Along-scan radiometric gradient in Landsat images: understanding the relation between floristic and remote sensing information in tropical forests

Javier Muro

UNIVERSITY OF TURKU
Faculty of Mathematics and Natural Sciences
Department of Biology

Muro, Javier:

Along-scan radiometric gradient in Landsat images: understanding the relation between floristic and remote sensing information in tropical forests

Master´s thesis,
48 pages, 25 figures,
7 tables, 3 Annexes
April 2014

The along-scan radiometric gradient causes severe interpretation problems in Landsat images of tropical forests. It creates a decreasing trend in pixel values with the column number of the image. In practical applications it has been corrected assuming the trend to be linear within structurally similar forests. This has improved the relation between floristic and remote sensing information, but just in some cases. I use 3 Landsat images and 105 floristic inventories to test the assumption of linearity, and to examine how the gradient and linear corrections affect the relation between floristic and Landsat data. Results suggest the gradient to be linear in infrared bands. Also, the relation between floristic and Landsat data could be conditioned by the distribution of the sampling sites and the direction in which images are mosaicked. Additionally, there seems to be a conjunction between the radiometric gradient and a natural east-west vegetation gradient common in Western Amazonia. This conjunction might have enhanced artificially correlations between field and remotely-sensed information in previous studies. Linear corrections may remove such artificial enhancement, but along with true and relevant spectral information about floristic patterns, because they can´t separate the radiometric gradient from a natural one.

Contents

1. Introduction

Earth observations systems (EOS) have been used during the last decades to provide us with information of our planet at scales too coarse to handle from an on-land perspective. They have allowed us to achieve an improved understanding of Earth as an integrated system. Their products have been used frequently in natural resources management. One of the most popular and widely used multispectral sensors are the ones on board the satellites Landsat 1-5, 7, and now Landsat 8. The Landsat project has been operating since 1972, creating the longest continuous global record of the Earth's surface. Among other topics, it has proven useful in local, regional, and continental scale ecological studies (Chambers et al. 2007), especially in areas of difficult accessibility such as the Amazon forest.

Amazonian rain forests and especially non inundated forests were traditionally considered rather homogeneous (Encarnación 1985; Pires & Prance 1985; Salo et al. 1986). Later on, several authors rejected that idea, proving that plant species composition and distributions were significantly related to physical and chemical characteristics of soils at both regional and landscape scales (Tuomisto et al. 1995; Ruokolainen et al. 1997; Tuomisto et al., 2002, 2003 a,b,c; Phillips et al. 2003). Since then, the necessity to truly understand the relation between plant species composition and environmental gradients has become an urgent issue regarding to biodiversity and land management of tropical forests. In other words, we want to predict how the floristic composition changes through space and what are the drivers of these changes. Such changes might be gradual (e.g. variation according to gradients of temperature or precipitation) or sudden (e.g. changes in species composition with changes in geological formations).

Variation observed in remotely sensed images has been correlated with variation of plant species compositions and environmental factors (Tuomisto et al. 1995; Ruokolainen & Tuomisto 1998; Tuomisto et al. 2003a, b; Rajaniemi et al. 2005). The use of now widely available remote sensing data of moderate spatial and spectral resolution in combination with relatively fast and economic sampling campaigns, has generated a better understanding of environmental gradients and how species are distributed through space (Salovaara et al. 2005, Thessler et al. 2005, Higgins et al. 2012, among others). However, it is essential to understand the functioning of remote sensing systems, their limitations, and the wide array of preprocessing techniques available in order to avoid drawing erroneous conclusions. This will allow to generate more truthful maps of ecological features or processes that can be used in conservation planning and land management.

3

The radiation registered by the sensors is dependent on the reflectance properties of the surface, but also on other factors such as the geometry formed by the sun, the surface and the sensor, atmospheric conditions, topography of the terrain, and the reflectance properties of neighboring surfaces. For a correct interpretation of the satellite information it is necessary to consider all these variables, especially when comparing images from different sensors or dates. In the nearly homogeneous surfaces of tropical lowland rain forests it is necessary to use the radiometric, spatial, and spectral resolution of the sensors to its limits in order to distinguish spatial patterns, tendencies or vegetation types (Toivonen el at. 2006). For these reasons, removing any extraneous and systematic pattern such as banding or memory effect is crucial for a right interpretation (Toivonen el at. 2006). One additional example of these systematic patterns is the so called "across-path gradient", or "east-west radiometric gradient". This phenomenon was noted first by Collet et al. (1998) and Danaher (2002), and Toivonen et al. (2006) provided an exhaustive description of it. I explain it in the following lines and more thoroughly in the section "Theoretical background".

The phenomenon causes an increase of brightness on the west side of the image, and a gradual darkening towards the east side. In Digital Numbers (DNs or pixel values) it means higher values at the west than at the east for the same type of surface (see Annex 1a as an example). Although very small, this tendency in the DNs becomes problematic when interpreting environmental and biological patterns in spectrally homogeneous regions such as Amazonia, leading in some cases to erroneous interpretations of spatial patterns. For example, after visual interpretation of a Landsat image of the Amazon River, Tuomisto et al. (1994) and Kalliola et al. (1997) assumed habitat difference between western and eastern sides. The field information collected later on (Ruokolainen & Tuomisto 1998) rejected that assumption.

The gradient is likely caused by differences in illumination conditions, the anisotropic reflectance of the surface, and atmospheric scattering (Toivonen et al. 2006). Radiance coming from illuminated areas is registered in the west part of the image, and from less illuminated areas in the east part (Figure 1). This difference in illumination conditions is produced in tropical forests by the cloud-shaped and emergent tree tops that cast shadows of their own. The gradient appears to match the scanning direction in the majority of the cases (Toivonen et al. 2006) and thus one can assume that it is function of the column number of the image.
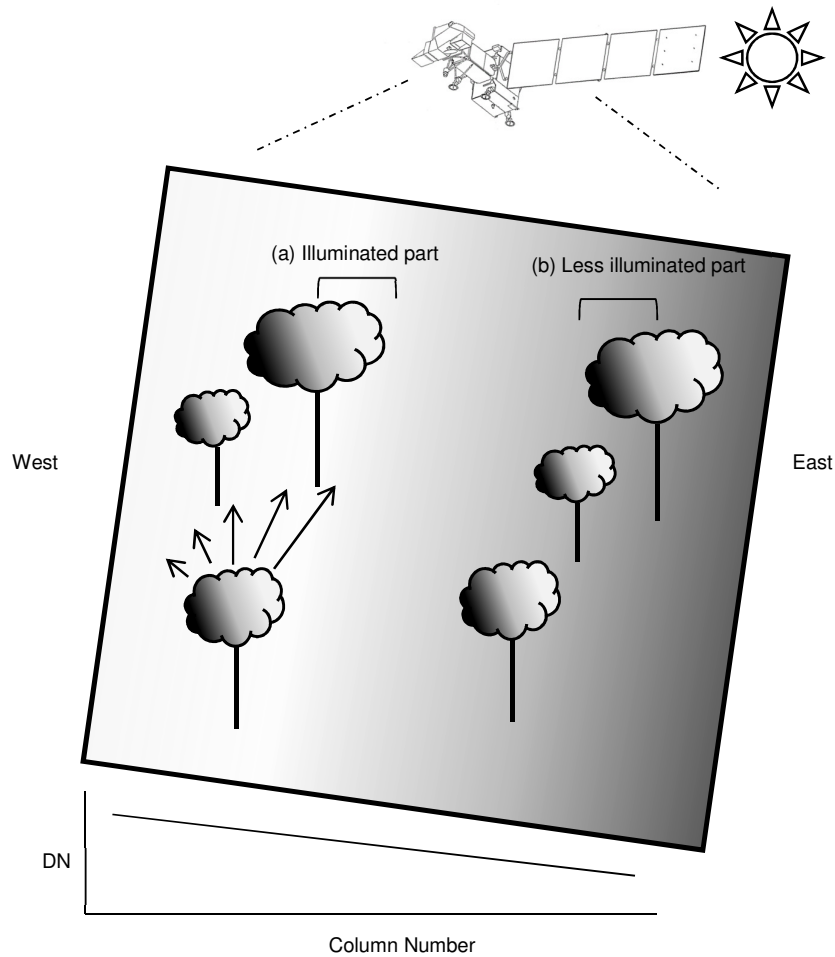
Figure 1: Conceptual representation of the effect of the gradient. The sensor registers the reflectivity coming from illuminated parts in the left side of the image (a), and the reflectivity coming from the less illuminated parts (b) in the right side. This translates into a decrease in DN values with the column number, i.e. a darkening towards the east. The arrows of different sizes over the lower-right tree simulate the different intensities with which light is reflected, depending on the angle of view.

Hansen et al. (2008) acknowledged the challenge that the gradient imposes and used bulk empirical linear adjustments to improve radiometric response and cover characterizations. Higgins et al. (2011, 2012) developed a correction called image detrending which models a regional trend in pixel values using surface interpolation and removes it from the image. The slope of the gradient has been found to be dependent on the land cover (e.g. steeper in forests than in open areas, Toivonen et al. 2006). Thus, any linear correction must be land cover specific. Image detrending is applied to the whole image and thus it over- or under corrects the gradient on other land covers. It assumes the gradient to behave linearly through the land cover studied; this is, the gradient has the same slope across the whole land cover studied within an image. In my research I test whether this assumption is correct or not for the case of *tierra firme* forests. They are closed canopy forest of the Amazonia, non-inundated, uniform-looking,

and very rich in species. They are relatively easy to identify and delineate in a medium spatial resolution satellite image, and occupy a large part of Amazonia (80% of the Peruvian Amazonia, Salo et al. 1986). My study area includes two geological formations, Pebas and Nauta, which determine the species composition of the forests (Higgins et al. 2011). Despite this variation in species composition, their structure remains similar and thus the radiometric gradient has been assumed to be linear.

Higgins et al. (2012) studied the relation between vegetation and remote sensing information in three study areas of *tierra firme* forests, before and after the detrending. The correction improved the correlation between the floristic data and the satellite image in two of the three sites surveyed. They also achieved different results depending on the band of the image. This brings up doubts about the linearity of the gradient and about the usefulness of image detrending as a gradient correction technique. If the gradient is indeed non-linear, the previous attempts of correction would be flawed and more complex corrections would be necessary.

2. Theoretical background

2.1. Landsat as a tool to study local and regional biological patterns

2.1.1. Landsat 4-5 and 7 technical characteristics

Landsat satellites move in a north-south orbit over the sunlit side of the Earth; i.e. they have a sun-synchronous and circular orbit. They orbit the Earth at 705 kilometers altitude and make a complete orbit every 99 minutes, make about 14 full orbits each day, and cross every point on Earth once every 16 days. The illumination conditions under which imagery is obtained vary due primarily to the north-south seasonal position of the sun relative to the Earth.

The Thematic Mapper (TM) sensors onboard Landsat 4 and 5 were designed with the following characteristics: three bands in the visible part of the spectrum, and four bands in the infrared part; spatial resolution of 30 meters for the visible, Infrared (IR), Near Infrared (NIR), and Short Wave Infrared (SWIR) bands; and the addition of a 120-meter thermal-IR band. Landsat 7 carries the Enhanced Thematic Mapper Plus (ETM+) that has all TM bands, plus a 15-meter panchromatic band (Table 1).

Table 1: Technical characteristics of each of the Landsat bands. Panchromatic band is only present in Landsat 7 (source http://landsat.usgs.gov)

Spectral bands	Wavelength (micrometers)	Resol. (m)	Use
1- Blue	0.45–0.52	30	Bathymetric mapping; distinguishes soil from vegetation; deciduous from coniferous vegetation.
2- Green	0.52–0.61	30	Emphasizes peak vegetation, which is useful for assessing plant vigor.
3- Red	0.63–0.69	30	Emphasizes vegetation slopes
4- IR	0.76–0.90	30	Emphasizes biomass content and shorelines.
5- NIR	1.55–1.75	30	Discriminates moisture content of soil and vegetation; penetrates thin clouds.
6- Thermal	10.40–12.50	120	Useful for thermal mapping and estimated soil moisture.
7- SWIR	2.08–2.35	30	Useful for mapping hydrothermally altered rocks associated with mineral deposits
8- Panchromatic	0.52–0.90	15	Useful in 'sharpening' multispectral images

2.1.2. Spectral response of vegetation

The reflectance properties of vegetated land covers are controlled by three parameters (Asner 1998):

1) Leaf biochemical properties (e.g., water, photosynthetic pigments, structural carbohydrates), which create wavelength-specific absorption features,

2) Leaf morphology (e.g., cell-wall thickness, air spaces, cuticle wax), which affects photon scattering

3) The three-dimensional architectural arrangement of foliage and non-photosynthetic components, that determines the amount of photon volumetric-scattering and attenuation within the crown

Different types of vegetation have different reflectance magnitudes for the different regions of the optical spectrum. Figure 2 shows examples of 4 land cover classes along 0.8 micrometers of the spectrum.
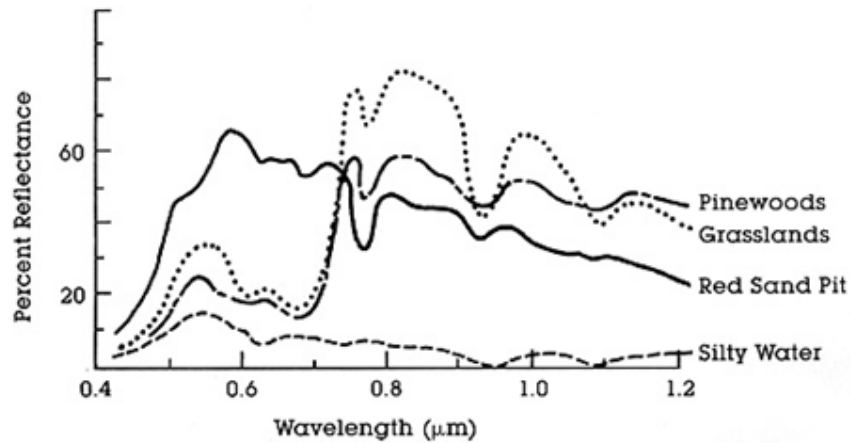
Figure 2: Book example of the spectral response curves for different targets. Extracted from Landsat User´s Handbook (Landsat Project Office 2002)

The Amazon forest has more homogeneous spectral responses than the ones shown in Figure 2. Figure 3 shows an example of spectral response of different high canopy tropical species. To distinguish meaningful spectral patterns within tropical forests it is necessary to make the most of the radiometric capabilities of the sensor performing certain enhancement techniques. It is after these enhancements when the effect of the gradient becomes clearly distinguishable.
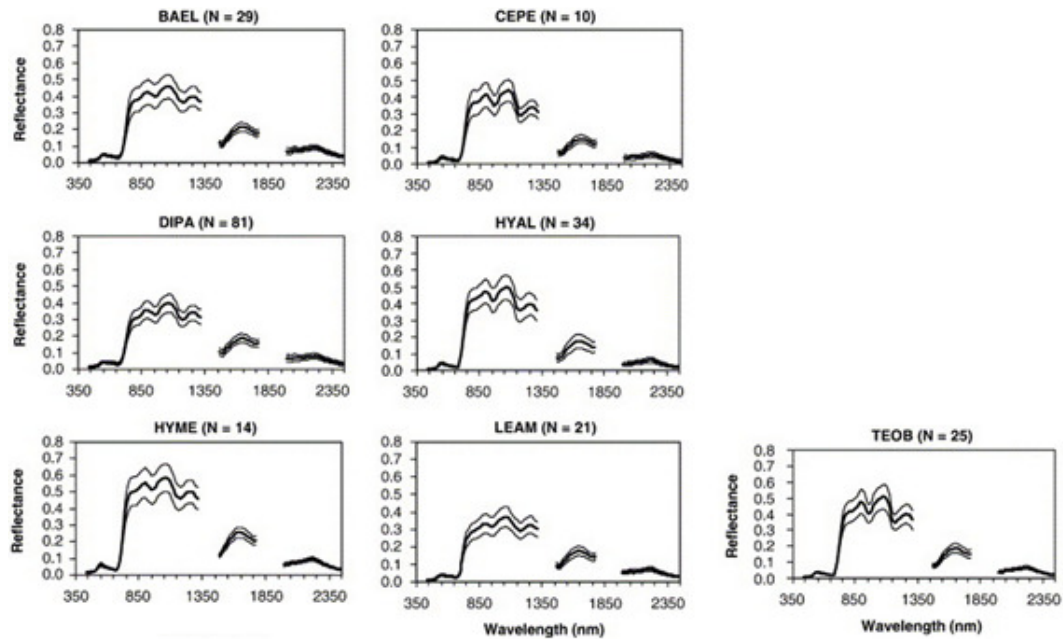
Figure 3: Spectral response of 7 different high canopy tropical tree species and their standard deviation. Extracted from Clark et al. (2005). BAEL: *Balizia elegans.* CEPE: *Ceiba pentandra.* DIPA: *Dipteryx panamensis.* HYAL: *Hyeronima alchorneoides.* HYME: *Hymenolobium mesoamericanum.* LEAM: *Lecythis ampla.* TEOB: *Terminalia oblonga.*

2.1.3. Radiometric enhancements

Often, raw images do not highlight the particular features or processes that one wishes to understand. Image enhancements are a variety of techniques that reveal such features, making images more interpretable for a given application. Subtle differences in vegetation types for instance, may be detected by the human eye only by increasing the contrast of the image; this is, by assigning a wider array of colors to our area of interest. One of these techniques and the one used to highlight the visual effect of the east-west gradient is histogram equalization.

An image histogram is an "x-y" graph where "x" represents all the possible pixel values and "y" their frequency. It contains only radiometric and not spatial information. Usually the most common pixel values are restricted to a small range and thus, contrast between different predominant features is low making them indistinguishable for the human eye. Histogram equalization modifies the pixel values of the image in such a way that the histogram presenting their frequencies becomes as even as possible. This allows the image to use a wider range of brightness values to represent the most frequent DNs, enhancing the visual contrast between the features of interest. Precisely because this enhancement increases the contrast between the most common pixel values, the effect of the gradient is more visible here. Figure 4 shows the study area using the default standard deviation technique for visualization (a), and histogram equalization (b).
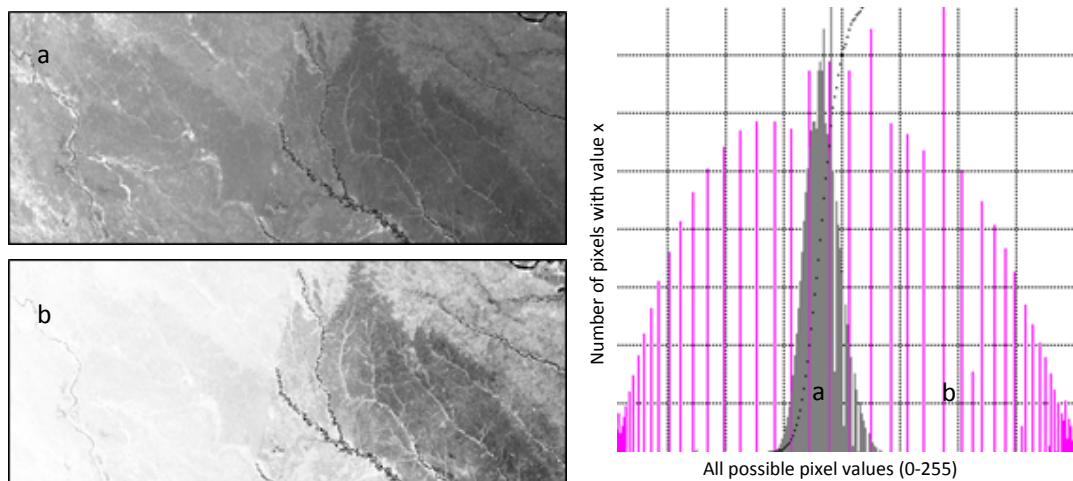

Figure 4: Study area with standard deviation technique (a) and histogram equalization (b). The histogram represents the original pixel values in grey, and the histogram equalized ones in stripes. Band 4 was used for the representation. It is visible how histogram equalization enhances dramatically the effect of the radiometric gradient.

2.1.4. Radiometric manipulation of the images: BRDF

The subtle (but enhanced in the histogram equalization) brightness artifact is caused by the non-Lambertian behavior of the Earth surface. This means that the radiation absorbed is later emitted with different intensities and directions according to the bidirectional reflectance distribution function (BRDF). This function describes how reflectance varies in each cover considering the angles of incidence and observation. Figure 5 shows an example of 4 different surfaces with different reflectance properties.

Mirror BRDF:
specular reflectance

Rough water surface BRDF:
sunglint reflectance

Volume scattering BRDF:
leaf/vegetation reflectance

Gap-driven BRDF (Forest):
shadow-driven reflectance
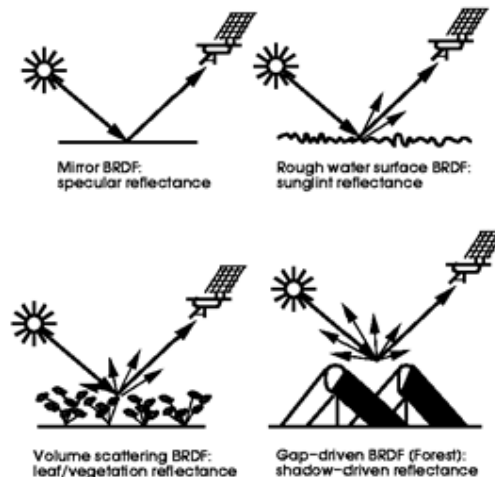
Figure 5: Bidirectional reflectance Distribution Functions: Causes. Extracted from Lucht (1997).

Modelling the BRDF is rather complex. Thus, in practical applications it is more convenient to assume a Lambertian behavior of the surface (Riaño et al. 2003).

Toivonen et al. (2006) provided a description of the strength of the gradient depending on the cover type, solar elevation and azimuth, and scattering angle. They normalized the sun elevation angle using the gain and bias given by Chander and Markham (2003) and atmospheric conditions using the COST method proposed by Chavez (1996) between 2 scenes. They also passed a BRDF correction proposed by Pellikka (1998) and found that the correction parameters for the radiometric gradient weren´t transferable between two scenes despite all the normalizations.

Before BRDF parameters can be fitted, atmospheric correction must be performed. One of the most used methods is the Dark Object Subtraction. It assumes that in every image there will always be features with reflectance values close to 0, such as deep waters. Therefore, if these objects have DNs higher than 0 it is due to the influence of atmospheric noise and that deviation from 0 is subtracted from all the pixels of the image. However, it implies that atmospheric conditions are the same throughout the image, and that´s usually not the case (Sheppard & Dymond 2000). I estimated that the corrections

applied in Toivonen et al. (2006) weren´t necessary in my case because images were acquired from the same season and hour (similar solar elevation and azimuth (Table 2), and the presence of clouds and haze is minimal. The image detrending applied in later stages performs as a BRDF correction.

Since I am using several images to conduct the quantitative and visual analyses it is still necessary to normalize them; this is, to set the DNs of the three images to a common scale. When using pairs of image of the same location but different date, histogram match is the most common procedure used. It manipulates the histogram of one image to match the other one so that the apparent distribution of brightness values in the two images is as close as possible. This procedure requires that the shapes of both histograms are similar. Because my images are adjacent, their histograms are too different and cannot be matched. The mosaic was then compiled matching the histograms of the overlapping area, and performing a relative standardization of the DNs of the rest of the image.

2.2. Pebas and Nauta geological formations and their spatial patterns

The humid Amazonian forests have been traditionally divided according to their drainage conditions: seasonally inundated forests, *tierra firme* forests, and permanently waterlogged swamp forests. Within *tierra firme* forests there have been several divisions mainly according to topographic criteria (Salovaara et al. 2005) and recent studies have demonstrated the classificatory power of digital elevation data in rain forests (Higgins et al. 2011). The latter study used also geological formations and their edaphic properties as division criteria to uncovered two distinct forest ecosystems: the ones growing on Pebas and on Nauta geological formations.

Pebas formation is a poorly weathered and cation-rich clay sediment layer deposited under the low energy semi-marine or lacustrine conditions of the Pebas Embayment (Räsänen et al. 1995). On the other hand, Nauta formation is a weathered sandy and cation-poor layer of fluvial sediments that overlies Pebas. Pebas soils have around 15 times higher cation content than Nauta (sum of Ca, Na, Mg and K; Higgins et al. 2011). This has consequences in the structure and composition of species. For example, on poorer soils plants invest in roots to make the most of the scarce nutrients and in leaf defenses in order to hold on to them. On the other hand, on richer soils plants invest more in growth and produce greater leaf mass (Grime 2001). This translates into a differentiation of 80-90 % in their plant species and in a higher reflectance in infrared

wavelengths from forests on richer soils (Higgins et al. 2011, 2012). Both forest types are clearly distinguishable in Landsat images as it can be observed in Figure 6.
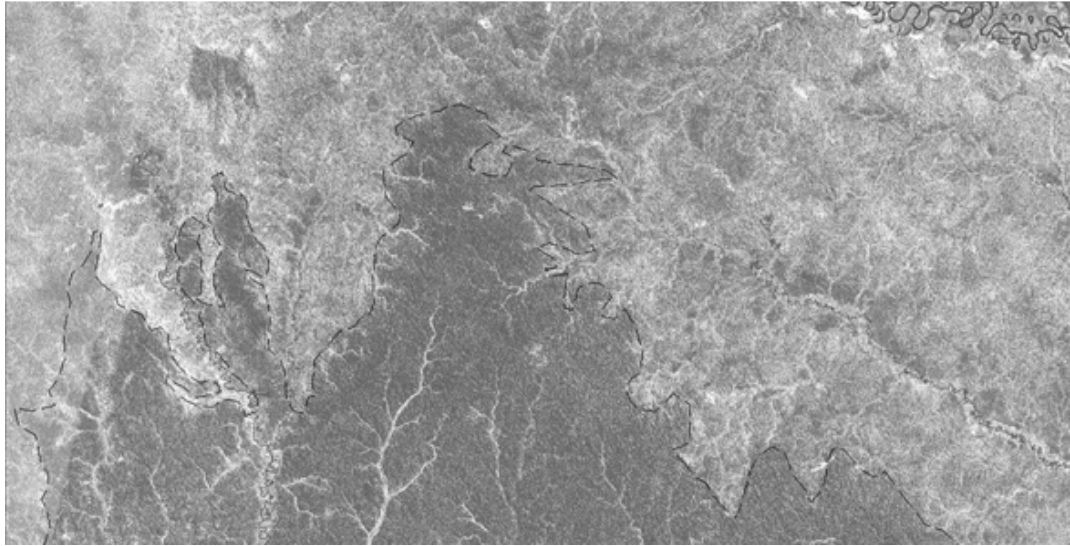
Figure 6: Detrended Band 4 image depicting the limit between Pebas (north) and Nauta (south), manually delineated out of Higgins et al. (2011).

In a broader scale, floristic patterns in western Amazonia are partly conditioned by the gradual change from recently deposited and nutrient-rich sediments of Andean origin in the west, to older, weathered and nutrient-poor soils in the east (Terborgh & Andresen, 1998; ter Steege et al., 2006; Pitman et al., 2008). Additionally, Higgins et al. (2011) proposed that trends in cation richness are also explained by the wide-spread removal of the cation-poor Nauta formation in the west, and the consequent exposure of the underlying cation-rich Pebas formation. Both research lines suggest the existence of a broad-scale trend of forests growing on richer soils in the west, and forests growing on poorer soils to the east.

2.3. Pteridophytes, melastomes, and remote sensing data

Indicator species have been traditionally used in temperate regions to classify vegetation types or site types according to edaphic or drainage conditions (Cajander 1926 cited by Salovaara et al. 2004). *Tierra firme* forests are difficult to subdivide using physiognomic criteria because they appear relatively uniform, and using indicators might prove problematic because of the high species richness. However, during the last years certain groups of plants have been successfully used to elaborate preliminary classifications such as Melastomataceae and Pteridophyta (Ruokolainen et al. 1997; Ruokolainen and

12

Tuomisto 1998; Tuomisto et al. 1995, 2002, 2003a, b; Salovaara et al. 2004; Higgins et al. 2011, 2012).

Variations in environmental drivers (e.g. soil chemistry and drainage conditions) are reflected in the distribution patterns of several plant species, resulting in a congruence of floristic patterns in taxonomically unrelated groups (Salovaara et al. 2004). An adequate indicator species must be a good representative of these environmental variations, and melastomes and pteridophytes along with palms have been proved to have strong soil-related distributions (Ruokolainen et al. 1997; Ruokolainen and Tuomisto 1998; Vormisto et al. 2000; Phillips et al. 2003; and Clark et al. 2005, Kristiansen et al. 2011). Consequently, these plant groups may work as adequate indicators of soil conditions, and by inference also of floristic patterns in other plant groups (Ruokolainen et al. 1997). In other words, if we know that pteridophyte and melastome species composition differ between sites, we can make rather certain inferences also about disparities in tree species compositions between sites. Therefore, although the spatial patterns we observe in understory plant species are not visible in Landsat data, the corresponding patterns in canopy species are visible. Such canopy patterns are then translated into digital patterns in the image by variations in leaf chemistry and morphology, and canopy structure, (Asner 1998, Higgins et al. 2012).
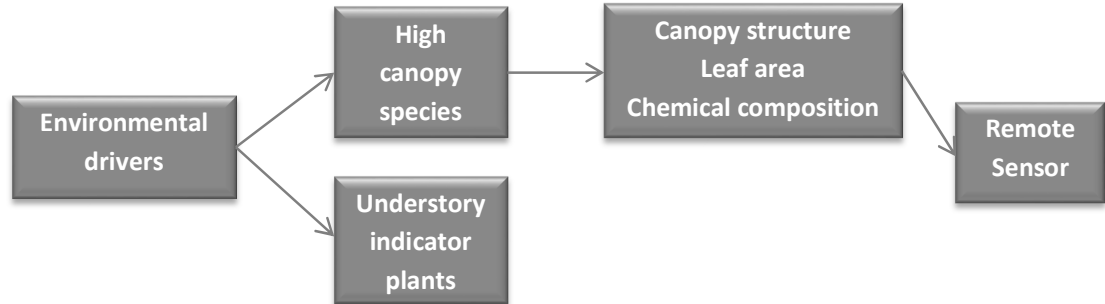
Figure 7: Link between data collected (understory plants) and remote sensing information. The concordant response between understory and canopy plants to environmental drivers allows to link understory species composition (easier to measure) with remote sensing information.

The linkage between soils, terrestrial pteridophytes and melastomes, and canopy plants (Figure 7) allows the use of remotely sensed data to separate rain forest types that have been recognized based on their indicator species composition with reasonable accuracies (Salovaara et al. 2005). However, due to the radiometric gradient these and even other spectrally more distinct forest types (swamp, flooded and unflooded forests, Toivonen et al. 2006) may be mixed if classified solely on the basis of their DN values.

Linear corrections for the gradient remove that artificial trend in pixel values. However, if floristic composition varies gradually with the column number, it would be difficult to

disentangle this natural variation in the image from the radiometric artifact. Image detrending could remove both gradients and relevant spectral information would be lost.

2.4. Research questions

Pixel values are function of the radiometric gradient and of the vegetated cover properties (assuming other factors to remain constant e.g. atmospheric conditions). The gradient is a function of the column number, and the vegetation cover is given in this case by the floristic composition. Linear corrections methods are designed to remove the influence of the column number in the pixel values. However, if there is a significant correlation between floristic composition and column number, the correction would eliminate such correlation. Consequently relevant spectral information about species composition could be lost. The effects of the radiometric gradient and linear corrections will be better understood when answering these questions:

> 1.- Can the radiometric gradient be considered linear across *tierra firme* forests of different floristic compositions? And is this floristic variation acting as a confounding factor when modelling the gradient?

> 2.- Why has image detrending improved the relationship between floristic data and DNs only in some cases?

To answer the first point I use Landsat and indicator species data to determine how much better the gradient can be explained by a curved model than by a linear one. Similar explanatory powers of both models would suggest a linear behaviour of the gradient. I also calculate and compare the slope of the gradient in different parts of the image using a much larger set of points but without floristic data. If the gradient is linear, the slopes should be similar regardless of what area of *tierra firme* is used to calculate it. To answer if floristic variation acts as a confounding factor I examine if a correlation between floristic data and column number exists, and how it affects the modelling.

To answer the second point, I detrend the images using several approaches and compare my results with those of Higgins et al. (2012). I also offer a possible explanation for them in the light of the results obtained.

3. Materials and methods

3.1. Study area

The study area occupies around 250 km × 350 km in northern Peru (Figure 8). It has a tropical, humid, and almost aseasonal climate, and the mean monthly temperature in the city of Iquitos (250 km away) is 25–27 °C throughout the year. Annual precipitation is about 3100 mm (Marengo, 1998). Elevation in the study area ranges from 98 to 433 m above sea level, and the landscape ranges from flat to precipitously hilly (Higgins et al. 2011).

Two important geological formations converge in the study area: the Pebas formation (equivalent to the Brazilian Solimões formation) in the north, and the overlain Nauta formation in the south (Higgins et al. 2011). Erosion processes have removed the Nauta formation in large areas across western and central Amazonia and in the north of our study area, exposing the Pebas formation. This has modified the landscape and the topography in certain aspects. For instance, patches of Nauta formation remain among large extensions of Pebas across western Amazonia. There is also a prominent topographic discontinuity of 20 m between north (lower elevation, Pebas) and south (higher elevation, Nauta) in my study area. Overlaid upon this local change in elevation there is a broad tilt from higher elevations in the north-west to lower elevations in the south-east (Higgins et al. 2011). The study area also contains portions of the Pastaza Fan in the west. Several floods and volcanic eruptions up the river through the history have changed the course of the river, and filled it with volcanoclastic sediments (Räsänen et al. 1990, 1992). Field sampling was carried out mostly on Nauta and Pebas formations, and only three samples were taken in Pastaza Fan.

In the study area the majority of the land is *tierra firme*. There the vegetation consists of closed-canopy evergreen rain forests typical of western and central Amazonia (Higgins et al. 2011). Some sporadically inundated zones are present along rivers, and the southernmost portion of the Pastaza Fan is inundated seasonally or permanently (Higgins et al. 2011).
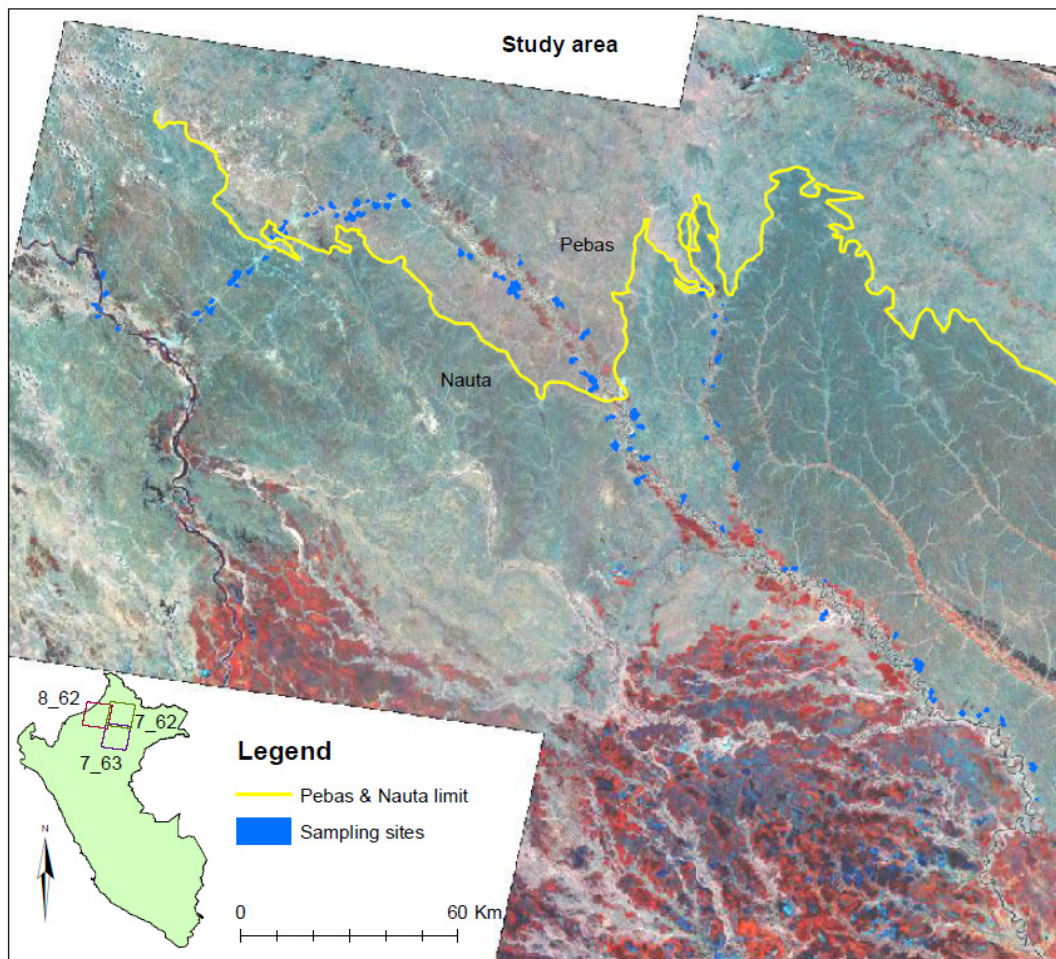
Figure 8: Mosaic of the study area formed by the three Landsat images used, assigning bands 4, 5 and 7 to red green and blue respectively. *Tierra firme* forests appear in several tones of green. Inundated areas, swamps, and bogs have red and bluish tones and are located to the south of the mosaic and along the main rivers. Blue segments correspond to the sampling sites where the floristic information was extracted. Divisor line between Pebas and Nauta was delineated manually out of the results of the SRTM analysis of Higgins et al. (2011).

3.2. Datasets

I counted on two sets of data: the first one consisted of floristic information about presence and absence of Pteridophyta (ferns and lycophytes) and Melastomataceae (a family of shrubs and small trees) along with its geographical location. The second one was a set of Landsat images of the study area.

3.2.1. Floristic data

The field data was collected between 2005 and 2006 by Mark Higgins, Hanna Tuomisto, and Glenda Cárdenas (pteridophytes), and Kalle Ruokolainen and Nelly Llerena

16

(melastomes), and used in Higgins et al. (2011, 2012). They used both field observations and SRTM elevation data to ensure that the sampling was limited to undisturbed *tierra firme* forests of the geological formations of interest explained above, excluding sites with alluvial soils in inundated areas.

The sampling strategy was designed to be sufficiently extensive to represent the width of large geological formations (300 km x 170 km), but also sufficiently dense to detect abrupt changes in floristic composition within the non-inundated forests. Data collection followed the standard procedure described in Tuomisto et al. (2003a). This means that plants were collected along a set of 139 linear transects of 5 m x 500 m each. Each transect followed a predefined compass bearing and the coordinates were noted using a GPS 2, 3, or 4 times along each transect, whenever possible. Later on, I used these coordinates and the compass bearings to interpolate the transect lines and locate them on the Landsat images. The software used for that was ArcMap 10 (ESRI). Presence-absence data was collected for each species, and for pteridophytes only individuals with at least one leaf longer than 10 cm were considered. Epiphytic and climbing individuals were included only if they had green leaves at a height ≤ 2 m above ground. All individuals of Melastomataceae excluding Olisbeoideae (syn. Memecylaceae) taller than 5 cm were recorded. Soil samples were also taken during the sampling campaign, but since they won´t be used in this analysis I won´t describe them here.

The result of the data collection was a matrix where rows represent transects and columns species, with 0 indicating absence and 1 presence.

3.2.2. Remote sensing data

I analyzed a set of three Landsat 5 and Landsat 7 scenes (Table 2) downloaded from the United States Geological Survey. An exhaustive search was carried out to find images free from clouds and haze, not too temporarily distant, and from the same season in order to minimize the influence of object-sun-sensor geometry (Landsat satellite covers that area every 16 days at around 15.00 h).

Table 2: Relevant remote sensing metadata

Path	Row	Date	Place name	Projection	Sun Azimuth	Hour	Satellite
8	62	8.9.06	Río Pastaza	UTM Zone 18 WGS 84	73.3150337	15:08	Landsat 5 TM
7	62	21.8.99	Ríos Chambira and Pintoyacu	UTM Zone 18 WGS 84	62.6001968	15:01	Landsat 7 ETM
7	63	21.8.99	Marañón with Tigre	UTM Zone 18 WGS 84	60.8536224	15:01	Landsat 7

Landsat images are systematically georectified using data generated by onboard computers during imaging events. This level of correction is called Level 1G and its geometric accuracy is estimated to be between 30-50 m (Landsat User´s Handbook 2002). The images I used incorporate in addition ground control points and a Digital Elevation Model (DEM) for topographic accuracy. This level of correction is called Level 1T Standard Terrain Correction and its geodetic accuracy depends on the accuracy of the ground control points and the resolution of the DEM. Therefore, the geometric accuracy of the images I used should be better than 30-50 m.

Images 7_62 and 7_63 were taken the same day and they can be considered as one single image. Image 8_62 corresponded to a different date and thus a radiometric normalization was required between 8_62 and the pair 7_62 and 7_63. This was needed to reduce differences arising from changing illumination, atmospheric conditions and sensors. The relative normalization adjusts the spectral values of all images to the values of one reference image using a simple linear relationship, so that both images are at the same reference level (Râsi et al. 2011). In this case, 8_62 was normalized to 7_62 and 7_63 merging them into a mosaic using histogram match for the overlapping areas. The software used was ERDAS IMAGINE (ESRI). All the bands were used to build the mosaic, but for visual interpretation and maps I used mostly bands 4, 5 and 7. These bands have been found to represent better floristic patterns (Tuomisto et al. 2003a).

3.3. Generation of the variables

The basic unit of observation for the analyses is the sampling sites where the floristic information was collected. Each sampling site has three variables:

- Column number of the image (CN)
- Average pixel value for each sampling site (DN)
- Position of each sampling site along a floristic gradient (NMDS score)

CN and NMDS score are the explanatory variables, and DN the response variable.

3.3.1. Column Number

Because column numbers are lost when the image is georectified I calculated the shortest distance between the center of each sampling site to the west margin of the scene (Figure 9). Although the gradient is called across-path, it is actually along-scan (Toivonen et al. 2006). Dividing this distance by the resolution (30 m) I produced a parameter equivalent to the CN.
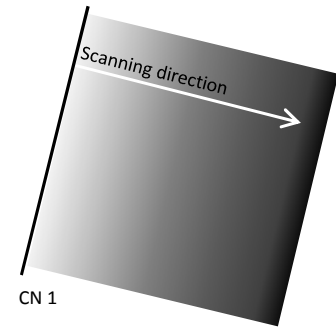
Figure 9: Scanning direction and starting point of the variable Column Number in a single georectified image.

3.3.2. Digital Numbers

DNs were extracted from the areas surrounding the transects where the floristic information was collected. To delimit these areas an object-based approach known as multispectral segmentation was considered advantageous in comparison to a pixel-based classification. In this approach, a segmentation algorithm delimits the boundaries of the objects according to the thresholds set, and pixels that are spatially connected and have similar values are grouped in a single segment. This depicts areas of similar color and textural characteristics as representative of homogeneous vegetation stands, in the same way as a photo interpreter would do. These vegetation stands are defined as areas in which vegetation is more homogeneous within the boundaries than across them (Woodcock 1994). Compared to traditional photointerpretation that uses only three bands at a time, one per cannon of color, I was able to use the six non-thermal bands to segment the mosaic. I run then several segmentations using different combinations of parameters to get segments of manageable sizes. The process was executed in eCognition v 8.8, using the following parameters: Scale 20, Shape 0.4, and Compactness 0.4.

From the total amount of segments generated, I extracted the ones that overlaid a transect, and calculated the average DN of the pixels falling inside each one. When two segments were crossed by one transect, the average DN of each one was weighted with the fraction of the transect that crosses it.

3.3.3. NMDS score

All the raw floristic data needed to be compressed into one dimension that represented as faithfully as possible sample and species relationship. This was done by ordering the sampling sites according to their similarity in species composition. Ordination techniques are used for such purpose and the one chosen here is the Non-metric Multi-Dimensional Scaling (NMDS). The method first creates a similarity matrix out of the floristic information matrix. This new matrix is square, symmetric, and has zeros in the diagonal. It is generated by comparing the similitude in species composition of each pair of samples. The aim of the NMDS is to arrange samples in a low dimensional ordination space (one dimension in this case) so that the between-point distances in the ordination have the same rank order as do the distances in the original similarity matrix (Whittaker 1987). A scatterplot of distances in the NMDS ordination against original dissimilarities is called a Shepard´s diagram (Figure 10). One of the advantages of the NMDS is that it replaces the assumption of linearity present in many ordination techniques with the weaker and less problematic assumption of monotonicity or rank order correlation (Whittaker 1987). The degree to which ordination space distances agree in rank-order with original dissimilarities can be determined by fitting a monotone regression of the ordination distances onto the dissimilarities. Distances between pairs increase monotonically as pairs are perceived to be more dissimilar.
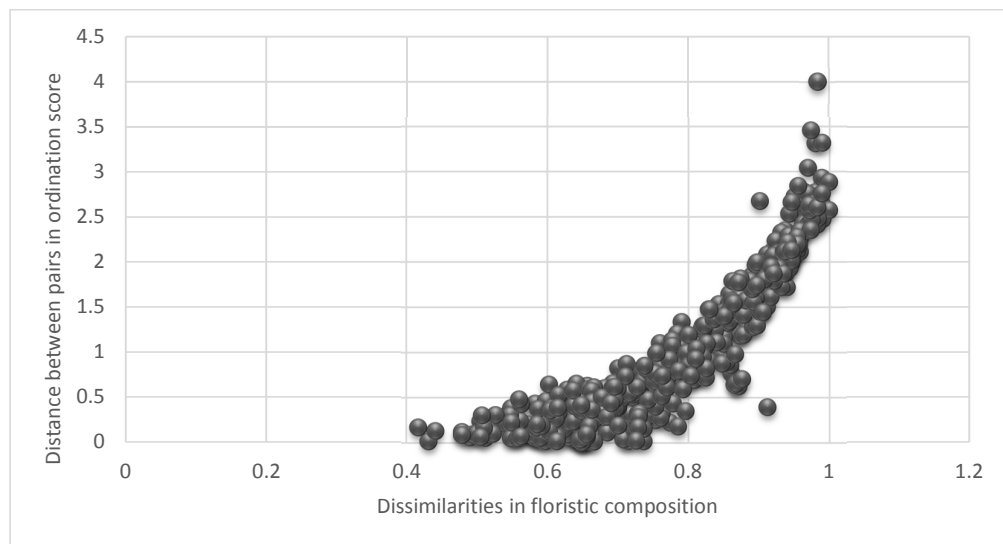

Figure 10: Shepard diagram plotting distances between points in the ordination scale vs. dissimilarities in the dissimilarity matrix. It shows a monotonic increasing tendency and represents only a random 10% of the pairs of data.

Perfect ordinal re-scaling of the data into distances is usually not possible. What is required is a scaling as good as possible. The goodness of fit of that configuration to the original similarities is measured by stress statistics, for instance Kruskal´s stress:

$$S = \sqrt{\frac{\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(\delta_{ij} - \hat{\delta}_{ij}\right)^{2}}{\sum_{i=2}^{n}\sum_{j=1}^{i-1}\delta_{ij}^{2}}}$$

Numerator is the residual sum of squares of monotone regression, and the denominator is the sum of squares of distances. By moving the samples in the low dimensional space we aim to reduce S as much as possible. As a rule of thumb, Kruskal stress values lower than 0.1 are considered good enough, and values higher than 0.2 are poor or misleading. The relation between the vegetation stands and environmental variables is measured by correlating the result of the NMDS with environmental variables.

There is a large variety of ordination techniques used to study ecological communities and carry out indirect gradient analyses. Minchin (1987) suggested Gaussian Ordination to be very sensitive to noise. He also argued that curvilinear distortions make linear techniques such as Principal Component Analyses or Principal Coordinates Analyses ineffective. He showed lack of robustness and important distortions in the ordinations when using Detrended Correspondence Analysis. Among all the possible scaling and indexes, NMDS was chosen because of the continuous nature of ecological communities. It achieves an adequate recovery of the underlying gradients that drive species composition and models them into few quantifiable dimensions (Minchin 1987).

The process was carried out combining the melastomes and pteridophytes data in PC-ORD v. 6 (MjM Software, Gleneden Beach, OR, USA) using the one-complement of the Jaccard similarity index as in Higgins et al. (2011, 2012). This dissimilarity measure expresses the number of species unique to either of the sites as a proportion of the total number of species in two sites. Melastomataceae were sampled only in 105 out of the 139 sampling sites, and using both plant groups for the analysis meant a reduction of samples from 139 to 105. Using both groups and fewer samples was preferred than using only pteridophytes but more samples for two reasons. The 34 sample sites discarded were situated along the road area where there is already a high density of floristic data per column number. The second reason is that using both plant groups would make a more robust floristic gradient. This single-axis ordination captured around 90% of the variation in floristic similarities in the original datasets, indicating

strong compositional patterns in these data and it is closely related with soil cation concentrations (Higgins et al. 2012).

3.4. Modeling the gradient

Once the three variables (DN, CN, and NMDS score) were processed, I compared how well a linear and a curved model explained the variability of DNs. All the analyses were carried out separately for each band in IBM SPSS Statistics version 20.

All the samples were taken from structurally similar forests, but of different floristic compositions. The detrending process used in Higgins et al. (2011, 2012) removes the gradient under the hypothesis in which the slope of the gradient is independent from the floristic composition. Thus the distribution of DNs in a three-dimensional space, being each dimension one of the variables, should be linear depicting a flat surface (Figure 11). If this hypothesis is confirmed, linear corrections of the image used before (Hansen et al. 2008, Higgins et al. 2011, 2012) could be valid to remove the noise produced by the radiometric gradient. Otherwise a more complex correction using exponential models would be required.

The linear model proposed is depicted as a flat surface determined by a simple polynomial:

$$F(DN) = a + b(NMDS) + c(CN)$$
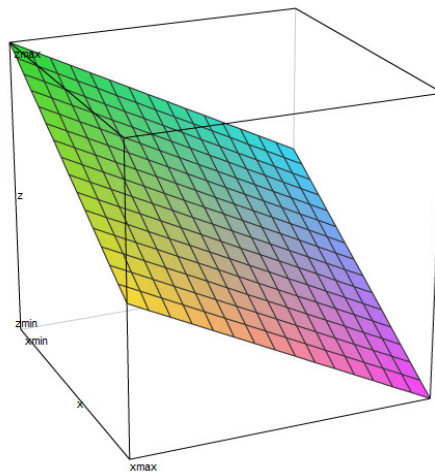
Figure 11: Distribution of DNs (z) on a three-dimensional space assuming linearity. Variable "x" is the Column Number and "y" the NMDS score. This model assumes that the slope in DNs along the East-West position is constant for any type of *tierra firme* forest (NMDS score) (LivePhysics 2013)

The linear model was compared to a cubic polynomial model which uses the terms suggested by Borcard et al. (1992). For the sake of simplicity I refer to this cubic

polynomial model as curved model, although theoretically it could adopt a staircase shape. The curved model is more complex than the linear one and is determined by the polynomial:

$$F(DN) = a + b(NMDS) + c(CN) + d(CN)*(NMDS) + e(CN)^2 + f(NMDS)^2 + g(CN)*(NMDS)^2$$
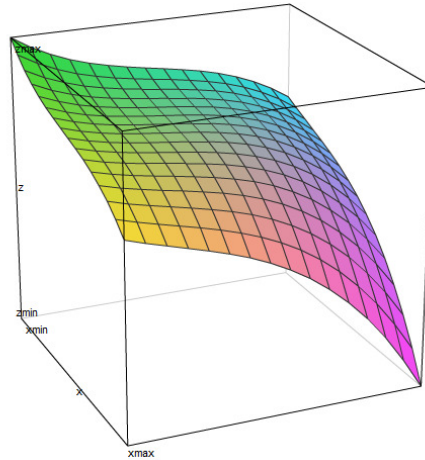$$+ h(CN)^2*(NMDS) + i(CN)^3 + j(NMDS)^3$$

Figure 12: Example of the distribution of DNs (z) on a three-dimensional space according to the curved model. Variable "x" is the Column Number and "y" the NMDS score. Here the slope in DNs along the east-west position caused by the radiometric gradient varies throughout the scene (LivePhysics 2013).

Because of its higher complexity, the curved model is expected to explain at least as much of the variation of DN values as the linear one. If this difference in explained variation is significantly larger in the curved model, it means that the gradient doesn´t behave equally throughout the image. Therefore the hypothesis of linearity on *tierra firme* forests cannot be accepted. In this case, the relation between the three variables would depict a curved surface, for example like the one in Figure 12. On the other hand, if both models explain a similar percentage of the variation it means that the gradient behaves linearly through the different *tierra firme* forests sampled. That is, its slope remains constant.

To create the curved model I used a step-wise method that identifies the group of predictors (terms of the regression of the cubic polynomial) that best fits the Landsat data sampled. This method conducts several consecutive regressions following a forward selection approach and a chosen model comparison criterion. For that it selects the predictor that explains better the distribution of DNs. Then it adds the next predictor that improves the model the most, repeating this process until none improves it more than a purely random predictor would do. Predictors chosen can be removed when another one is included to maximize the explanatory power and minimize the number of predictors. If

23

the exponential transformations of NMDS score and CN can explain a significant part of the variation that the original variables can´t, it means that there is a non-linear relationship between DN and CN. Also, if the interaction term (NMDS*CN) has a significant explanatory power it means that the slope of the regression between CN and DN varies according to the NMDS score. In both cases the gradient couldn´t be modeled accurately with a linear function. Mathematically more complex methods would be then required, and the probability of committing errors in the estimation increases with the complexity of the method.

3.5. Comparison of the slopes of the gradient

Parallel to the analyses of the curved and linear models, I calculated the slope of the gradient in different parts of the image using 40000 pixels scattered all over the study area, but with no relation to the floristic data. This was done for two reasons. First, because the DNs extracted from the segments cover just a small part of the variability of textures and colors of *tierra firme* forests, limited to the areas the researchers could reach when sampling the vegetation. These areas do not cover a large extension and chance variation may affect the conclusions. The second reason is because the multivariate regressions were executed using DNs from three scenes, two different dates, and two geological formations. Significant different slopes between images or geological formations may cause the gradient to show a non-linear trend despite normalizations.

I divided the study area into four regions (Figure 13); one geological formation per image. Remember that images 7_62 and 7_63 are actually one single image. The north part of image 8_62 showed some haze, and therefore that area was excluded from all analyses. To characterize the slope of each region I selected randomly 10000 random pixels for each one (therefore 40000 pixels altogether), and extracted their corresponding DN and CN.
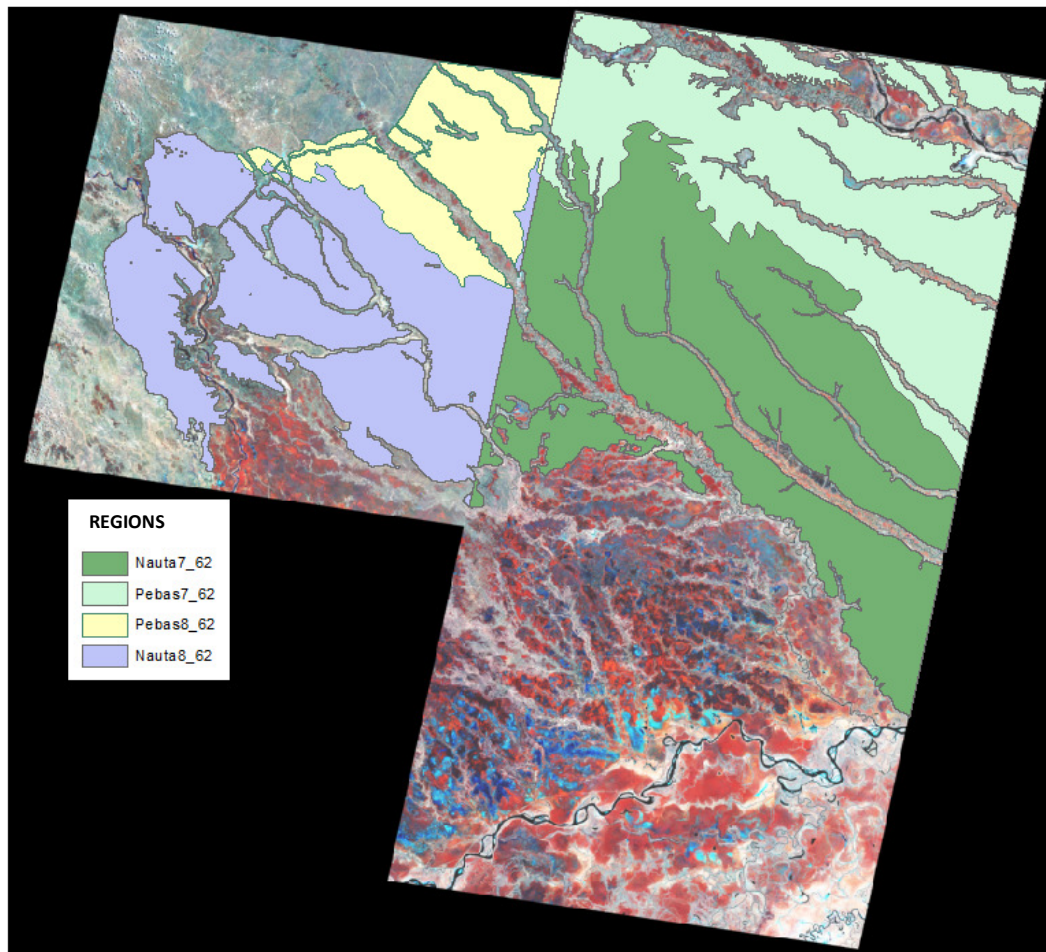
Figure 13: Map of the four regions defined according to the image and geological formation.

3.6. Analysis of the relation between NMDS score and CN

I calculated the correlation between NMDS score and CN and measured the significance of such relation by running a Monte Carlo permutation test. This method builds two matrices; one containing the floristic distances of each pair of transects, and another one containing the distances in column number of each pair of transects. The procedure involves the construction of 999 null distributions where the floristic distance matrix is held rigid and the other one is randomly permuted. A result of $p<0,001$ would mean a non- random distribution of the NMDS score in relation to the CN; i.e. the floristic composition varies gradually with the CN.

3.7. Detrending the surface

The removal of the gradient was done following two different detrending approaches. The first one corrects the average DN of each sampling site using a simple linear regression as in Hansen et al. (2008) generating new detrended DNs:

$$F(DN) = a + cCol\ Num$$

where "*c*" is the slope of the linear regression of the linear model, and "*a*" is the intercept. This approach works on the data extracted from the image, not on the actual image. Here only a trend with the scanning direction is removed and it is only valid for the sampling sites, not for the rest of the image.

The second approach is called image detrending (Higgins et al. 2011, 2012). It is based on trend surface analysis, which uses a scattered set of points to separate a mapped variable into two components; the regional trend and the local features. In this case it is used to separate the radiometric artifact (regional trend) from the DNs given by different forests (local features). Mathematically, the method is an interpolation that fits a polynomial surface by least-squares regression through the geographic coordinates of the sample points. It ensures that the sum of the square deviations from the trend surface is minimum. The polynomial can be expanded to any degree desired but since I am testing the efficiency of a linear model I will keep it at degree 1, generating a dipping flat plane. This is a simple method for describing large variations although it is susceptible to outliers. Note that unlike the former method, the trend modelled in this one is not bound to the scanning line. The pixels selected to interpolate the gradient were restricted to Nauta formation to avoid including in the modeled trend the difference between the high DNs of the Pebas formation in the north, and the low DNs of Nauta to the south. The process was carried out in Arc Map 10 and its workflow can be seen in Annex 3.

Once the trend surface was created, I subtracted it from the original layer (Figure 14). This generated values distributed around a mean of 0 which made it necessary to add to equation a constant value (values of unsigned 8 bit images must range from 0 to 255). This constant was the average between the mean values of the trend surface, and the original mean values of the sampling sites:

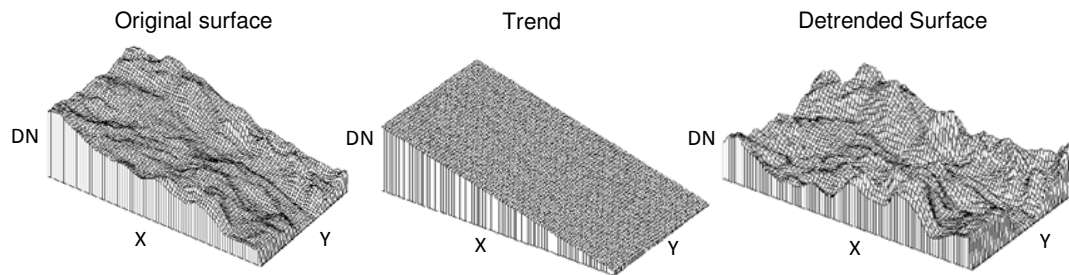$$Detrended\_bandX = Original\_bandX - Trend\_surface\_bandX + Constant$$

Figure 14: Graphic explanation of image detrending. First a raster representing the trend caused by the gradient is created out of the original image. Then that new image is subtracted from the original one, and a constant is added to eliminate negative values. This will produce a third image that depicts and enhances the local differences in DNs.

To understand better the effects of the detrending methods on the relation between floristic and Landsat data, I compared Hansen et al. (2008) equation and Higgins et al. (2012) image detrending. I performed regression analyses of the NMDS score and the DNs and used the parameter adjusted R^2 to test which detrending approach improved more the relation Landsat-floristic data, i.e. which one increased more the explanatory power of the NMDS score.

In Higgins et al. (2012) image detrending was applied on an image per image basis and without mosaicking, and in Higgins et al. (2011) the correction was applied to the whole mosaic. I have also applied the correction after mosaicking, although the images and the order they were mosaicked in differs from Higgins et al. (2011). Thus before presenting the main results, I analyze the outcomes of the processing and enhancement techniques used regarding to the DNs, colors, textures, and seams.

4. Results

4.1. Implications of the preprocessing of the images

4.1.1. Composition of the mosaic

Textural and color differences between adjacent images 7_62 and 8_62 were apparently only visual and more prominent in Pebas forests. They presented different tones between both images (see color composite image in Annex 1a and b). No sudden changes were visible in the DNs of the sampling sites with the change of scene for any of the bands when plotting DN vs. CN (Annex 1a and b). Spatial profiles of the image didn´t show any sudden changes in DNs around the seam lines either. There was

however a slight textural difference between the images noticeable mostly in band 2 (less in the other bands) in a surface profile (Figure 15).
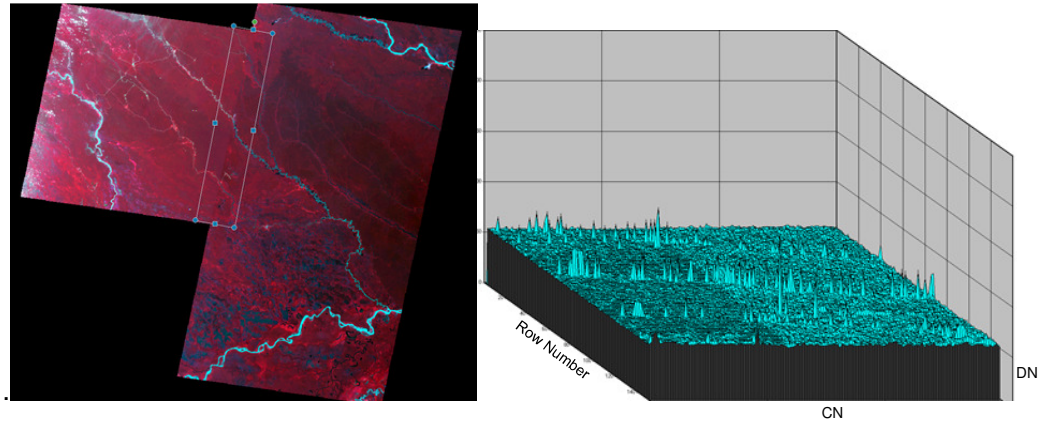
Figure 15: Spectral profile of the seam that links images 7_62 and 8_62 for band 2, the one that shows the greatest change in texture.

4.1.2. Segmentation

In spite of the textural and color difference observed between scenes 8_62 and 7_62, the seam between scenes didn´t seem to affect the generated segments (Figure 16). Boundaries between segments didn´t follow the seam between scenes.
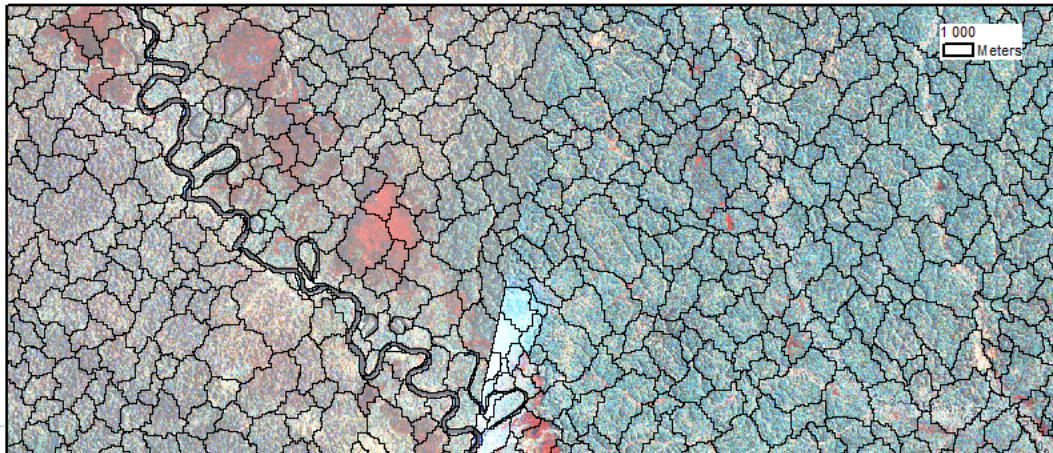
Figure 16: Detail of the seam in Tigre River. Segments make a distinction in the seam only where there are circumstantial differences between images, such as clouds, not between homogeneous stands of vegetation

4.2. Modeling the gradient

Table 3 summarizes the results of the regressions of the linear model and curved model.

Table 3: Results of the multivariate regression. Right side shows the results of a simple linear regression among the three main variables. DNs of each band are the dependent variables and NMDS score and CN the predictors. Left side represents the regressions resulting of the step-wise method and it uses the terms of the cubic polynomial as predictors. One predictor is added in each regression in decreasing order of added explanatory power. Adjusted R^2 represents the percentage of the variance explained by each model and it is the parameter used to compare their performances.

| Curved Models		Linear model	
Model Summary Band 1		**Band 1**	
Predictors	Adjusted R^2	Predictors	Adjusted R^2
CN	0,922	CN, NMDS	0,922
CN, CN^2	0,929		
CN, CN^2, CN^3	0,949		
CN, CN^2, CN^3, NMDS	0,951		
CN, CN^2, CN^3, NMDS, CN^2*NMDS	0,953		
Model Summary Band 2		**Band 2**	
Predictors	Adjusted R^2	Predictors	Adjusted R^2
CN	0,853	CN, NMDS	0,853
CN, CN^2	0,858		
CN, CN^2, CN^3	0,914		
CN, CN^2, CN^3, NMDS	0,929		
Model Summary Band 3		**Band 3**	
Predictors	Adjusted R^2	Predictors	Adjusted R^2
CN	0,841	CN, NMDS	0,847
CN, CN^2	0,866		
CN, CN^2, CN^3	0,889		
Model Summary Band 4		**Band 4**	
Predictors	Adjusted R^2	Predictors	Adjusted R^2
CN	0,774	CN, NMDS	0,878
CN, NMDS	0,878		
CN, NMDS, CN*NMDS	0,885		
CN, NMDS, CN*NMDS, CN*NMDS^2	0,900		
Model Summary Band 5		**Band 5**	
Predictors	Adjusted R^2	Predictors	Adjusted R^2
CN	0,934	CN, NMDS	0,955
CN, NMDS	0,955		
CN, NMDS, CN*NMDS	0,956		
CN, NMDS, CN*NMDS, CN^2*NMDS	0,959		
CN, NMDS, CN*NMDS, CN^2*NMDS, NMDS^3	0,961		
Model Summary Band 7		**Band 7**	
Predictors	Adjusted R^2	Predictors	Adjusted R^2
CN	0,931	CN, NMDS	0,949
CN, NMDS	0,949		
CN, NMDS, CN*NMDS	0,951		
CN, NMDS, CN*NMDS, CN^2*NMDS	0,955		

The exponential transformations of CN explained more variability than the linear model in the visual bands. Indeed, NMDS score didn´t contribute at all to the linear model in bands 1 and 2 and practically nothing in band 3. In the infrared bands CN and NMDS score explained around 90% of the variability. The curved model explained 2% more with respect to the linear one in band 4, but virtually nothing more in bands 5 and 7.

The multivariate regression was repeated individually for 8_62 and the pair 7_62-7_63. The results were very similar to the ones in table 3 regarding to the seleted predictors and their adjusted R^2.

A third regression was performed forcing the model to incorporate the predictors NMDS, CN, and CN^2, in that order. This revealed that the increase in explanatory power when adding CN^2 was only 2% in band 3, and nothing in the other bands. The same was done using the group of predictors NMDS, CN, and CN*NMDS obtaining that CN*NMDS practically didn´t increase the adjusted R^2 for any band. This can be seen in Table 3 for the infrared bands because of a coincidence with the results of the stepwise regression.

I executed another two stepwise regressions but this time selecting the segments located on Pebas and on Nauta formations separately. Here CN was still the first predictor chosen in all the bands. The NMDS score was chosen as the second one only in band 4 when looking at Nauta segments, and increased the adjusted R^2 2%. However, unique explanatory powers of the NMDS score were around 30% for the infrared bands in each geological formation.

4.3. Comparison of the slopes of the gradient

Table 4 shows the result of the analysis of the slopes of the four regions created and their comparisons. Differences in slopes between regions in the visual bands were 6%-30%. They were smaller in the infrared bands (less than 6%) except for band 4 in Pebas 8_62 which were around 15%. This region showed also similar differences when compared to Nauta 8_62.

Table 4: Slopes of each region and their differences in percentage. Slopes were calculated by regression of the DNs vs. CN of 10000 randomly picked pixels. The two last columns show the differences in slope between Pebas and Nauta within each scene. P=Pebas, N=Nauta

Band	P 8_62	P 7_62	P 8_62 & 7_62 difference (%)	N 8_62	N 7_62	N 8_62 & 7_62 difference (%)	P & N 8_62 difference (%)	P & N 7_62 difference (%)
1	-0,00118	-0,00101	14,24	-0,00153	-0,00108	29,33	22,78	6,30
2	-0,00067	-0,00072	5,98	-0,00090	-0,00082	8,28	24,41	12,33
3	-0,00078	-0,00062	20,76	-0,00094	-0,00066	30,43	17,28	5,79
4	-0,00154	-0,00180	14,51	-0,00180	-0,00188	3,83	14,80	4,15
5	-0,00180	-0,00180	0,38	-0,00184	-0,00185	0,37	2,65	2,64
7	-0,00074	-0,00077	4,53	-0,00075	-0,00080	5,77	1,86	3,14

The same process was executed using the DNs of the original images instead of the DNs of the mosaic (Table 5). I found similar proportions within images and much higher

differences in slopes between images, which means that the mosaicking adjusted also the slopes of the gradient to some extent.

Table 5: Differences in the slopes between units using the DNs of the images before mosaicking. P=Pebas, N=Nauta.

Band	P 8_62	P 7_62	P 8_62 & P 7_62 difference (%)	N 8_62	N 7_62	N 8_62 & N 7_62 difference (%)
1	-0,001273	-0,001036	18,61	-0,001586	-0,001064	32,91
2	-0,000460	-0,000727	36,72	-0,000624	-0,000819	23,80
3	-0,000387	-0,000615	37,07	-0,000445	-0,000617	27,87
4	-0,001359	-0,001862	27,01	-0,002167	-0,001875	13,47
5	-0,001794	-0,001789	0,27	-0,001951	-0,001826	6,40
7	-0,000658	-0,000737	10,71	-0,000618	-0,000775	20,25

4.4. Analysis of the relation between NMDS score and CN

Figure 17 shows a scatter plot of the NMDS score vs. CN. According to the Pebas-Nauta boundary detected in Higgins et al. (2011), sampling sites located on Pebas soils are depicted in green and the ones in Nauta in blue. The NMDS divides all the Nauta segments to one side and Pebas to the other, but for 4 exceptions rounded with a dash line. The upper circle contains two sampling sites that occur on Nauta formation but their floristic composition is more similar to the samples collected on Pebas soils. The lower circle contains two other sampling sites within the Pebas side whose floristic composition corresponds to Nauta soils. See Figure 18 to locate them in the study area and in detail in Figure 19 and Figure 20.
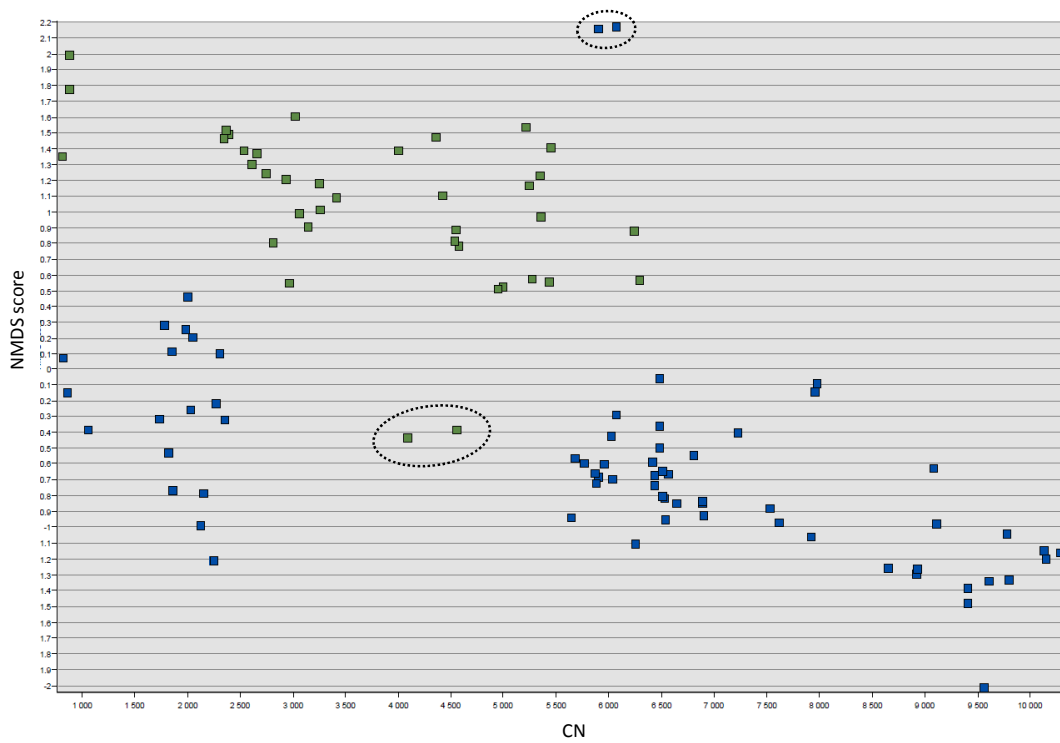
Figure 17: Scatterplot NMDS score vs. CN. Blue squares corresponds to the Nauta area and green to Pebas. Note here the uneven distribution of the sampling sites. Most of Nauta points are to the east, and most of Pebas to the west, creating a trend where the NMDS score decreases with CN. That trend also happens within each geological formation.
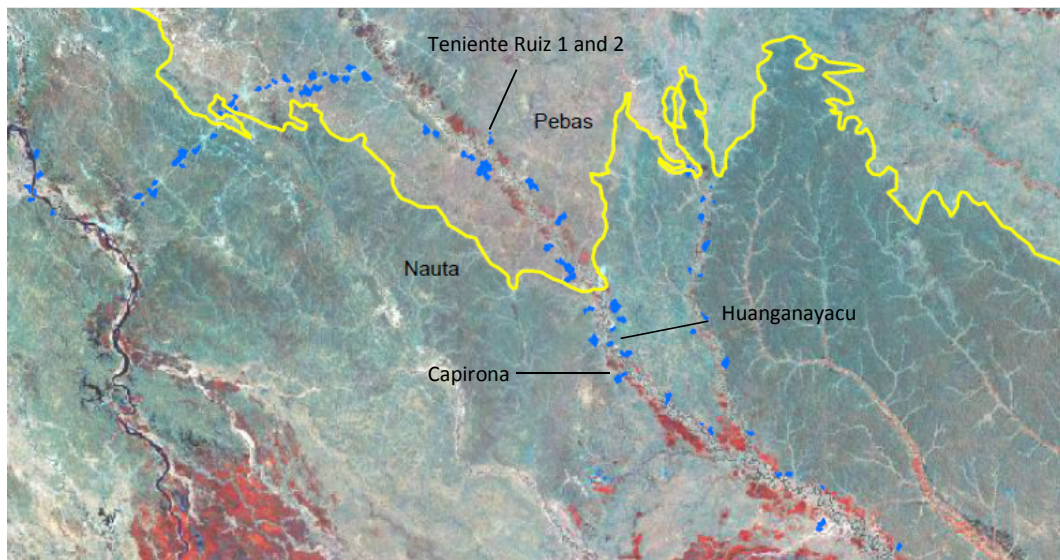
Figure 18: Study area with the location of the 4 sampling sites whose NMDS score doesn´t correspond to the geological formation where they are located.
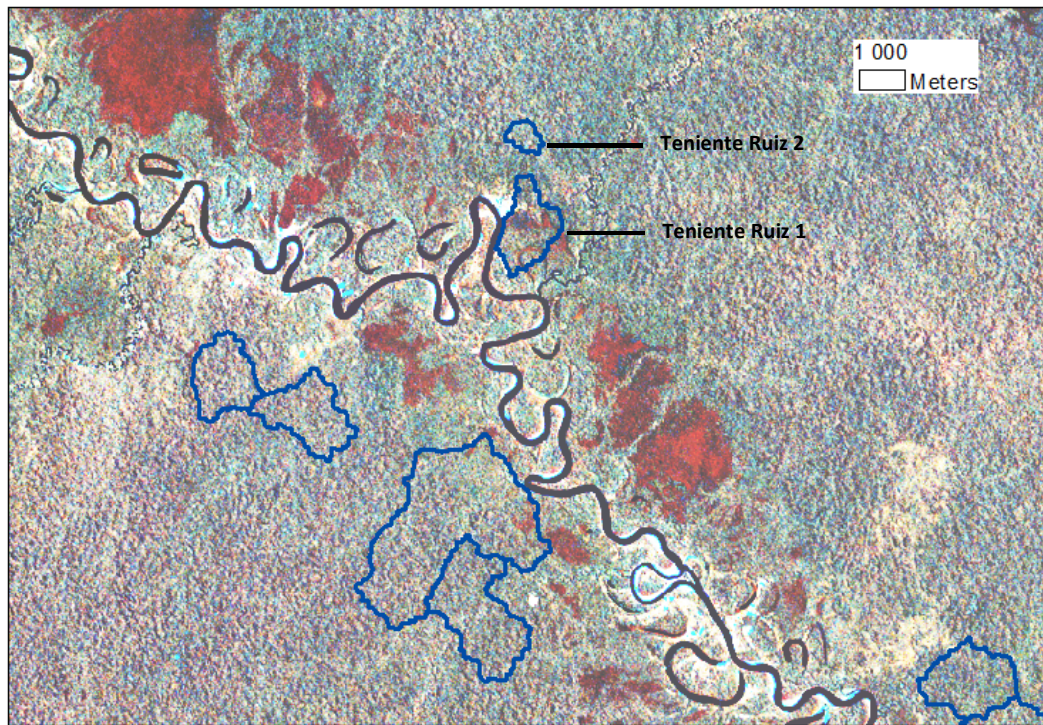
Figure 19: Detail of Annex 1b showing the two sampling sites in Pebas soils, but with Nauta-like floristic composition.

Figure 20: Detail of Annex 1b showing the two sampling sites in Nauta soils, but with Pebas-like floristic composition.

In Higgins et al. (2011), a cation concentration of 2.09 cmol(+)/kg (sum of the concentrations of Ca, Mg, Na and K) is considered to be a threshold between Nauta (below that threshold) and Pebas (above it). According to the soil analysis in Higgins et al. (2011), the transects Teniente Ruiz 1 and 2 had concentrations of 0.34 cmol(+)/kg and 2.10 cmol(+)/kg respectively, corresponding to Nau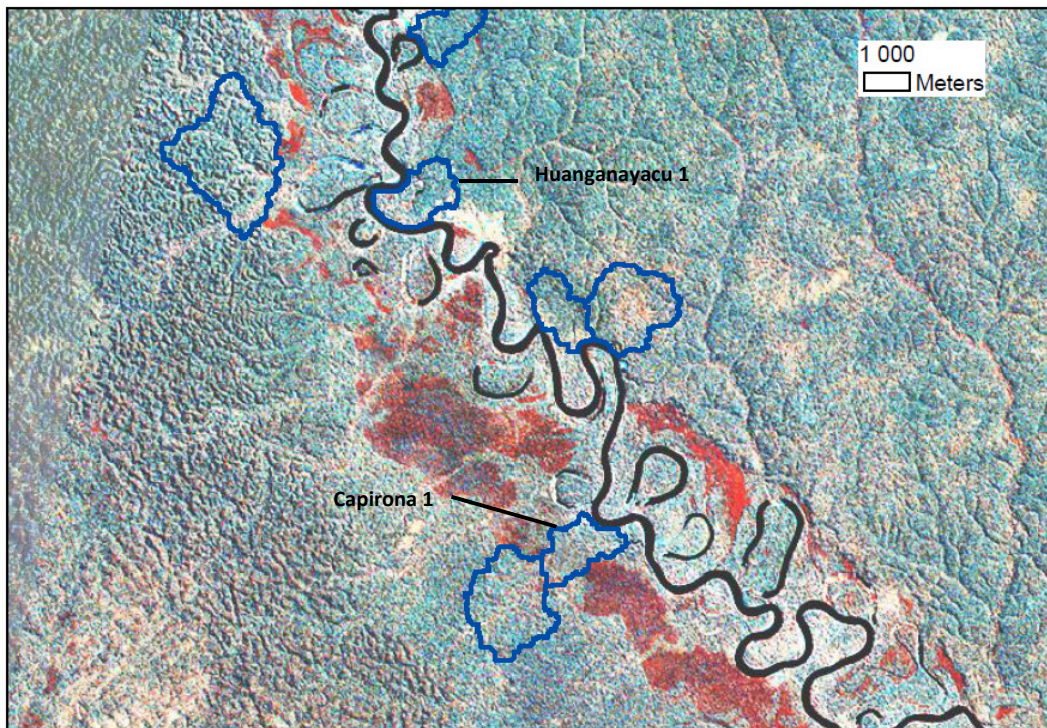ta soils. Furthermore, Huanganayacu 1 and Capirona 1 had concentrations of 8.1 cmol(+)/kg and 5.8 cmol(+)/kg respectively, similar to other Pebas soils.

CN explained in a linear regression 33% of the variability of the NMDS scores and this dependence was statistically significant (the Monte Carlo test with 999 permutations gave a $p<0.001$).

4.5. Detrending the surface

The results of image detrending can be seen in detail in Annex 1b, visually in the color composite map and graphically in the scatter plot DN vs. CN.

Figure 21 shows two scatter plots of the original DNs and detrended DNs vs. the NMDS score. Distinct forest types (Pebas or Nauta) yield more different DNs with the original than with the detrended data. That is, different types of *tierra firme* forests have more similar DNs after the detrending. This might reduce their separability in the image. However, Annex 1 shows that visual contrast between both geological formations is higher after detrending.

To know more about the effect of the detrending on the floristic information I compared the relation between the floristic and remote sensing data using both detrending approaches (Table 6).

When I removed the effect of the gradient using Hansen et al. (2008) equation, I found a dramatic decrease in the explanatory power of the NMDS score in the visual bands, and a moderate decrease in infrared bands. The detrending almost eliminated the explanatory power of the NMDS score in the visual bands, meanwhile the reduction ranges only between 3% and 9% for infrared bands.

When detrending the surface using the 10000 random pixels to interpolate the trend as in Higgins et al. (2012), I found that it worsened even more the relationship between remote and floristic data. I detrended here only the infrared bands to reduce unnecessary work.
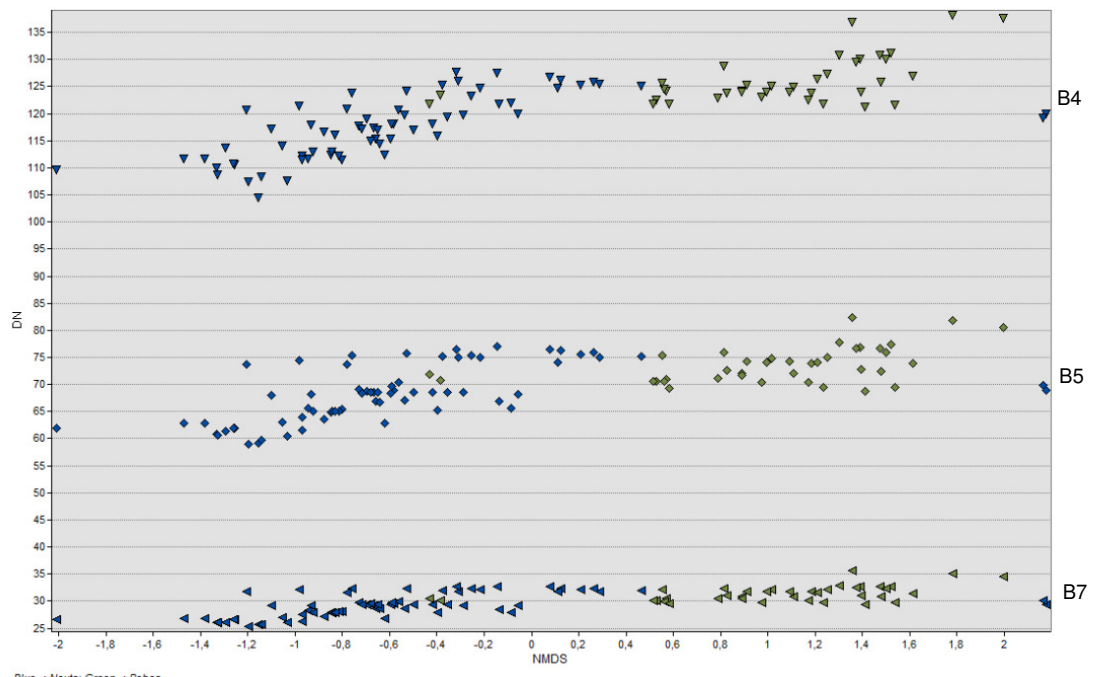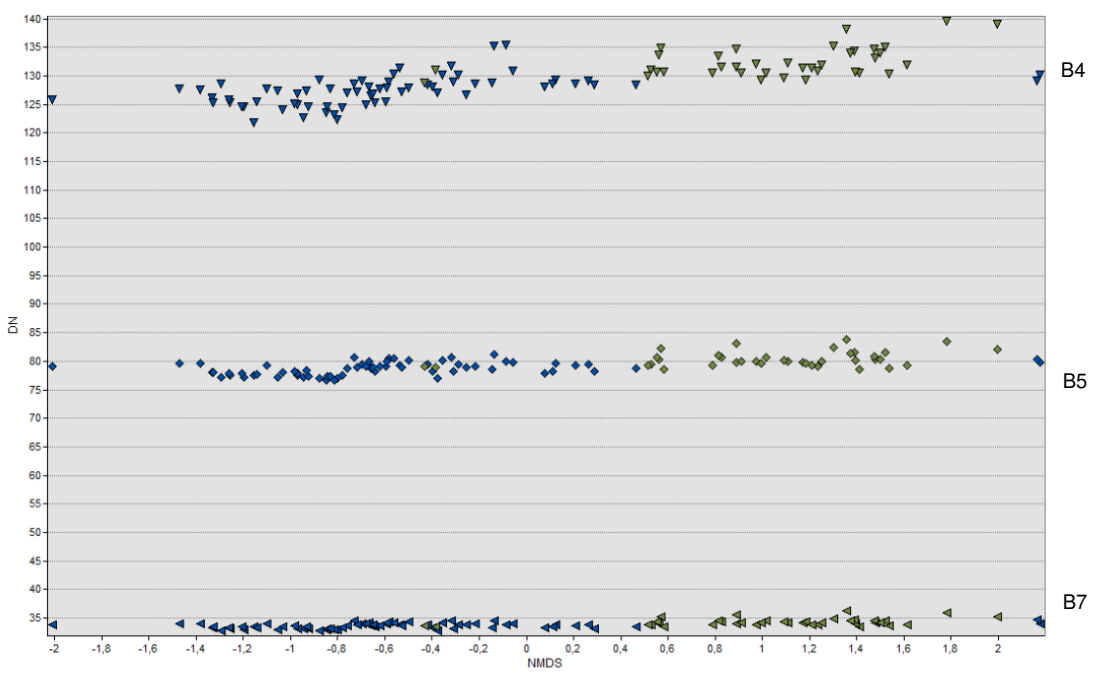
Figure 21: Scatter plot of sampling sites´ DNs vs NMDS score with original (a) and detrended (b) values, being blue points Nauta, and green Pebas. Note how the detrending compresses the range of DNs.

Table 6: Relationship between Landsat and floristic data after different types of detrending treatments. The adjusted R^2 shows the explanatory power of the NMDS score in the variance of DNs for the 6 Landsat bands.

Band	Original adjusted R^2	Detrended using Hansen et al. (2008) equation adjusted R^2	Detrending using 10000 random pixels adjusted R^2
B1	0,301	-0,009	
B2	0,321	0,009	Analyses not performed
B3	0,207	0,061	
B4	0,600	0,572	0,504
B5	0,459	0,407	0,341
B7	0,445	0,359	0,303

The analysis was repeated with a histogram equalization of the detrended mosaic. This returned generally lower NMDS score explanatory powers: 0.505, 0.336 and 0.230 for bands 4, 5 and 7 respectively.

In infrared bands, reductions in NMDS score adjusted R^2 were found also when studying individually each geological formation; from around 30% it decreased to 5%.

The trends modeled with each approach had different directions (Figure 22). These directions were also different between bands when using the 10000 random pixels despite these were the same for all the bands.
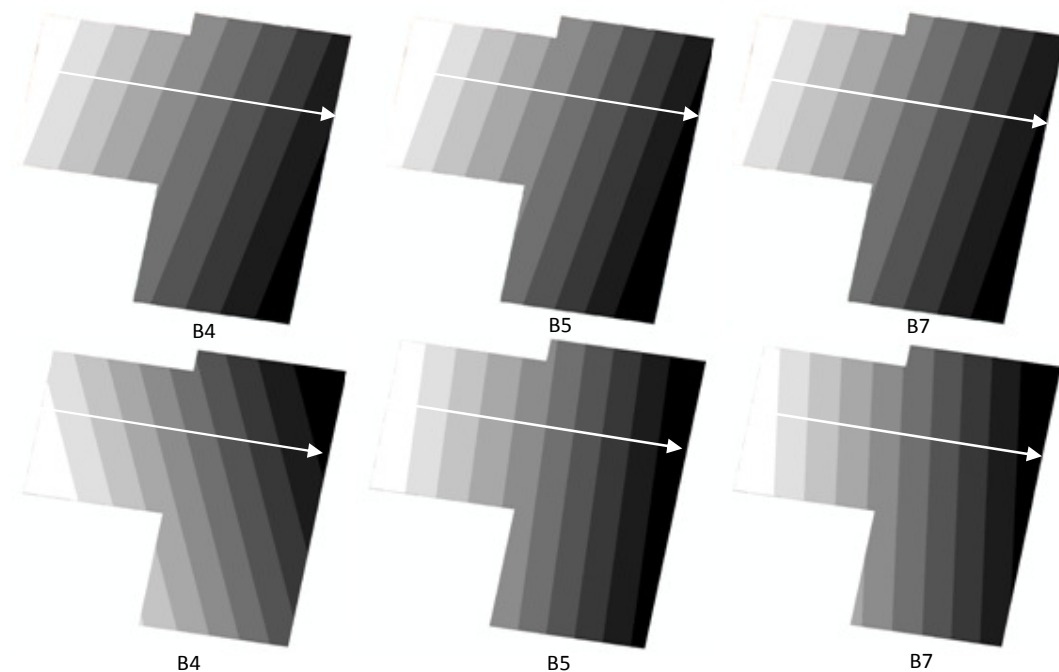
Figure 22: The upper row shows the trends modeled using Hansen et al (2008) equation. This is, using the DN and position of the segments. The lower row shows the trends modeled using the 10000 random pixels scattered all over Nauta forests. The white arrow represents the scanning direction. The first approach generated slopes tilted to the right of the scanning direction. This was expected because Pebas sites are to the north and have higher DNs. The second approach generates slopes with a similar direction than the scanning for bands 5 and 7, but considerably tilted to the left in band 4.

5. Discussion

5.1. Implications of the preprocessing of the images

The textural difference between the scenes seen in Figure 15 might be caused by multiple reasons. The first one is differences in the sensor; 8_62 corresponds to Landsat 5 (left image, more homogeneous texture) and 7_62 and 7_63 correspond to Landsat 7. Another possibility for this difference is that 7_62 and 7_63 were taken the year after the El Niño event 97-98, the strongest drought ever recorded which might have increased the variability of the spectral responses of the vegetation cover, showing then a more heterogeneous texture. A third possibility is the existence of different atmospheric conditions since images were taken in two different dates.

Although in this first inspection I didn´t find notorious evidence that the preprocessing might distort the results of the analysis, the way the mosaic is composed produced an interesting effect on the results of the statistical analyses. This is discussed in chapter 5.3, Comparison of the slopes of the radiometric gradient: the possible relevance of mosaicking direction.

5.2. Modeling the gradient

Over all, the curved model was 3%-8% better than the linear one in visual bands. The results also show that the NMDS score and the visual bands information are basically unrelated. This is in accordance with the results of Higgins et al. (2012) who found that "no information about floristic patterns is lost by excluding bands 1-3 from image display and manual image interpretation". With respect to the infrared bands, the curved model worked only from 0.6% to 2% better than the linear one. This improvement is given by the interaction term (NMDS*CN) and its exponential transformations. This suggests that the slope of the gradient varies slightly according to the NMDS score.

In practical terms the higher explanatory powers achieved by the curved model in the infrared bands seem insignificant. A third order function naturally fits better through the sampling points than a first order function. Therefore it should explain the variability of the data at least as well as the first order function. The small improvement of the curved model with respect to the linear one can be disregarded. Thus, the gradient shows here a rather certain linear behavior across different types of *tierra firme* forests for the images used. Consequently, a linear correction such as image detrending may be a possible solution. However, such correction implies other additional considerations that I develop in the following sections.

37

5.3. Comparison of the slopes of the radiometric gradient: the possible relevance of mosaicking direction.

Nauta soils are generally less fertile and showed steeper slopes than Pebas. Image 7_62 appears to contain less vigorous vegetation (darker tones in the infrared bands) and both geological formations showed steeper slopes there than in 8_62 (Table 4). Besides, the results of the multivariate step-wise regression showed a small but still significant explanatory power of the term NMDS*CN. These results suggest that the radiometric gradient could be stronger in less fertile/vigorous forests. This hypothesis should be checked in other areas with other images, but if the tendency repeats, the contrast between forests growing on richer and on poorer soils would be enhanced towards the east. It is possible though that the steeper slope detected in Nauta compared to Pebas is only a spurious observation; it is naturally unlikely to achieve the exact two same slopes.

In the following lines I explain with a graphical simulation the importance that the direction of the mosaicking has had in this case. For that I take as an example band 4 in image 7_62. The average DNs of Pebas and Nauta are 113 (± 6,70) and 110 (± 7,12) respectively, which accounts for a difference of 3 DNs between pixels of both formations. Being their respective slopes 0,00180 and 0,00188, in one single image pixels in Pebas and Nauta would differ 0,48 DN more on the east margin respect to the west. This difference is insignificant when dealing with one or two images but it might become relevant when increasing the number of images. Let´s suppose that we mosaicked 6 scenes eastwards and that all the images followed the same pattern of slopes and average DNs than in image 7_62. As figure 23 shows, the difference between Pebas and Nauta at the west margin of the reference image would be 3 DNs. By the end of the 5th image added this difference would be more than doubled if mosaicking eastwards.

Conversely, if more images were added westwards Pebas and Nauta DNs would get closer and closer, and the initial difference of 3 DNs would be halved by the end of the 5th image added. These relative differences would persist after the linear detrendings I performed.
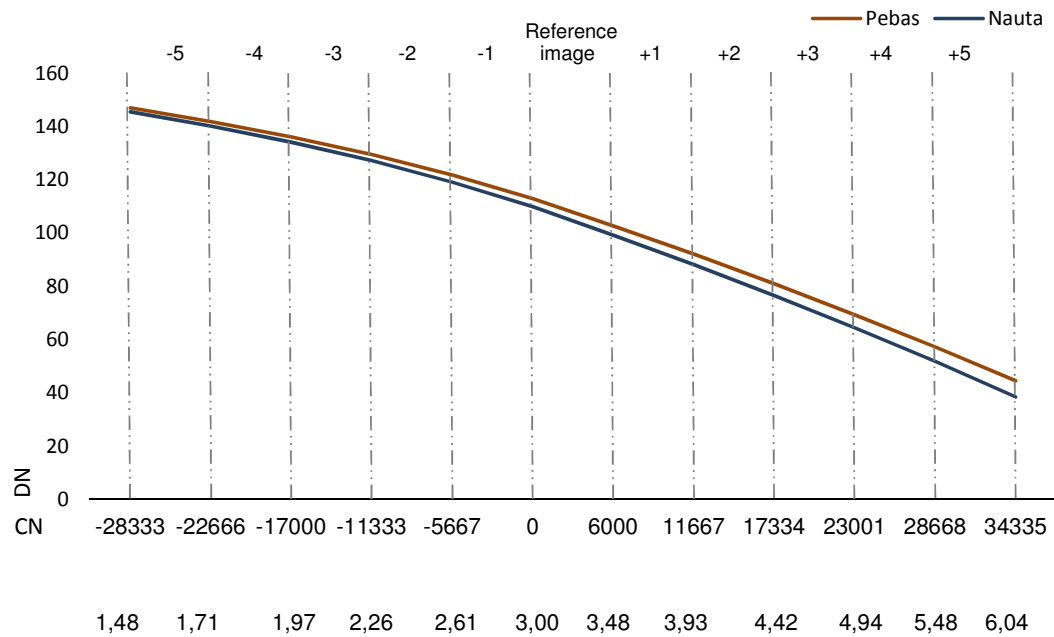
Figure 23: Effect of the difference in slopes between Pebas and Nauta when mosaicking east- and westwards, starting at the reference image. First average DNs and slopes were calculated using band 4 from image 7_62. The rest is a hypothetical extrapolation assuming that all the images follow the same DN and slope patterns than 7_62. Each column represents one image added, which means 5667 columns added to account for the overlapping areas between images. The lower row of numbers shows the difference between Pebas and Nauta at each overlapping area. It increases eastwards and decreases westwards.

The mosaic in Higgins et al. (2011) was constructed eastwards which could have improved the separability of different forest types. My mosaic was built westwards which might have compressed the DNs of image 8_62. Note that this phenomenon wouldn´t occur only between Pebas and Nauta forests, but between high and low pixel values. Therefore, it might be possible to use this phenomenon as a "trick" to enhance the spectral differences between forest types by mosaicking eastwards.

To corroborate my prediction I intended to perform the same regression mosaicking the images eastwards (that is, setting 8_62 as reference image). However, the created mosaic presented a systematic error that made impossible any analysis. Some of the pixels got value 1 for some of the bands with no apparent logical reason. Figure 24 shows an east-west spatial profile to illustrate the phenomenon. The seam between scenes is at 5000 m, where the error starts. The pixels with value one occur only in the adjusted image and their frequency increase toward the east. The error was found to happen in 3 out of 6 pairs of other images analyzed, always when mosaicking eastwards. No explanation was found until the writing of this thesis, but in the future other software could be used to mosaic instead of ERDAS IMAGINE.
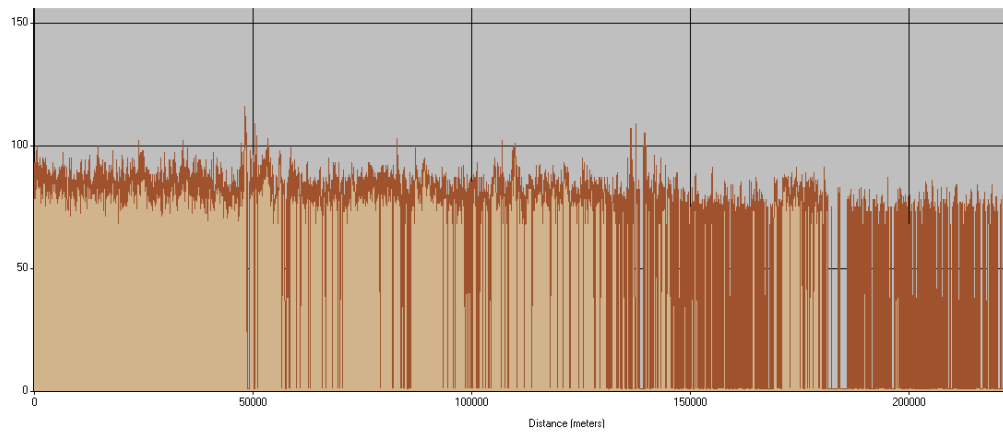
Figure 24: Spatial profile of the mosaic that shows the phenomenon produced when mosaicking eastwards in ERDAS IMAGINE. In the adjusted image some of the pixels get value 1.

5.4. Analysis of the relation between NMDS score and CN

All but 4 of the sampling sites were placed at opposite sides in the NMDS axis depending on the geological formation they belong to according to the satellite image. The four exceptions were explained with their soil cation concentrations. This supports the results of Higgins et al. (2011, 2012), where geological formations were found to determine the species composition and the NMDS score was strongly associated with the soil cation content.

Geological formations and other environmental factors that drive species composition may vary also along space and in this study case they seem to have a relevant spatial component in the scanning direction. This might pose a problem when detrending because the process identifies a regional trend in DNs and removes it without distinguishing whether it is a natural trend or product of the radiometric artifact.

As it was shown in Figure 22 the trends modeled with image detrending tilted a bit to the left of the scanning line in bands 5 and 7, and considerably more in band 4. The modelled trend should in theory reflect only the scanning direction. Contradictorily, band 4 is the one that shows the highest explanatory powers of the NMDS score for both approaches. Although the trend in band 4 deviates that much from the scanning direction, its angle matches with one of the two most frequent directions of the gradient found in the visual analyses of Toivonen et al. (2006). If one wants to constrain the slope of the trend generated to the scanning line, a script in Arc Map could be designed for such purpose. However it wouldn´t help if a natural trend coincides with the scanning line, as apparently happens in my case (the Monte Carlo permutation test between CN and NMDS score gave p<0.001). In my study area soil cation content decreases generally eastwards. Besides, the distribution of the segments over the image also follows this pattern: 39 of

the 105 segments belong to Pebas (with higher cation content and DNs) and are on the west half of the mosaic, and 47 segments belong to Nauta and are on the east half (Figure 17). This causes a natural gradient in pixel values with the same direction than the radiometric gradient. This natural gradient has probably been included in the modeled trend and removed with the detrending, decreasing the explanatory power of the NMDS score in both detrending approaches. In the next section I explain more thoroughly the challenges that it poses when linking Landsat and floristic data.

5.5. Detrending the surface

Image detrending improved the visual contrast of the image (Annex 1). However, both detrending approaches decreased the explanatory power of the NMDS score in infrared bands between 4% and 10%. These results are unlike the ones got by Higgins et al. (2012) who achieved increments for most of the cases. The explanation for it could be a sampling effect. The NMDS score decreased with CN considering both geological formations simultaneously and also within each of them (Figure 17). Apparently, the higher is my NMDS score the higher is the DN value. Therefore the higher the CN the lower the DNs for two reasons: because of the radiometric gradient and because of the distribution of the sampling sites (lower cation content to the east). This causes detrending operations to bring back closer the DN values of different forest types. Put it in another way, the radiometric gradient was enhancing artificially the spectral differences between different types of forests because of the distribution of the sampling sites. This could help to explain the results obtained by Higgins et al. (2012). They compared the relation between Landsat and floristic data in three sites of similar soil and vegetation characteristics in the north of Peru, before and after the detrending. The results were different in each of the sites and are summarized in Table 7.

Table 7: Results of Higgins et al. (2012). The table shows the NMDS score explanatory power (called coefficient of determination in their article) before and after detrending, and their differences (grey columns).

	Pastaza-Tigre			Curaray			Sucusari		
Band	Raw	Detrended	Diff. (%).	Raw	Detrended	Diff. (%)	Raw	Detrended	Diff. (%)
1	0,32	0,11	-0,21	0,01	0,01	0	0,27	0,24	-0,03
2	0,07	0,4	0,33	0,04	0,06	0,02	0,56	0,75	0,19
3	0,40	0,23	-0,17	0,10	0,11	0,01	0,17	0,11	-0,06
4	0,68	0,81	0,13	0,55	0,56	0,01	0,34	0,24	-0,1
5	0,40	0,55	0,15	0,03	0,20	0,17	0,55	0,39	-0,16
7	0,47	0,58	0,11	0	0,12	0,12	0,64	0,50	-0,14

The detrending increased the explanatory power of the NMDS score in Pastaza-Tigre and Curaray, but decreased it in Sucusari. The decrease in Sucusari was similar to the one I observed when detrending my mosaic, so it was the distribution of sampling sites: Nauta to the east and Pebas to the west. In their study site Pastaza-Tigre the detrending improved the explanatory power of the NMDS score. There, Nauta sites were found to the west half and Pebas to the east, contrary to my case. The radiometric gradient was enhancing artificially the spectral differences among sampling sites in Sucusari, and mixing them at Pastaza-Tigre. Figure 25 illustrates how image detrending can improve or worsen the spectral separability of the sampling sites depending on their spatial distribution in relation to the column number.
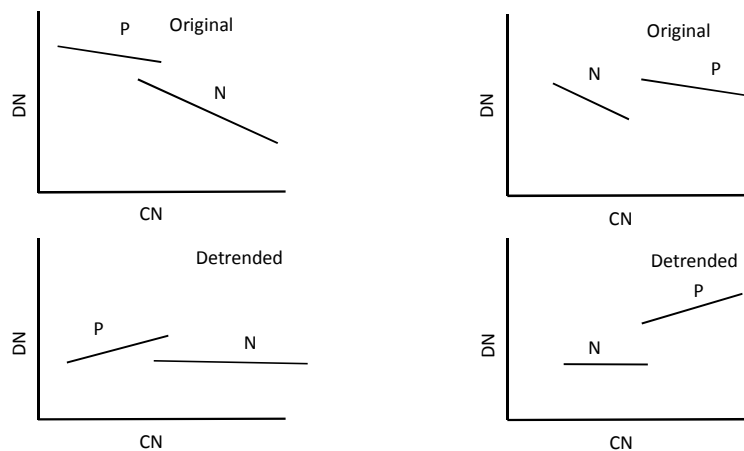
Figure 25: Image detrending may enhance or worsen the spectral separability of different forest types depending on the distribution of the sampling sites. The left graphs show the effect of image detrending when Nauta sampling points are in the east of the image and Pebas to the west. Such is the case of this experiment and that of Sucusari in Higgins et al. (2012). Image detrending reduces the artificial diferences between both geological formations. Pebas and Nauta have been drawn with different slopes, but the effect would be the same if they were similar.

Right graphs show the same representation but when Pebas are found to the east of the image and Nauta to the west. Such is the case of Pastaza-Tigre in Higgins et al. (2012). Here the detrending improves the spectral separability of Pebas and Nauta.

In Curaray sampling wasn´t that stratified and Pebas sites were overrepresented. Thus, it wasn´t included here.

The detrending I performed worsened also the NMDS score and DNs relation when executing the regression within each geological formation. That artificial "stretching" produced by the radiometric gradient seems to have also enhanced the differences between sampling sites within geological formations. Because of a conjunction between radiometric and natural gradients, when the artifact is removed the image is also "overcorrected". Along with the artifact, natural environmental variation between, and probably within, geological formations may also be eliminated.

A decreasing cation content towards the east in Amazonia is common (ter Steege et al, 2006, Pitman et al. 2008, Higgins et al. 2012). Some studies (Higgins et al. 2012, Sirens et al. 2013 among others) have presented strong correlations between DN values and soil cation content (directly measured or inferred through indicator species). These correlations might be overestimated because of this conjunction between vegetation and radiometric gradients and it would be something worth looking at.

Another approach that hasn´t been tested here is the use of ratios instead of single bands to reduce the effect of the gradient. For instance, band 7 showed a much smaller slope in the gradient and lower DN values than bands 4 and 5. Perhaps a ratio 4/7 could remove the artifact to some extent. Bands 4 and 5 seem to have very similar slopes for Pebas and Nauta, and substracting band 5 from 4 could also remove the effect of the gradient to some extent.

6. Conclusions

Visual bands showed weak relationships with the NMDS score, and different slopes in different parts of the mosaic. Infrared bands can be used to model the gradient and apply a linear correction. If the difference in slope found between Pebas and Nauta forests is a constant pattern and not only a spurious observation, mosaicking images towards the east could improve the separability of forests growing on poor and richer soils, or perhaps according to their health status or vigor.

The existence of a natural vegetation gradient with the same direction as the scanning line poses an additional challenge when studying vegetation patterns. Trend surfaces can´t distinguish what part of the DNs variation is due to the radiometric artifact and what part is due to environmental drivers. Moreover, it is possible that this conjunction between natural and artificial variation in DNs had overestimated previous correlations between field information and DN values.

7. References

Asner GP (1998) Biophysical and Biochemical Sources of Variability in Canopy Reflectance. Remote Sensing of Environment 3: 234-253

Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. Ecology 73: 045-1055.

Chambers JQ, Asner GP, Morton DC, Anderson LO, Saatchi SS, Espírito-Santo FDB, Palace M, Souza CJr (2007) Regional ecosystem structure and function: ecological insights from remote sensing of tropical forests. Trends in Ecology and Evolution 22: 414-423

Chander G & Markham B (2003) Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. IEEE Transactions on Geoscience and Remote Sensing 41: 2674-2677.

Chao A, Chazdon RL, Colwell RK, Shen T-J (2005) A new statistical approach for assessing compositional similarity based on incidence and abundance data. Ecology Letters 8: 148-159.

Chavez PS (1996) Image-based atmospheric corrections—revisited and revised. Photogrammetric Engineering and Remote Sensing 62: 1025-1036.

Clark ML, Roberts DA, Clark DB (2005) Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. Remote Sensing of Environment 96: 375-398.

Collett LJ, Goulevitc BM, & Danaher TJ (1998) SLATS Radiometric Correction: A semi-automated, multi-stage process for the standardization of temporal and spatial radiometric differences. Proceedings of the 9th Australasian remote sensing and photogrammetry conference, Sydney, Australia, July 1998.

Danaher TJ (2002). An empirical BRDF correction for Landsat TM and ETM+imagery. Proceedings of the 11th Australasian remote sensing and photogrammetry conference, Brisbane, Australia, September 2002.

Encarnación F (1985) Introducción a la flora y vegetación de la Amazonia peruana: estado actual de los estudios, medio natural y ensayo de una clave de determinción de las formaciones vegetales en la llanura Amazónica. Candollea 40: 237-252

Gauch HG (1982) Multivariate Analysis in Community Ecology, Cambridge Studies in Ecology.

Grime JP (2001) Plant strategies, vegetation processes, and ecosystem properties. Second edition. John Wiley and Sons, Chichester, UK.

Hansen MC, Roy D P, Lindquist E, Adusei B, Justice CO, 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.

Higgins MA, Ruokolainen K, Tuomisto H, Llerena N, Cardenas G (2011) Geological control of floristic composition in Amazonian forests. Journal of Biogeography 38: 2136-2149.

Higgins MA, Asner GP, 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.

Hill RA & Foody GM (1994) Separability of tropical rain-forest types in the Tambopata-Candamo Reserved Zone, Peru. International Journal of Remote Sensing 15:2687-2693.

INGEMMET (2000) Mapa geológico del Perú. Lima: Instituto Geológico Minero y Metalúrgico.

Kalliola R, Ruokolainen K, Tuomisto H, Linna A, Mäki S (1998) Mapa geoecológico de la zona de Iquitos y variación ambiental. - In: Kalliola R & Flores Paitán S (eds.) Geoecología y desarrollo Amazónico: estudio integrado en la zona de Iquitos, Perú. Annales Universitatis Turkuensis Ser. A II. pp. 443-457.

Kalliola R & Flores Paitán S (Eds.) Geoecología y desarrollo Amazónico: estudio integrado en la zona de Iquitos, Perú. Annales Universitatis Turkuensis Ser A II, vol. 114. (pp. 443–457).

Kastner TP & Goñi MA (2003) Constancy in the vegetation of the Amazon Basin during the late Pleistocene: evidence from the organic matter composition of Amazon deep sea fan sediments. Geology 31: 291-294.

Kristiansen T, Svenning J-C, Pedersen D, Eiserhardt WL, Grández C, Balslev H (2011) Local and regional palm (Arecaceae) species richness patterns and their cross-scale determinants in the western Amazon. Journal of Ecology 99: 1001-1015.

Landsat User´s Handbook (2002). Landsat Project Science Office at NASA's Goddard Space Flight Center, Maryland. http://landsathandbook.gsfc.nasa.gov/pdfs/Landsat7_ Handbook.pdf. Accessed 29.09.2013.

Live Phisics (2013) http://www.livephysics.com/tools/mathematical-tools/online-3-d-function-grapher/.

Marengo JA (1998) Climatología de la zona de Iquitos, Perú In: Paitán RKASF, editor. Geoecología y desarrollo Amazónico: estudio integrado en la zona de Iquitos, Perú. Finland: University of Turku; pp. 35-57.

Minchin PR (1987) An evaluation of the relative robustness of techniques for ecological ordination- Vegetatio 69: 89-107.

Pellikka P (1998) Development of correction chain for multispectral airborne video camera data for natural resource management. Fennia 176: 1-110.

Pires JM & Prance GT (1985) The vegetation types of the Brazilian Amazon. Pp. 109-145 in Prance GT & Lovejoy TE (eds). Key environments: Amazonia. Pergamon, Oxford.

Pitman NCA, Mogollón H, Dávila N, Ríos M, García-Villacorta R, Guevara J, Baker TR, Monteagudo A, Phillips OL, Vásquez-Martinez R, Ahuite M, Aulestia M, Cardenas D, Ceron CE, Loizeau PA, Neill DA, Núñez VP, Palacios WA, Spichiger R, Valderrama E. (2008) Tree community change across 700 km of lowland Amazonian forest from the Andean foothills to Brazil. Biotropica 40: 525-535.

Phillips OL, Núñez Vargas P, Monteagudo AL, Cruz AP, Zans MEC, Sánchez WG, Yli-Halla M, Rose S (2003) Habitat association among Amazonian tree species: a landscape-scale approach. Journal of Ecology 91: 757-775.

Rajaniemi S, Tomppo E, Ruokolainen K, Tuomisto H (2005) Estimating and mapping pteridophyte and Melastomataceae species richness in Western Amazonian rain forests International Journal of Remote Sensing 26: 475-493.

Raši R, Bodart C, Stibig H-J, Eva H, Beuchle R, Carboni S, Simonetti D, Achard F (2011) An automated approach for segmenting and classifying a large sample of multi-date Landsat imagery for pan-tropical forest monitoring, Remote Sensing of Environment 115: 3659-3669.

Riaño D, Chuvieco E, Salas J, Aguado I (2003) Assessment of different topographic corrections in Landsat-TM data for mapping vegetation types. Geoscience and Remote Sensing, IEEE Transactions, vol. 41, no.5, pp.1056-1061.

Ruokolainen K & Tuomisto H (1998) Vegetación natural de la zona de Iquitos. In: Kalliola R & Flores Paitán S (eds.) Geoecología y desarrollo amazonico: estudio integrado en la zona de Iquitos, Perú. Annales Universitatis Turkuensis Ser A II, vol. 114, pp. 253-365.

Räsänen M, Neller R, Salo J, Jungner H (1992) Recent and ancient fluvial deposition systems in the Amazonian foreland basin, Peru. Geological Magazine 129: 293-306.

Räsänen M, Salo JS, Jungner H, Pittman LR (1990) Evolution of the western Amazonian lowland relief: impact of Andean foreland dynamics. Terra Nova 2: 320-332.

Räsänen ME, Linna AM, Santos JCR, Negri FR (1995) Late Miocene tidal deposits in the Amazonian foreland basin. Science 269: 86-390.

Salo J, Kalliola R, Hakkinen I, Makinen Y, Niemela P, Puhakka M, Coley PD (1986) River dynamics and the diversity of Amazon lowland forest. Nature 322: 254-258.

Salovaara KJ, Thessler S, Malik RN, Tuomisto H (2005) Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery. Remote Sensing of Environment 97: 39-51.

Shepherd JD & Dymond JR (2000) BRDF Correction of Vegetation in AVHRR Imagery. Remote Sensing of Environment 74: 397-408.

Sirén A, Tuomisto H, Navarrete H (2013) Mapping environmental variation in lowland Amazonian rainforests using remote sensing and floristic data. International Journal of Remote Sensing 34: 1561-1575

ter Steege H, Pitman NCA, Phillips OL, Chave J, Sabatier D, Duque A, Molino JF, Prévost MF, Spichiger R, Castellanos H, von Hildebrand P, Vásquez R (2006) Continental-scale patterns of canopy tree composition and function across Amazonia. Nature 443: 444-447.

Terborgh J & Andresen E. (1998) The composition of Amazonian forests: patterns at local and regional scales. Journal of Tropical Ecology 14: 645–664.

Thessler S, Sesnie S, Ramos Bendaña ZS, Ruokolainen K, Tomppo E, Finegan B (2008) Using k-nn and discriminant analyses to classify rain forest types in a Landsat TM image over northern Costa Rica. Remote Sensing of Environment 112: 2485-2494.

Toivonen T, Kalliola R, Ruokolainen K, Malik RN (2006) Across-path DN gradient in Landsat TM imagery of Amazonian forests: A challenge for image interpretation and mosaicking. Remote Sensing of the Environment 100: 550–562.

Tuomisto H, Linna A, Kalliola R (1994) Use of digitally processed satellite images in studies of tropical rain-forest vegetation. International Journal of Remote Sensing 15: 1595-1610.

Tuomisto H, Poulsen AD, Ruokolainen K, Moran RC, Quintana C, Celi J, Cañas G (2003a) Linking floristic patterns with soil heterogeneity and satellite imagery in Ecuadorian Amazonia. Ecological Applications 13: 352-371.

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.

Tuomisto H, Ruokolainen K, Yli-Halla M (2003c) Dispersal, environment, and floristic variation of western Amazonian forests. Science 299: 241-244.

Tuomisto H & Ruokolainen K (1994) Distribution of Pteridophyta and Melastomataceae across an edaphic gradient in an Amazonian rain forest. Journal of Vegetation Science 5: 25-34.

Tuomisto H, Ruokolainen K, Kalliola R, Linna A, Danjoy W, Rodriguez Z (1995) Dissecting Amazonian biodiversity. Science 269: 63-66.

Vormisto J, Phillips O, Ruokolainen K, Tuomisto H, Vasquez R (2000) A comparison of fine-scale distribution patterns of four plant groups in an Amazonian rainforest. Ecography 23: 349-359.

Walker JJ, de Beurs KM, Wynne RH, Gao F (2012) Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. Remote sensing of the environment 117: 381-393.

Woodcock CE, Collins JB, Gopal S, Jakabhazy VD, Li X, Macomber S, Ryherd S, Harward VJ, Levitan J, Wu Y, Warbington R (1994) Mapping forest vegetation using Landsat TM imagery and a canopy reflectance model. Remote Sensing of Environment 50: 240-254.

Whittaker RH (1967) Gradient analysis of vegetation. Biol. Rev. 42: 207-264.

Annex 1a: Original mosaic and DNs of sampling sites; histogram equalization

Pebas segments are indicated in green and Nauta in blue on the map as well as in the graph (bands 4, 5 and 7 from up to down).
The three highest DNs to the left correspond to the Pastaza Fan sites.
The sudden change in DN to the left corresponds to the western Pebas-Nauta limit.
There isn´t any visible sudden change in DNs for the Pebas-Nauta limit at the centre of the image
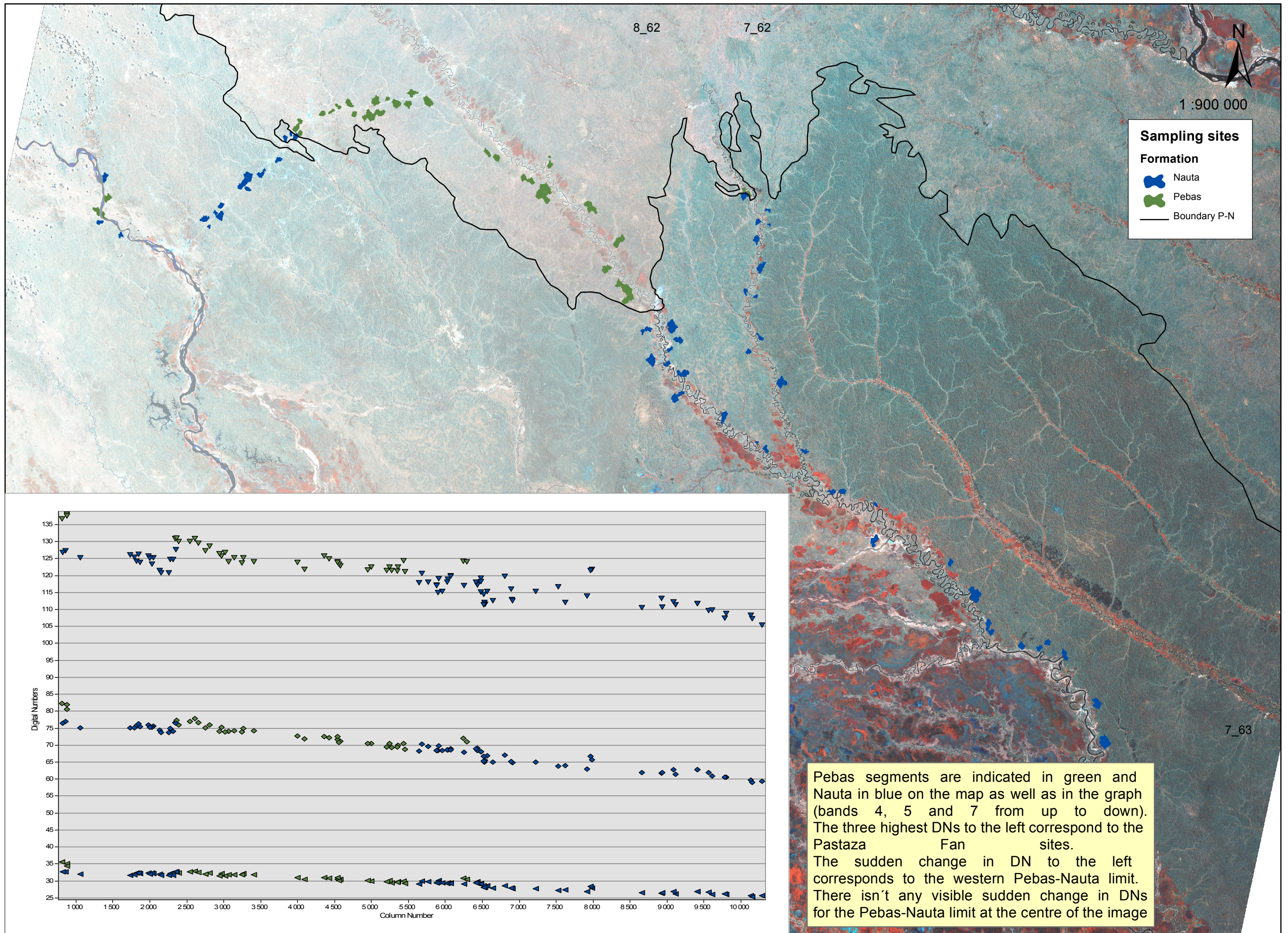
Annex 1b: Detrended mosaic and DNs of sampling sites; histogram equalization

Sampling sites

Formation

- Nauta
- Pebas
- Boundary P-N

1 : 900 000

8_62   7_62

7_63

Visual contrast between Pebas and Nauta formations has increased after image detrending. The graph shows how the range of DN values has been compressed in the three bands, making all the forest types to have more similar DNs.

Detrended image

1.

Histogram match of 8_62 to 7_62-63

Mosaic.img

Segmentation

Segm.sph

Spatial join of transects to segments (one to many and uncheck "keep all target features" box)

Segm_sp.shp

Feature to Raster

Segm_sp.shp → Segm_rast.img

Sample pixel values

Sample.txt

Display XY

Sample_events

Sample_join.shp

2.

Display XY data of the GPS measurements

XYevents

Interpolate transects using compass direction and *bearing line*

Transects

Join based on spatial relation Sample_events to Segm_sp.shp

Intersect to calculate fraction of each transect on each segment

Weigh avge DNs with fraction of transect crossing each segment

Dissolve and sum up weighted average pixel values

3.

7_62
7_63
8_62

Set null values = 0

Img_null

Raster domain to get the boundaries of each img, and select the left boundary (column 1)

Img_domain

From Feature to Point

Segm_centre

DNs_dissolve

Near to calculate nearest distance from each point to left boundary

Segm_near.shp

Divide distance by resolution

CN

Join NMDS score

CN_NMDS

Join

CN_NMDS_DN.shp

Annex 3: Workflow of image detrending