et al. (2012) who stated that "no information about floristic patterns is lost by excluding bands 1–3 from image display and manual image interpretation". Data in the visible bands can, however, assist in discriminating primary forest from other land cover types (Sesnie et al., 2010).

In the near and shortwave infrared bands, addition of NMDS scores (which represent position along the main floristic gradient observed in the forests) in the iterative regression analysis increased $\bar{R}^2$ in both study areas. In other words, differences in floristic composition can be observed using the relatively limited spectral and radiometric resolution of the Landsat TM/ETM+ sensors. This has implications for simple, scene-based empirical BRDF correction methods such as those suggested by (Hansen et al., 2008). It was shown in Fig. 3 that a

floristic east-west gradient exists in both study sites. This creates a natural, real trend of reflectance values in the near and shortwave infrared wavelengths that coincides with the reflectance gradient resulting from reflectance anisotropy. Such covariation can result in an over- or underestimation of the magnitude of either one gradient if it is studied without considering the other.

When the empirical reflectance gradient was derived over forests covering the Pebas and Nauta formation separately, pooling all the images, no significant difference between their slopes could be found for any of the visible or infrared bands. Significant differences in the infrared bands were detected for two of the three scenes, albeit with different signs. These may result from scene-specific properties such as subtle topographic or floristic patterns within a single geological formation, residual atmospheric contamination in some images, or unmasked clouds and cloud shadows. In conclusion, even though plant species compositions on the two formations are very different, this seems to have little or no effect on the directional scattering properties of the vegetation canopy for the limited swath width of the Landsat sensor.

Even though reflectance gradients obtained for the Pebas and Nauta formation are quite similar, it is important to notice that there is a difference in the absolute reflectance of the forest types covering these formations. If such floristic patterns are not taken into account when calculating an empirical reflectance gradient for a Landsat scene, the obtained model may differ significantly from what is needed for a correct normalization of viewing angle effects. In this study, the maximum difference between the unstratified and the stratified reflectance gradient was approximately 3×10^{-6}, or more than 50 % of the stratified gradient. A single Landsat image has a width of approximately 8000 columns. When using the unstratified instead of the more reliable stratified reflectance gradient for normalization to zenith view angle (maximum 4000 columns), an error in corrected surface reflectance of up to 1.2 % is introduced in the NIR band. This can be enough to hamper image interpretation or classification when studying

subtle floristic patterns.

The magnitude of the difference between stratified and unstratified slopes can be expected to be related to the orientation of the boundary between forest types. To remove the reflectance gradient from Landsat images over regions characterized by a north-south-oriented interface between geological/floristic formations, an empirical gradient should be determined after stratification of the satellite image according to geology. Since no consistent significant difference in the slope was found for the two different formations, the average value over these strata could be used to normalize the entire scene. In the case of east-west-oriented boundaries between formations, such stratification is not necessary, since the difference between the stratified and unstratified reflectance gradients will be small. Effects of gradual vegetation changes within a single geological formation are not accounted for using this approach. These can be expected to be smaller than those between across-formations effects (Fig. 3), and in addition their verification would require more extensive field data.

6. Conclusions

This paper aimed at investigating the validity of empirical, linear, scene-based view angle normalization approaches for Landsat data over Amazonian rainforest. Such a radiometric correction is indispensable for the correct classification of different primary *terra firme* forest types, which are radiometrically very similar. We investigated the interrelation between floristic gradients and reflectance gradients introduced by the bidirectional reflection distribution function with the help of an extensive dataset of in situ field observations of indicator species. Results showed that the reflectance of a pixel is largely defined by its position in the Landsat image in the east-west direction, and that this relationship can be considered to be linear for the limited swath width of Landsat TM/ETM+. For the infrared bands, but not for the visible light bands, surface reflectance was also related to floristic composition. Therefore, empirical view

angle corrections need to take the floristic variation in the forests into account to avoid producing erroneous results. A second part of this study investigated whether different forest types, identified by the geologic formation on which they grow, are characterized by a similar magnitude of the BRDF-induced reflectance gradient. No consistent differences could be found for the reflectance gradient in any spectral band.

Extensive floristic datasets, as the one used in this study, are generally not available for image preprocessing over extensive areas. Whether simple regression-based angular corrections that ignore floristic variability are appropriate depends on the application. If the aim of a study is to map deforestation, the radiometric errors introduced by ignoring floristic variation are probably trivial. If the aim is to map the floristic variation itself, the errors can be detrimental. A possible improvement for empirical view angle normalization methods could be to first stratify the images based on geological data, which can serve as a rough proxy for vegetation composition, and deriving view angle correction coefficients based on these strata.

Acknowledgments

The study was technically supported by the University of Turku Laboratory of Geoinformatics (UTU-LCC). We thank the Peruvian and Brazilian authorities and the local villages for permission to carry out field work, and Universidad Nacional de la Amazonía Peruana (UNAP, Iquitos - Peru), Instituto de Investigaciones de la Amazonía Peruana (IIAP, Iquitos - Peru) and Instituto Nacional de Pesquizas da Amazônia (INPA, Manaus - Brazil) for logistic support. We are grateful to numerous persons who collaborated in the field, especially Glenda Cárdenas and Nelly Llerena. The study was funded by Academy of Finland grants to HT. Financial support for MH was provided by the National Science Foundation, the American-Scandinavian Foundation, and Duke University. At the time of field work in Brazil, GM was based at INPA.

References

Bohlman, S.A., Laurance, W.F., Laurance, S.G., Nascimento, H.E., Fearnside, P.M., Andrade, A., 2008. Importance of soils, topography and geographic distance in structuring central Amazonian tree communities. Journal of Vegetation Science 19, 863–874.

Borcard, D., Legendre, P., Drapeau, P., 1992. Partialling out the spatial component of ecological variation. Ecology 73, 45–1055. doi:10.2307/1940179.

Costa, F.R.C., Magnusson, W.E., Luizao, R.C., 2005. Mesoscale distribution patterns of amazonian understorey herbs in relation to topography, soil and watersheds. Journal of Ecology 93, 863–878.

Duque, A.J., Duivenvoorden, J.F., Cavelier, J., Sánchez, M., Polanía, C., León, A., 2005. Ferns and Melastomataceae as indicators of vascular plant composition in rain forests of Colombian Amazonia. Plant Ecology 178, 1–13.

Encarnación, F., 1985. Introducción a la Flora y Vegetación de la Amazonía peruana: estado actual de los estudios, medio natural y ensayo de claves de la determinación de las formaciones vegetales en la llanura amazónica. Candollea 40, 237–252.

Flood, N., Danaher, T., Gill, T., Gillingham, S., 2013. An operational scheme for deriving standardised surface reflectance from Landsat TM/ETM+ and SPOT HRG imagery for eastern australia. Remote Sensing 5, 83–109. doi:10.3390/rs5010083.

Hansen, M.C., Roy, D.P., Lindquist, E., Adusei, B., Justice, C.O., Altstatt, A., 2008. A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. Remote Sensing of Environment 112, 2495–2513. doi:10.1016/j.rse.2007.11.012.

Higgins, M.A., Asner, G.P., Perez, E., Elespuru, N., Tuomisto, H., Ruokolainen, K., Alonso, A., 2012. Use of Landsat and SRTM data to detect broad-scale

biodiversity patterns in Northwestern Amazonia. Remote Sensing 4, 2401–2418. doi:10.3390/rs4082401.

Higgins, M.A., Ruokolainen, K., Tuomisto, H., Llerena, N., Cardenas, G., Phillips, O.L., Vásquez, R., Räsänen, M., 2011. Geological control of floristic composition in Amazonian forests. Journal of Biogeography 38, 2136–2149. doi:10.1111/j.1365-2699.201.02585.x.

Hoorn, C., Wesselingh, F., ter Steege, H., Bermudez, M., Mora, A., Sevink, J., Sanmartín, I., Sanchez-Meseguer, A., Anderson, C., Figueiredo, J., Jaramillo, C., Riff, D., Negri, F., Hooghiemstra, H., Lundberg, J., Stadler, T., Särkinen, T., Antonelli, A., 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science 330, 927–931.

Irion, G., Kalliola, R., 2009. Long-term landscape development processes in Amazonia, in: Hoorn, C., Wesselingh, F.P. (Eds.), Amazonia: Landscape and Species Evolution: A look into the past. Wiley-Blackwell Publishing Ltd., p. 185.

Jaccard, P., 1912. The distribution of flora in the alpine zone. New Phytologist 11, 37–50.

Ju, J., Roy, D.P., Vermote, E., Masek, J., Kovalskyy, V., 2012. Continental-scale validation of MODIS-based and LEDAPS Landsat ETM+ atmospheric correction methods. Remote Sensing of Environment 122, 175–84. doi:10.1016/j.rse.2011.12.025.

Legendre, P., Legendre, L., 1998. Numerical ecology. 2nd ed., Elsevier, Amsterdam.

Li, F., Jupp, D.L.B., Thankappan, M., Lymburner, L., Mueller, N., Lewis, A., Held, A., 2012. A physics-based atmospheric and BRDF correction for Landsat data over mountainous terrain. Remote Sensing of Environment 124, 756–770. doi:10.1016/j.rse.2012.06.018.

Marengo, J.A., 1998. Climatología de la zona de Iquitos, Perú. Geología y desarrollo Amazónico, estudio integrado de la zona de Iquitos, Perú , 35–57.

Masek, J.G.M., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T., 2006. A Landsat surface reflectance dataset for North America, 1990–2000. IEEE Geoscience and Remote Sensing Letters 3, 68–72. doi:10.1109/lgrs.2005.857030.

Phillips, O.L., Núñez Vargas, P., Monteagudo, A.L., Cruz, A.P., Zans, M.E.C., G., S.W., Yli-Halla, M., Rose, S., 2003. Habitat association among amazonian tree species: a landscape-scale approach. Journal of Ecology 91, 757–775. doi:10.1046/j.1365-2745.2003.00815.x.

Pires, J.M., Prance, G.T., 1985. The vegetation types of the Brazilian Amazon, in: Prance, G.T., Lovejoy, T.E. (Eds.), Key enironments: Amazonia. Pergamon, Oxford, pp. 109–145.

Poulsen, A.D., Tuomisto, H., H., B., 2006. Edaphic and floristic variation within a 1-ha plot of lowland Amazonian rain forest. Biotropica 38. doi:10.1111/j.1744-7429.2006.00168.x.

Räsänen, M.E., Linna, A.M., Santos, J.C.R., Negri, F.R., 1995. Late Miocene tidal deposits in the Amazonian foreland basin. Science 269, 386–390. doi:10.1126/science.269.5222.386.

Rebata-H., L.A., Gingras, M.K., Rasanen, M.E., Barberi, M., 2006. Tidal-channel deposits on a delta plain from the Upper Miocene Nauta Formation, Maranon Foreland Sub-basin, Peru. Sedimentology 53, 971–1013. doi:10.1111/j.1365-3091.2006.00795.x.

Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M., Loveland, T.R., Vermote, E., Zhang, C., 2010. Web-enabled Landsat Data (weld): Landsat ETM+ composited mosaics of the conterminous United States. Remote Sensing of Environment 114, 35–49. doi:10.1016/j.rse.2009.08.011.

Roy, D.P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., Lindquist, E., 2008. Multi-temporal MODIS-Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. Remote Sensing of Environment 112, 3112–3130. doi:10.1016/j.rse.2008.03.009.

Roy, D.P., Qin, Y., Kovalskyy, V., Vermote, E.F., Ju, J., Egorov, A., Hansen, M.C., Kommareddy, I., Yan, L., 2014. Conterminous United States demonstration and characterization of MODIS-based Landsat ETM+ atmospheric correction. Remote Sensing of Environment 140, 433–449. doi:10.1016/j.rse.2013.09.012.

Ruokolainen, K., Linna, A., Tuomisto, H., 1997. Use of Melastomataceae and pteridophytes for revealing phytogeographical patterns in Amazonian rain forests. Journal of Tropical Ecology 13, 243–256.

Ruokolainen, K., Tuomisto, H., 1998. Vegetación natural de la zona de Iquitos, in: Kalliola, R., Flores, S. (Eds.), Geoecología y desarrollo amazonico: estudio integrado en la zona de Iquitos, Perú. Annales Universitatis Turkuensis, Ser A II, vol 114, pp. S253–S365.

Ruokolainen, K., Tuomisto, H., Macía, M.J., Higgins, M.A., Yli-Halla, M., 2007. Are floristic and edaphic patterns in Amazonian rain forests congruent for trees, pteridophytes and Melastomataceae? Journal of Tropical Ecology 23, 13–25. doi:10.1017/S0266467406003889.

Salo, J., Kalliola, R., Häkkinen, I., Mäkinen, Y., Niemelä, P., Puhakka, M., Coley, P.D., 1986. River dynamics and the diversity of Amazon lowland forest. Nature 322, 254–258.

Schaaf, C.B., Gao, F., Strahler, A.H., Lucht, W., Li, X.W., Tsang, T., Strugnell, N., Zhang, X.Y., Jin, Y.F., Muller, J.P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entremont, R.P., Hu, B.X., Liang, S.L., Privette, J.L., Roy, D., 2002. First operational BRDF, albedo and nadir reflectance products from MODIS. Remote Sensing of Environment 83, 135–148. doi:10.1016/S0034-4257(02)00091-3.

Sesnie, S.E., Finegan, B., Gessler, P.E., Thessler, S., Ramos Bendana, Z., Smith, A.M.S., 2010. The multispectral separability of Costa Rican rainforest types with support vector machines and Random Forest decision trees. International Journal of Remote Sensing 31, 2885–2909. doi:10.1080/01431160903140803.

Silva, F.B., Shimabukuro, Y.E., Aragão, L.E.O.C., Anderson, L.O., Pereira, G., Cardozo, F., Arai, E., 2013. Large-scale heterogeneity of Amazonian phenology revealed from 26-year long AVHRR/NDVI time-series. Environmental Research Letters 8, 024011. doi:10.1088/1748-9326/8/2/024011.

Thessler, S., Ruokolainen, K., Tuomisto, H., Tomppo, E., 2005. Mapping gradual landscape-scale floristic changes in Amazonian primary rain forests by combining ordination and remote sensing. Global Ecology and Biogeography 14, 315–325. doi:10.1111/j.1466-822x.2005.00158.x.

Toivonen, T., Kalliola, R., Ruokolainen, K., Malik, R.N., 2006. Across-path DN gradient in Landsat TM imagery of Amazonian forests: A challenge for image interpretation and mosaicing. Remote Sensing of Environment 100, 550–562. doi:10.1016/j.rse.2005.10.006.

Tuomisto, H., . Unpublished.

Tuomisto, H., Poulsen, A.D., Ruokolainen, K., Moran, R.C., Quintana, C., Celi, J., Cañas, G., 2003a. Linking floristic patterns with soil heterogeneity and satellite imagery in Ecuadorian Amazonia. Ecological Applications 13, 352–371. doi:10.1890/1051-0761(2003)013[0352:LFPWSH]2.0.CO;2.

Tuomisto, H., Ruokolainen, K., Aguilar, M., Sarmiento, A., 2003b. Floristic patterns along a 43-km long transect in an Amazonian rain forest. Journal of Ecology 91, 743–756. doi:10.1046/j.1365-2745.2003.00802.x.

Tuomisto, H., Ruokolainen, K., Kalliola, R., Linna, A., Danjoy, W., Rodriguez, Z., 1995. Dissecting Amazonian biodiversity. Science 269, 63–66. doi:10.1126/science.269.5220.63.

Tuomisto, H., Ruokolainen, K., Poulsen, A.D., Moran, R.C., Quintana, C., Canas, G., Celi, J., 2002. Distribution and diversity of pteridophytes and melastomataceae along edaphic gradients in Yasuni National Park, Ecuadorian Amazonia. Biotropica 34, 516–533. doi:10.1111/j.1744-7429.2002.tb00571.x.

Tuomisto, H., Ruokolainen, K., Yli-Halla, M., 2003c. Dispersal, environment, and floristic variation of western Amazonian forests. Science 299, 241–244. doi:10.1126/science.1078037.

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. Remote Sensing of Environment 122, 2–10.

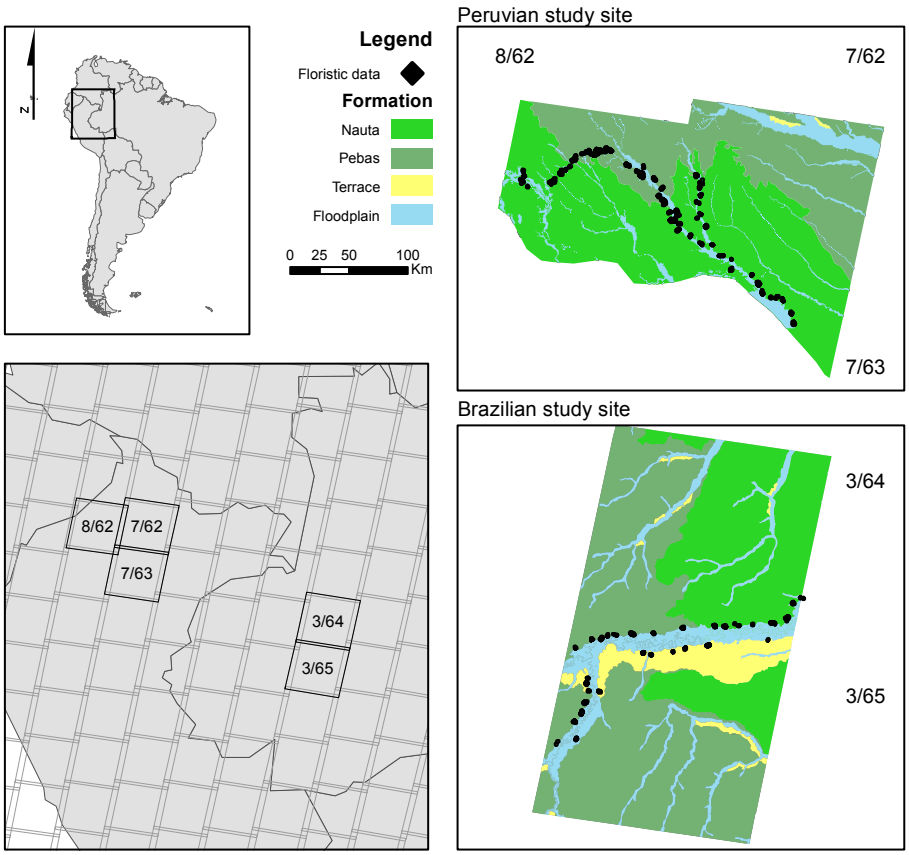

Figure 1: Locations of the Landsat images used, with indication of path/row combination, the main geological formations in the study area, and the sampling sites where floristic information was collected.

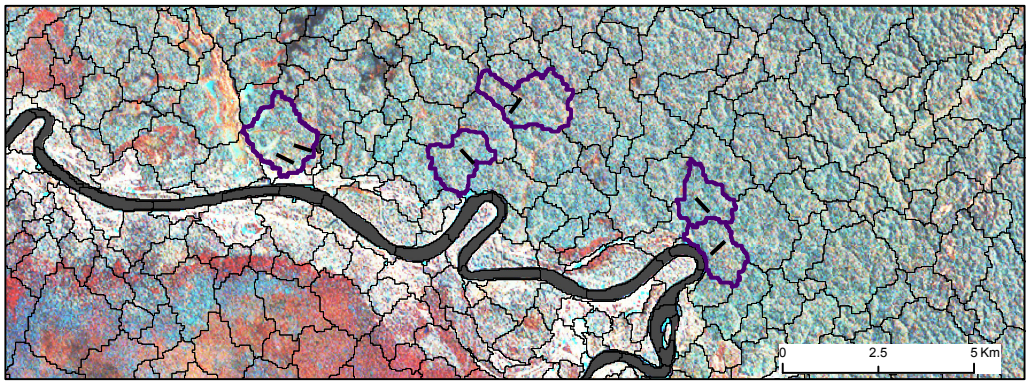

Figure 2: Example of image segmentation for the Peruvian study site (red=band 4, blue=band 5, green=band 7). Straight black lines indicate the field sampling transects and blue polygons the segments used to derive digital numbers corresponding to each transect.

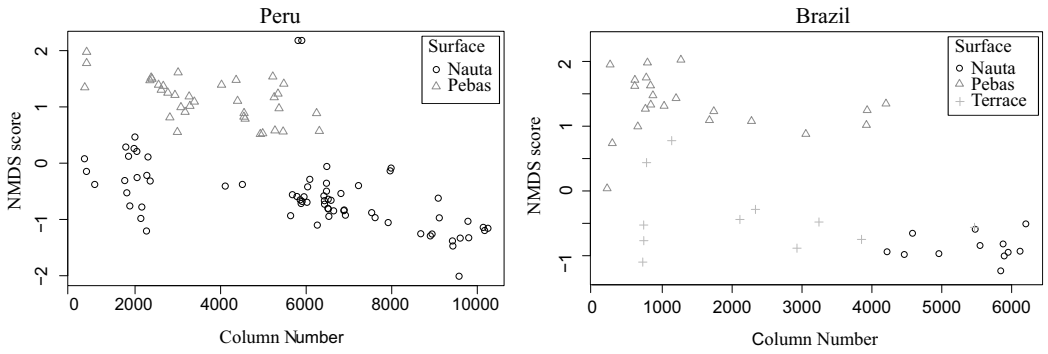

Figure 3: NMDS score versus image column number for the Peruvian and Brazilian study sites, with indication of geological formation.

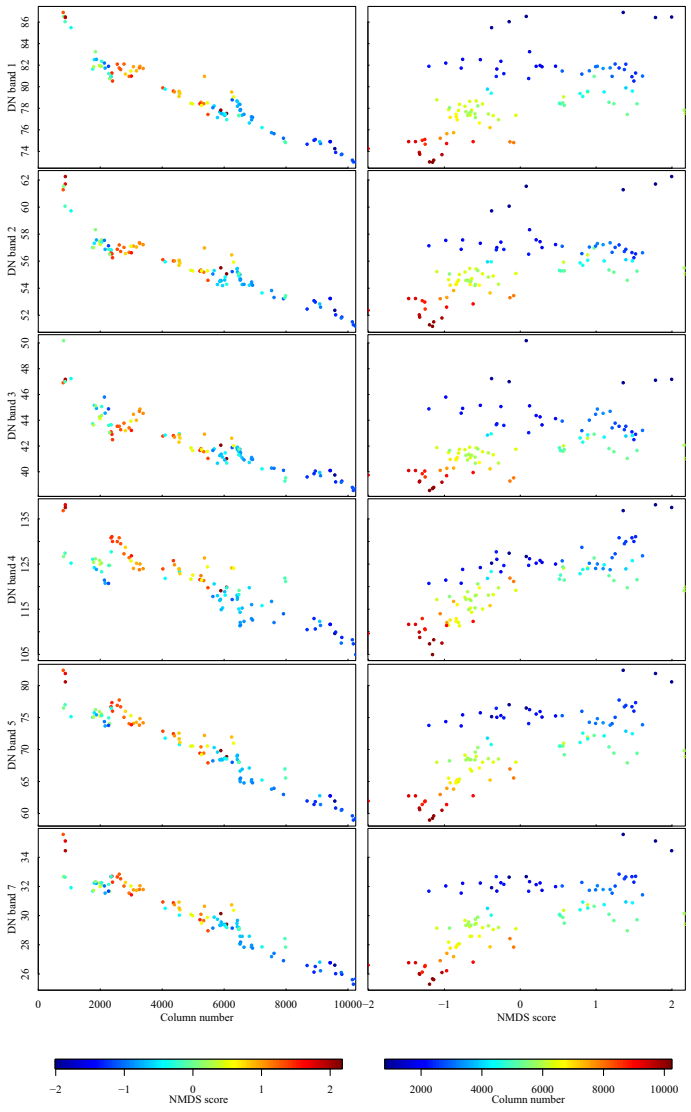

Figure 4: *DN* versus column number (left) and *DN* versus NMDS score (right) for 105 transects in Peru.

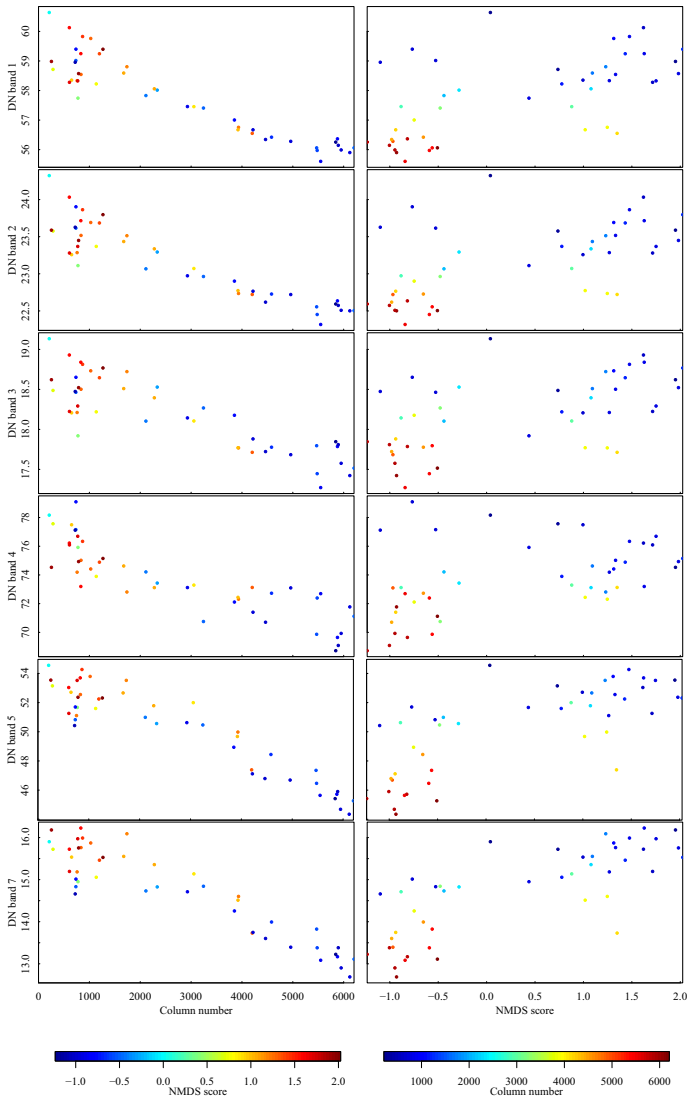

Figure 5: *DN* versus column number (left) and *DN* versus NMDS score (right) for 45 transects in Brazil.

Table 2: $\bar{R}^2$ values for the iterative regression analyses between digital number and column number (Γ) and NMDS score (ν), for the Peruvian and Brazilian Landsat TM/ETM+ mosaics.

	Peru		Brazil	
	Predictors	$\bar{R}^2$	Predictors	$\bar{R}^2$
Band 1	Γ	0.924	Γ	0.848
	$\Gamma + \nu$	0.923	$\Gamma + \nu$	0.845
	$\Gamma + \nu + \Gamma^2$	0.940		
	$\Gamma + \nu + \Gamma^2 + \Gamma^3$	0.972		
Band 2	Γ	0.857	Γ	0.823
	$\Gamma + \nu$	0.857	$\Gamma + \nu$	0.820
	$\Gamma + \nu + \Gamma^2$	0.875		
	$\Gamma + \nu + \Gamma^2 + \Gamma^3$	0.927		
	$\Gamma + \nu + \Gamma^2 + \Gamma^3 + \nu^3$	0.929		
Band 3	Γ	0.842	Γ	0.723
	$\Gamma + \nu$	0.848	$\Gamma + \nu$	0.717
	$\Gamma + \nu + \Gamma^2$	0.879		
	$\Gamma + \nu + \Gamma^2 + \Gamma^3$	0.905		
Band 4	Γ	0.779	Γ	0.729
	$\Gamma + \nu$	0.881	$\Gamma + \nu$	0.746
	$\Gamma + \nu + \Gamma\nu$	0.890	$\Gamma + \nu + \Gamma\nu$	0.885
	$\Gamma + \nu + \Gamma\nu + \Gamma\nu^2$	0.903	$\Gamma + \nu + \Gamma\nu + \Gamma\nu^2$	0.900
Band 5	Γ	0.936	Γ	0.869
	$\Gamma + \nu$	0.956	$\Gamma + \nu$	0.891
	$\Gamma + \nu + \Gamma\nu$	0.958	$\Gamma + \nu + \Gamma^2$	0.908
	$\Gamma + \nu + \Gamma\nu + \Gamma^2\nu$	0.961		
	$\Gamma + \nu + \Gamma\nu + \Gamma^2\nu + \nu^3$	0.963		
Band 7	Γ	0.933	Γ	0.848
	$\Gamma + \nu$	0.951	$\Gamma + \nu$	0.899
	$\Gamma + \nu + \Gamma\nu$	0.953	$\Gamma + \nu + \Gamma^2$	0.907
	$\Gamma + \nu + \Gamma\nu + \Gamma^2\nu$	0.958		

Table 3: Means and standard deviations of the slopes of linear regression of reflectance versus column number over Pebas and Nauta formation forests and the difference between their means (all unitless and all $\times 10^6$) for each Landsat TM/ETM+ band, per scene and for all scenes pooled. Asterisks indicate non-normal distribution according to Shapiro-Wilks test, superscripts ^{t} and ^{w} indicate significant differences for the t-test and Wilcoxon test, respectively (all at p=0.05).

		Pebas		Nauta		
Scene	Band	Mean	St. dev.	Mean	St. dev.	Difference
003/064	1	-0.9299	1.1181*	-0.5280	0.7060	-0.4020
	2	-0.7840	0.8208*	-0.4204	0.5329	-0.3636
	3	-0.2456	0.3708	-0.1574	0.4669	-0.0882
	4	-6.1178	2.5038	-3.3766	1.0596	$-2.7412^{t,w}$
	5	-4.0358	1.3477*	-2.9620	0.7976	$-1.0738^{t,w}$
	7	-1.8789	0.6976*	-1.2619	0.4385	$-0.6170^{t,w}$
007/062	1	-1.0706	0.5193	-0.8320	0.6532	-0.2386
	2	-0.9853	0.4352	-0.7761	0.5046	-0.2093
	3	-0.5138	0.4091	-0.2923	0.5388	-0.2215^{t}
	4	-6.9189	2.1561	-7.0125	1.4243	0.0936
	5	-4.2102	0.7976	-4.5051	0.7818	0.2949
	7	-1.8172	0.5909	-1.9635	0.6023	0.1463
008/062	1	-0.6820	1.5101*	-0.7758	0.8906*	0.0938
	2	-0.6617	1.1511*	-0.8976	0.6292*	0.2359
	3	-0.1517	0.9404*	-0.2192	0.5173*	0.0676
	4	-6.6532	1.7389	-8.9520	1.8638	$2.2987^{t,w}$
	5	-3.9889	0.9475	-4.5632	1.0401	$0.5742^{t,w}$
	7	-1.9235	0.5365	-2.0963	0.4808	$0.1728^{t,w}$
Combined	1	-0.8966	1.1103*	-0.6996	0.7486	-0.1969
	2	-0.8086	0.8423*	-0.6795	0.5827*	-0.1290
	3	-0.2998	0.6192*	-0.2186	0.4973	-0.0812
	4	-6.5336	2.1599	-6.2423	2.7806	-0.2913
	5	-4.0755	1.0594	-3.9402	1.1516	-0.1353
	7	-1.8736	0.6058	-1.7398	0.6245	-0.1338

Table 4: Means and standard deviations of the slopes of linear regression of reflectance versus column number, either stratified based on geological formation or unstratified, and the difference between their means (all unitless and all $\times 10^6$) for each Landsat TM/ETM+ band, per scene. Asterisks indicate non-normal distribution according to Shapiro-Wilks test, superscripts ^{t} and ^{w} indicate significant differences for the t-test and Wilcoxon test, respectively (all at p=0.05).

		Stratified		Unstratified		
Scene	Band	Mean	St. dev.	Mean	St. dev.	Difference
003/064	1	-0.7289	0.5625	-0.7373	0.4008	0.0084
	2	-0.6022	0.4596	-0.7815	0.3730	$0.1793^{t,w}$
	3	-0.2015	0.3091*	-0.2053	0.2714*	0.0038
	4	-4.7472	1.5206	-7.6987	1.7555	$2.9514^{t,w}$
	5	-3.4989	1.0215	-5.0451	1.1189	$1.5461^{t,w}$
	6	-1.5704	0.5403	-2.4040	0.6357	$0.8336^{t,w}$
007/062	1	-0.9513	0.5429	-0.9248	0.4326	-0.0264
	2	-0.8807	0.4151	-0.7552	0.3266	$-0.1256^{t,w}$
	3	-0.4031	0.4421	-0.3920	0.3441	-0.0110
	4	-6.9657	1.5497	-5.6448	1.4250	$-1.3209^{t,w}$
	5	-4.3576	0.6684	-3.7911	0.6432	$-0.5665^{t,w}$
	6	-1.8904	0.5392	-1.6300	0.5198	$-0.2604^{t,w}$
008/062	1	-0.7289	1.1833*	-0.7317	1.1708*	0.0028
	2	-0.7797	0.8746*	-0.7016	0.8481*	$-0.0780^{t,w}$
	3	-0.1855	0.7149*	-0.2206	0.7004*	0.0351
	4	-7.8026	1.5830	-6.4770	1.6745	$-1.3257^{t,w}$
	5	-4.2760	0.9587	-3.7931	1.0045	$-0.4829^{t,w}$
	6	-2.0099	0.4930	-1.7369	0.5099	$-0.2730^{t,w}$