

11-26-2002

Land-cover detection and landscape structure analysis in the Pachitea Basin, Peruvian Amazon

Daniel Gann

Florida International University

DOI: 10.25148/etd.FI15082600

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>



Part of the [Environmental Sciences Commons](#)

Recommended Citation

Gann, Daniel, "Land-cover detection and landscape structure analysis in the Pachitea Basin, Peruvian Amazon" (2002). *FIU Electronic Theses and Dissertations*. 2009.

<https://digitalcommons.fiu.edu/etd/2009>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

LAND-COVER DETECTION AND LANDSCAPE STRUCTURE ANALYSIS IN
THE PACHITEA BASIN, PERUVIAN AMAZON

A thesis submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

ENVIRONMENTAL STUDIES

by

Daniel Gann

2003

to: Dean Arthur W. Herriott
College of Arts and Sciences

This thesis, written by Daniel Gann, and entitled Land-Cover Detection and Landscape Structure Analysis in the Pachitea Basin, Peruvian Amazon, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

David Bray

Dean Whitman

Michael E. McClain, Major Professor

Date of Defense: November 26, 2002

The thesis of Daniel Gann is approved.

Dean Arthur W. Herriott
College of Arts and Sciences

Dean Douglas Wartzok
University Graduate School

Florida International University, 2003

© Copyright 2003 by Daniel Gann

All rights reserved

DEDICATION

I dedicate this thesis to Marisol

ACKNOWLEDGEMENT

I wish to thank the members of my committee for their support, Dr Dean Whitman for his guidance in the aspect of remote sensing, and spatial analysis, and Dr David Bray for keeping me focused on the importance of the anthropogenic dimension of the subject. I thank specifically Dr. Michael McClain my Major Professor for trusting in my ability to successfully accomplish the goals of this research and giving me the opportunity to freely experiment and therefore gain valuable experience in the field of remote sensing. He also financially supported the field work of this research and provided the necessary equipment generously.

I would also like to thank Dr. Tom Philippi for guiding me in the statistical aspect of the thesis, not only pointing to the right direction in theory, but also providing valuable ideas and instruction in the use of statistical software. I further thank Jennifer Fu the director of the Geographic Information Systems and Remote Sensing Center at FIU who was always looking out for ways to support this research in terms of GIS and image processing software as well as a powerful computing environment. I also thank Paola Agudelo, Roxanna Ayllon, and Carlos Garnica, for their assistance during the field work in Peru. I acknowledge the Instituto del Bien Comun and specifically Dr. Dick Smith, for providing valuable information about the native communities of the region and the provision of GIS data including the boundaries of native communities of the Pachitea Basin.

ABSTRACT OF THE THESIS

LAND-COVER DETECTION AND LANDSCAPE STRUCTURE ANALYSIS IN THE PACHITEA BASIN, PERUVIAN AMAZON

by

Daniel Gann

Florida International University, 2003

Miami, Florida

Professor Michael E. McClain, Major Professor

Classification procedures, including atmospheric correction satellite images as well as classification performance utilizing calibration and validation at different levels, have been investigated in the context of a coarse land-cover classification scheme for the Pachitea Basin. Two different correction methods were tested against no correction in terms of reflectance correction towards a common response for pseudo-invariant features (PIF). The accuracy of classifications derived from each of the three methods was then assessed in a discriminant analysis using crossvalidation at pixel, polygon, region, and image levels. Results indicate that only regression adjusted images using PIFs show no significant difference between images in any of the bands. A comparison of classifications at different levels suggests though that at pixel, polygon, and region levels the

accuracy of the classifications do not significantly differ between corrected and uncorrected images.

Spatial patterns of land-cover were analyzed in terms of colonization history, infrastructure, suitability of the land, and landownership. The actual use of the land is driven mainly by the ability to access the land and markets as is obvious in the distribution of land cover as a function of distance to rivers and roads. When considering all rivers and roads a threshold distance at which disproportional agro-pastoral land cover switches from over represented to under represented is at about 1km. Best land use suggestions seem not to affect the choice of land use. Differences in abundance of land cover between watersheds are more prevailing than differences between colonist and indigenous groups.

TABLE OF CONTENTS

CHAPTER	PAGE
INTRODUCTION	1
Description of the Pachitea Basin	1
Alteration of Ecosystems and their Consequences	4
CHAPTER 1	
LAND - COVER DETECTION FROM SATELLITE IMAGERY	7
Classification Scheme, Scale and Choice of Imagery	8
Geometric and Atmospheric Correction	10
Theoretical Background of Atmospheric Correction for Data from Remote Sensors	11
Consideration of Correction Methods	18
Data preparation	21
Radiometric Correction	22
Test for Reflectance Differences	25
Supervised Classification and Accuracy Assessment	30
Conclusion	52
CHAPTER 2	
ASSOCIATIONS BETWEEN LANDSCAPE STRUCTURE AND NATURAL AND ANTHROPOGENIC FACTORS	55
Landscape Composition and Configuration	57
Theoretic Concept of Land-Cover Variability	58
The Selva Central	60
Variability in the Peripheral and Expansion Areas of the Pachitea	65
Bio-Geophysical Factors	65
Ethnicity – Native Communities versus Colonists in the Pachitea	69
Land and Market Accessibility	75
Summary of Research Questions	79
Land-Cover Abundance	79
Land Cover as a Function of Distance from Rivers and Roads	79
Data and Methodology	80
Dependent Variables	80
Independent Variables	80
Watershed Boundaries	80
Indigenous Communities	81
Bio-Geophysical entities	81
Infrastructure	82
Results and Discussion	85
Land-Cover Abundance within the Study Area	86
Differences in Overall Distribution of Land Cover between Watersheds	88

Differences in Overall Distribution of Land Cover between Native Communities and Colonists	90
Distribution of Land Suitable for Agriculture, Pasture, Forestry and Protection by Watershed and Management Type	93
Pattern of Actual Land Cover versus Recommended Best Land Use Practices by Watershed and Management Type	94
Land-Cover Distribution in Relation to Rivers and Roads	98
Distribution of Conflicting and Potential Land Uses as a Function of Distance to Rivers and Roads	104
Conclusion	105
REFERENCES	107
APPENDICES	112

LIST OF TABLES

CHAPTER	PAGE
Table 1. Regression coefficients for bands 1–5 and 7.	25
Table 2. Results of matched-pair t-tests.	28
Table 3. Contingency table for uncorrected images crossvalidated at the pixel level.	48
Table 4. Accuracy for all classes at four different levels of crossvalidation for different correction methods.	49
Table 5. Kappas for different correction methods.	50
Table 6. p-values of the Chi-Square for comparison of pairs of correction methods at all levels.	51
Table 7. Land cover by best land use category indicating conflicts and underutilization.	69
Table 8. Parameters for distance raster.	84
Table 9. Area covered by the major watersheds.	86
Table 10. Land cover distribution within the study area.	87
Table 11. Land cover by watershed within the study area.	89
Table 12. Land-cover distribution by management.	91
Table 13. Land-cover distribution by watershed and management.	92
Table 14. Best land use (BLU) suggestions by watershed.	93
Table 15. Best land use suggestions by management.	94
Table 16. Actual land cover by best land use category.	95
Table 17. Distribution of land-use conflicts and potential.	95
Table 18. Distribution of land-use conflicts and potential by watershed.	97
Table 19. Distribution of land-use conflicts and potential by management.	97

LIST OF FIGURES

CHAPTER	PAGE
Figure 1. Location of the study region.	2
Figure 2. Watershed of the Pachitea River and major mountain ranges.	2
Figure 3. Major sub-basins of the Pachitea watershed.	3
Figure 4. Districts that coincide with the Pachitea watershed.	3
Figure 5. Path descriptions for radiance at-sensor.	12
Figure 6. Study area and the four major watersheds.	24
Figure 7. Scatterplots of reflectance values by band and corresponding regression lines.	26
Figure 8 Mosaics of the three differently corrected images.	29
Figure 9. Crossvalidation concept for different levels of calibration.	33
Figure 10. Ellipsoids, means and standard error graphs of COST corrected image for all band combinations of bands 1,3,4 and 7.	37
Figure 11. Ellipsoids, means and standard error graphs of COST corrected image for all band combinations of bands 1,3,4 and 7.	40
Figure 12. Ellipsoids, means and standard error graphs of COST corrected image for all band combinations of bands 1,3,4 and 7.	43
Figure 13. Spectral signatures for COST corrected images.	46
Figure 14. Spectral signature for regression corrected images.	46
Figure 15. Spectral signatures for uncorrected images.	47
Figure 16. Classified map of the study area.	54
Figure 17: Five life zones according to the classification system by Holdridge.	56
Figure 18: Best land use categories based on soil, slope and precipitation.	67
Figure 19: Matrix of current land cover and best land use practices.	68

Figure 20: The legal boundaries of Native Communities.	73
Figure 21: Colonization progression and infrastructure expansion.	78
Figure 22. Distance raster for different combinations of infrastructure features.	85
Figure 23. Land-cover distribution for the study area.	87
Figure 24. Land cover distribution by watershed.	90
Figure 25. Land-cover distribution of native communities versus colonist managed land.	91
Figure 26. Comparison of land cover distribution between native communities and colonists by watershed.	92
Figure 27. Expected proportions for random distribution of land-cover types as a function of distance to infrastructure features.	99
Figure 28. Expected proportions for random distribution and actual proportions of all land-cover types as a function of distance to all rivers and roads.	100
Figure 29. Cumulative percentage for land-cover distributions as a function of distance to all rivers and roads.	101
Figure 30. Deviation from expected random distribution of land-cover categories as a function of distance to rivers and roads in meters.	102
Figure 31. Deviation from expected random distribution of land-cover proportions and cumulative percent of land cover as a function of distance to different combinations of infrastructure features.	103
Figure 32. Expected proportions for random distribution and actual proportions of (1) inappropriate use of best land use categories forestry and protection (Red) and (2) potential versus underutilized land (Blue).	105

INTRODUCTION

The primary purpose of this research is to assess the current (1999) state of land cover in the Pachitea basin of Peru and to explicitly describe and quantify spatial patterns of the land cover in association with natural and anthropogenic factors.

This research consists of a two-step analysis. The first step includes mapping of land cover from remotely sensed data and selection of appropriate procedures for future monitoring by evaluating different methods at multiple stages of the process. The second step focuses on the quantitative description of landscape patterns in relation to cultural and historic parameters as well as infrastructure features.

Description of the Pachitea Basin

The Pachitea river is located between the Longitudes 74°07'W and 75°01'W and the Latitudes 8°33'S and 10°48'S and about 300 km northeast of Lima (Fig.1). The catchment occupies an area of approximately 29,027 km². The western border lies along the Sierra Central, the main range of the Andean Mountains, and the Cordillera del Sira describes the eastern boundary of the basin. The maximum altitude of the watershed is the Nevado Tarata at about 5,720 meters above sea level (masl), and the lowest elevation is below 170 masl, at the confluence of the Pachitea and the Ucayali Rivers. Due to this steep

elevation gradient the basin encompasses eleven ecological life zones according to the Holdridge classification system. It ranges from very humid tropical forest to humid alpine steppe known as *puna* or *páramo*. Glaciers cover several mountain peaks. Within the Pachitea basin the Cordillera de Yanachaga with peaks of just about 4000 masl divides the sub-basins of the Pozuzo Valley and the Palcazu Valley. The Cordillera de San Matías is the dividing range between the Palcazu Valley and the Pichis Valleys (Fig.2).

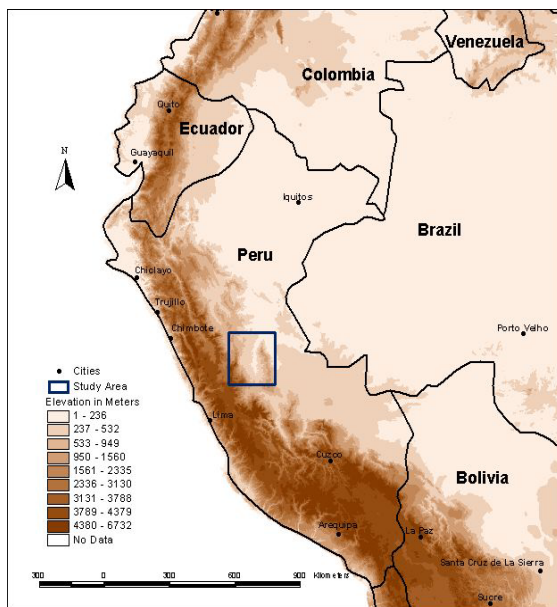


Figure 1. Location of the study region showing national boundaries, major cities and Elevation.

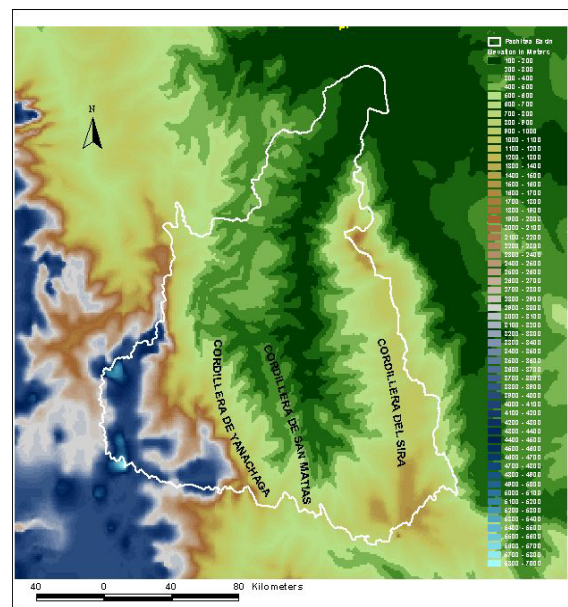


Figure 2. Elevation map showing the delineated basin of the Pachitea River as well as major mountain ranges. Data Source: GTOPO30 1km GRID.

The predominant flow direction of the river system is from South to North with a secondary trend to the East. The Pozuzo in the southwest joins the Palcazu, and about 40 km further east they merge with the Pichis to form the

Pachitea. Hence the basin can be subdivided into four sub-catchments, the Pozuzo in the west, the Palcazu in the center, the Pichis in the East, and the Pachitea in the North (Fig.3). The whole watershed is named after the Pachitea, the river that finally unites with the Ucayali, one of the major Peruvian rivers contributing water to the Amazon, the largest river system in the world.

The political jurisdiction boundaries subdivide Peru into Departments, Provinces and Districts. The Pachitea basin covers the major parts of 14 districts (Fig.4). Seven of these districts (Pozuzo, Huancabamba, Chontabamba, Oxapampa, Palcazu, Villa Rica and Puerto Bermudez) are in the Province of Oxapampa, two districts (Chaglla and Pano) and five districts (Codo del Pozuzo, Yuyapichis, Puerto Inca, Tournavista and Honoria) are within the Provinces of Pachitea and Puerto Inca, respectively.

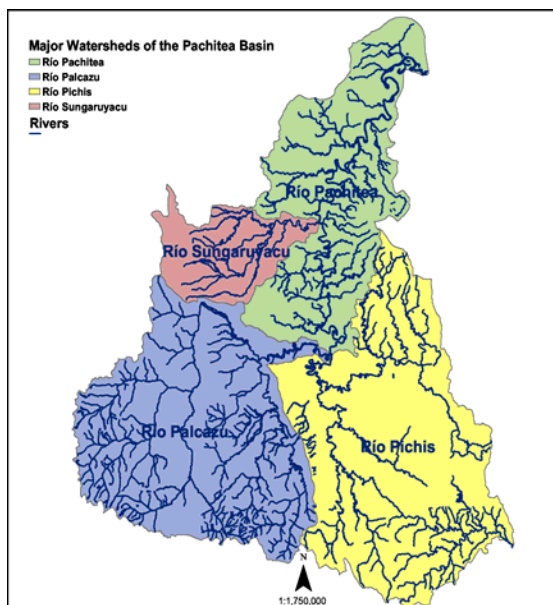


Figure 3. Major sub-basins of the Pachitea watershed.



Figure 4. Districts that overlap the study area.

The different types of forests that cover approximately 775,650 Km² of Peru can be roughly described as *Selva Baja* at elevations below 800 masl, *Selva Alta* between 800 masl and 3800 masl, Transition forest from 800 masl to 1600 masl, Cloud Forest from 1600 masl to 2000 / 3000 masl and *Achaparrados*, *puna* or *páramo* above 2500 / 3000 masl. The deforestation rate for Peru is not well known but was estimated to be 350 000 ha/year in the late 80's (Ocaña 1990). Deforestation patterns throughout the Selva Central and specifically the Pachitea Basin have been neither temporally nor spatially homogeneous. Studies conducted in the past to explain deforestation rates implicated land tenure systems, and economic activities as the main controls (Bedoya 1996).

Alteration of Ecosystems and their Consequences

International concern and efforts in the field of land-cover and land-use change are reflected in the research programs of the International Geosphere-Biosphere Project (IGBP) and the International Human Dimensions Programme on Global Environmental Change (IHDP) (Lambin et al. 1999).

Ecosystem alterations can be recognized at local, regional and global scales and are taking place at a threatening speed. The deforestation of primary forest in the tropics is of special concern because of fragmentation of ecosystems, the high losses of biodiversity, erosion and losses of fertile land, degradation of water quality and associated health problems and changes in local and global climate.

Once human colonization takes place the natural landscape is fragmented and soon it can be described as a patchwork of agricultural fields, pastures, and forests used for diverse purposes and urbanized areas. The land-cover composition of a landscape or region is a function of natural circumstances and cultural land uses. It is obvious that there are innumerable possibilities for the development of different land-use compositions and changes, and yet not all options are equally likely to occur, nor are they equally beneficial for the integrity of ecosystems (Pearson 1994). Therefore it is crucial to understand the dynamics of land-cover changes, the forces that are driving them as well as the impact they have on the ecosystems and their various agents.

Modeling efforts that aim to determine optimal arrangements of landscapes for a multitude of different agents in the natural as well as cultural domain depend on the analysis of spatially explicit data containing information on structure, function and change of ecosystems. Structure describes the spatial patterns or relationships between landscape elements such as animal and plant species, nutrients or energy in relation to the composition and configuration of the landscape. Function explains the flows of objects between landscape elements, and change describes the alterations in structure and function of the mosaic through time (Forman 1995).

Understanding the patterns of energy flow, material transport and movement of species including humans is based on the structural arrangement of the landscape. Landscape patterns that can be expressed in terms of composition and configuration can serve as indicators for ecological processes

and possible stress factors. The interrelation between spatial patterns of land cover and ecosystem processes is therefore a growing subject in natural resource management and associated scientific disciplines such as hydrology (Krysanova 1998, Su 2000), ecology (Forman 1995, Hughes 2000), and island biogeography (MacArthur and Wilson 1961).

Since the structural arrangement of the landscape determines the function and quality of ecosystems and their interactions with each other, it is essential to be able to also quantitatively describe and monitor landscape mosaic composition and configuration patterns over large spatial extents and to determine the sources and parameters that can be associated with them.

Questions that need to be addressed in this context are: Which methods can be used to retrieve reliable information on land cover? How can the spatial patterns of land cover and land use be described in a quantitative and a spatially explicit fashion? Are there natural or anthropogenic variables that can serve as indicators for certain landscape patterns?

The answer to these questions is crucial for the study of landscape dynamics and especially the sustainable management of natural resources.

CHAPTER 1

LAND – COVER DETECTION FROM SATELLITE IMAGERY

Spatial and temporal modeling and monitoring of land cover and land use and their changes over time at a large spatial extent and at high temporal resolution depend to a great amount on the availability of appropriate data sets. The availability of remotely sensed data from satellites allows for such extensive mapping of large areas at multiple points in time, especially for regions that are very remote and not easily accessible. If mapping of land cover or land use is an ongoing process, as in the case of monitoring changes over time, it is useful to test different classification approaches at various stages of the procedures involved in order to determine how different approaches impact mapping accuracy. The results can help to take future decisions on a procedure considering not only the accuracy of the product but also the necessity of steps that are required and possible alternatives that would yield similar accuracies but are less expensive not only in monetary but also in terms of time.

Remotely sensed satellite imagery has been utilized successfully in mapping land-cover in the Amazon (Brondizio et al. 1994, Novo 1997, Saatchi et al. 2000). Basic steps of a multivariate supervised classification procedure of multi-spectral imagery include the defining of a classification scheme, choice of image type, geometric and atmospheric correction of the imagery, collection of reference data for the calibration and validation process, an analysis to discriminate the various classes, classification of the entire image using a

statistical classifier and finally an accuracy assessment to determine the quality of the final map.

In the context of this study one goal is to evaluate different approaches of atmospheric corrections and the use of different reference data sets during the calibration process, and their impact on classification accuracy. Especially for remote areas that are inaccessible it is crucial to find a calibration and validation process that makes wise use of scarce reference data.

Classification Scheme, Scale and Choice of Imagery

The classification scheme and the mapping scale depend to a large degree on the purpose of the study and the desired mapping detail. The classification scheme for this project is a close adaptation of the one used by the local government agencies, responsible for monitoring ongoing land-use changes in the region of the Pachitea. The classification scheme is rather coarse, distinguishing only closed canopy forest, succession forest, purma, grassland, agriculture, bare surface, and water. The classes are not determined through measurements but rather through subjective observations.

Closed canopy forest is described as a forest with a minimum age of about 20 years that has a closed canopy. Succession forest is defined as forest that is at an advanced stage of secondary succession or forest that has been thinned by extraction of individual trees and therefore opening the canopy

whereas purma is a secondary re-growth of forest at an early stage of succession where the majority of plants are not older than 10 years.

Grassland includes natural grassland and man-made pasture as well as otherwise used grassy areas such as soccer fields and open areas in towns.

Agriculture collapses perennial crops such as platanos, yucca and annual crops such as corn, soybean into one category. Bare exposed surface can be found primarily on roads and riverbanks whereas rivers and other water bodies such as lakes make up the category water. Cloud shadows as well as sporadic shadow pixels in areas that are close to elevated features needed to be classified in a separate class, so that none of the desired classes would be polluted with shadow pixels, water and primary forest being the closest classes in signature resemblance.

Extent and grain size define the scale of mapping, where the grain size corresponds to the minimum mapping unit. The largest extent of this study covers the Pachitea Basin, an area of 29,027 km². The purpose of the study is to detect areas where anthropogenic activities alter the landscape. These activities include agriculture and cattle ranching. It is expected that the smallest features to be detected are agricultural plots. Surveys of the region suggest that the smallest agricultural plots are in the order of 0.5 hectares (Cossio-Solano 2001). Therefore the minimum mapping unit that would allow for detection of these small plots was determined to be 0.4 hectares.

The choice of sensor depends on the classification scheme and the anticipated spatial and temporal variability to be detected, modeled or monitored. The spatial and temporal scale determines which sensor delivers images at the appropriate spatial resolution and temporal frequency and the classification scheme determines the spectral resolution necessary to discriminate the target classes. For the given classification scheme and the scale of the study, Landsat 7 Enhanced Thematic Mapper (ETM) imagery with a spatial resolution of 28.5 meters, a temporal resolution of 14 days and a spectral resolution of 6 spectral bands in the visible, near infrared (nIR) and mid infrared (mIR) spectrum of the light is an appropriate data source.

Four ETM images are necessary to cover the entire watershed of the Pachitea (rows 66 and 67 of paths 6 and 7) The minimum mapping unit of 0.4 hectares required for this study translates to the area covered by five ETM imagery pixels, which should be sufficient to eliminate single pixel noise from the classification.

Geometric and Atmospheric Correction

There are several challenges that need to be met when dealing with classification across multiple images using a single classification algorithm. Issues that need to be addressed are geographic referencing and rectification of the images, which would deliver a seamless dataset. The classification of mosaics composed of several images taken at multiple points of time pose a

potential challenge due to atmospheric variability at the times the images are acquired. In order to get consistent and reliable classification results across space as well as making them comparable through time it is necessary to address the issue of atmospheric correction of spectral reflectance values.

Questions that need to be answered are: Is atmospheric correction necessary and do different correction methods yield different classification accuracies? Which correction method, if necessary, is the most appropriate one when considering expenditure, time and accuracy? A theoretic background of atmospheric effects on satellite data will lay the foundation to discuss different approaches of atmospheric correction.

Theoretical Background of Atmospheric Correction for Data from Remote Sensors

Remote sensors mounted on a satellite orbiting in space record radiance values (L_s), picking up radiation that is reflected from earth's surface or the target radiance (L_T) as well as path radiance (L_p) that is caused by scattering effects in the path of the sun's radiation through earth's atmosphere (Fig.5). Before reaching earth's surface, electromagnetic radiation emitted by the sun passes through earth's atmosphere, which has an impact on the intensity of the radiation reaching the target on the ground. Part of the radiation is scattered other parts are absorbed (Fig.5). The degree to which this scattering and absorption takes place depends on wavelength and the conditions of the

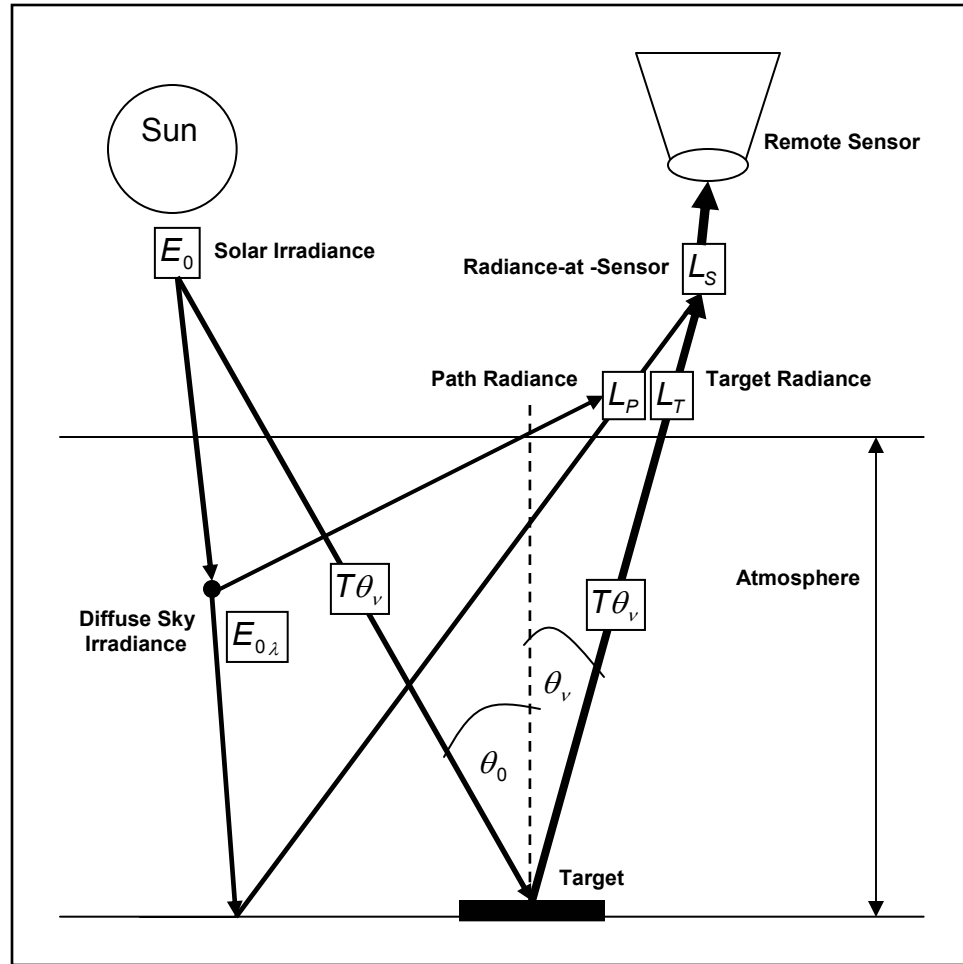


Figure 5. Path descriptions for radiance at-sensor.

atmosphere (Jensen 1996). After the radiation is reflected by earth's surface it has to traverse the atmosphere a second time to reach the sensor (Fig.5). Radiance of a target recorded by a remote sensor therefore is not equal to the radiance measured on the ground. The radiance recorded at the sensor can vary significantly even for ground features such as rock outcroppings and deep water bodies, that do vary only very little or not at all through time.

These radiance differences can cause major difficulties when it comes to differentiating target features for different classes based on the spectral reflectance properties of the targets. The radiance-at-sensor can be described by the following equations, which are adopted from (Markham and Barker 1986, Chavez 1996, Jensen 1996), and (NASA).

$$L_S = L_T + L_P \quad (1)$$

where

L_S at-sensor radiance in $Wm^{-2}sr^{-1}\mu m^{-1}$

L_T target radiance in $Wm^{-2}sr^{-1}\mu m^{-1}$

L_P path radiance in $Wm^{-2}sr^{-1}\mu m^{-1}$

W = watts, m = meters, sr = steradians, μm = micrometers

Due to atmospheric attenuation, solar irradiance E_g reaching earth's surface can be expressed as:

$$E_g = \int_{\lambda_1}^{\lambda_2} (E_{0_\lambda} T_{\Theta_z} \cos \Theta_z + E_{d_\lambda}) d\lambda \quad (2)$$

where

λ_1 wavelength of lower limit

λ_2 wavelength of upper limit

E_{0_λ} spectral solar exoatmospheric irradiance corrected for Earth-sun distance in $Wm^{-2}sr^{-1}\mu m^{-1}$

Θ_z solar zenith angle in radians

T_{Θ_z} atmospheric transmittance from sun to ground at solar zenith angle (unitless)

E_{d_λ} diffuse sky irradiance in $Wm^{-2}sr^{-1}\mu m^{-1}$

Exoatmospheric irradiance differs throughout the year according to the distance of the earth to the sun.

$$E_{0_\lambda} = \frac{E_{0(s)_\lambda}}{d} \quad (3)$$

where

$E_{0(s)_\lambda}$ spectral solar exoatmospheric irradiance in $Wm^{-2}sr^{-1}\mu m^{-1}$

d earth-sun distance in astronomic units at time of data acquisition

Atmospheric transmittance is the proportion of radiant energy that reaches the ground in comparison to absent atmosphere.

$$T_{\Theta} = e^{\tau_\lambda / \cos \Theta_z} \quad (4)$$

where

$$\tau_\lambda = \tau_m + \tau_p + \tau_a \quad (5)$$

and

τ optical thickness of the atmosphere

τ_m Rayleigh scattering

τ_p Mie scattering

τ_a atmospheric absorption

with

$$\tau_a = \tau_{H_2O} + \tau_{O_2} + \tau_{O_3} + \tau_{CO_2} \quad (6)$$

Therefore assuming that the earth is a diffuse reflector

$$L_T = \frac{1}{\pi} \int_{\lambda_1}^{\lambda_2} R T_{\Theta_v} (E_{0_\lambda} T_{\Theta_z} \cos \Theta_z + E_{d_\lambda}) d\lambda \quad (7)$$

where

R reflectance of target on the ground (no units)

T_{Θ_v} atmospheric transmittance from ground to satellite at nadir view
angle of the satellite (no units)

The integrated radiance for a wavelength interval then is

$$L_S = \frac{1}{\pi} R T_{\Theta_v} (E_{0_\lambda} T_{\Theta_z} \cos \Theta_z + E_{d_\lambda}) + L_P \quad (8)$$

further Equation (8) solved for Reflectance R can then be described as

$$R = \frac{\pi(L_S - L_P)}{T_{\Theta_v} (E_{0_\lambda} \cos \Theta_z T_{\Theta_z} + E_{d_\lambda})} \quad (9)$$

Satellite imagery is delivered as 8-bit data with a value range of 0-255. These digital numbers (DN) can be transformed to radiance-at-sensor using satellite specific and data post-processing parameters according to equation (10).

$$L_S = L_{\min_{\lambda}} + \left(\frac{L_{\max_{\lambda}} - L_{\min_{\lambda}}}{DN_{\max} - DN_{\min}} \right) (DN - DN_{\min}) \quad (10)$$

where

L_S spectral radiance at-sensor in $Wm^{-2}sr^{-1}\mu m^{-1}$

$L_{\max_{\lambda}}$ maximum spectral radiance in $Wm^{-2}sr^{-1}\mu m^{-1}$

$L_{\min_{\lambda}}$ minimum spectral radiance $Wm^{-2}sr^{-1}\mu m^{-1}$

DN_{\max} maximum digital number

DN_{\min} minimum digital number

DN digital number

The unknown variables in equation (9) that vary with atmospheric conditions are L_P , T_{Θ_v} , T_{Θ_z} and $E_{d_{\lambda}}$. Several correction methods have been suggested to correct at-sensor radiance values to eliminate image differences due to those variables. The methods proposed to radiometrically calibrate or rectify images can be grouped into two major categories; absolute correction and relative correction.

Absolute correction intends to generate reflectance values of the target area under investigation by adjusting radiance values using information on the characteristics of the atmosphere at the time of acquisition. The atmospheric conditions are preferably in-situ measurements or otherwise simulated conditions used with radiative transfer codes (RTC) such as LOWTRAN7 (Kneizys et al. 1988), 5SC (Tanré et al. 1990, Teillet and P. 1991), and HBC (Herman and Browning 1965).

Other models that utilize minimal external image information are models such as Dark-Object-Subtraction (DOS), (Chavez 1988) and improved DOS models such as the COST model, which uses the COSine of the solar zenith angle as an approximation for atmospheric Transmittance (Chavez 1996). Comparisons of performance of these models or combinations of them indicate that RTC approaches yield better results than DOS when compared to ground reflectance measurements (Moran et al. 1992). The COST model approximates path transmittance between the sun, earth and satellite and therefore combined with the DOS model can improve the product (Chavez 1996).

Relative correction on the other hand attempts to adjust the reflectance values of a target images to those of a reference image. This approach would not yield absolute corrected reflectance values but rather approximate the reflectance values to similar conditions as when the reference image was taken. The advantage of relative correction is that only image inherent data are used and no external data are necessary. Relative correction methods that have been

proposed are for example, Pseudo Invariant Feature correction (PIF) (Schott et al. 1988), rectification using dark-bright (DB) control sets extracted from the Kauth-Thomas greenness and brightness scattergrams (Hall et al. 1991), simple regression (Jensen 1996) and no-change regression (Yuan and Elvidge 1996).

Consideration of Correction Methods

Some of the proposed methods are very costly, labor and computation intensive others require specific features and settings within an image and therefore are not feasible for this project.

In the case of absolute correction methods, RTC methods are computationally rather exhaustive and in-situ measurements are prohibitively expensive for the remoteness of the study area. Reliable simulated atmospheric conditions for the Andean Amazon are not well documented and performance with RTC for this region still needs to be demonstrated. Since the DOS model performs rather poorly due to neglecting atmospheric transmittance the refined COST model was chosen as a possible alternative for an absolute correction method.

For the DB relative correction methods is recommended to choose control sets from the extremes of the KT greenness-brightness scattergram that are not in the vegetated areas. This approach is useful if the data set includes dark lakes and bright man-made structures or rock outcroppings in the bright spectrum. For

the study area of this project neither is available. The darkest pixels are those of closed-canopy forest and the brightest values are sandy riverbanks.

The PIF method performs best just as recommended for the DB method using invariant features such as man-made structures that are large enough to extract pure pixels (Schott et al. 1988) and has to be rejected for the same reasons as the DB method.

Regression and no-change regression use either the entire image or part of it in order to derive coefficients for a linear adjustment by band assuming that at least 50 percent of the images include no-change pixels. These methods are generally used to adjust images of the same location from different times using images from the same sensor or even across sensors (Yuan and Elvidge 1996). Since images for this study are adjacent images that share only a relatively small number of pixels this approach is not optimal either. A combination of several described approaches was therefore chosen. Pixels for PIF were selected using the KT scattergram. The brightest pixels turned out to be the elevated areas of sandy and stony riverbanks. These are little susceptible to changes in wetness, which allows for the assumption of linearity in reflectance. The darkest pixels fell all within closed-canopy forest (well within the greenness region) as a result the equations for the DB method could not be applied. Instead these pixels together with the selection of bright pixels were used in a simple regression model.

Since a poor correction is not always better than no correction at all (Moran et al. 1992) the third option, computationally and money wise by far the easiest and simplest, is not to correct for atmospheric conditions.

The three correction methods then were:

1. Radiometric Correction using the COST model (absolute)
2. PIF regression model (relative)
3. no atmospheric correction

The fitness of the correction models for the images of the study area and the proposed classification scheme was tested in two ways. A test for differences in terms of reflectance values for pseudo-invariant features in overlap areas indicates if the corrected images are sufficiently alike so that signatures could be extended across image boundaries. In a second step a linear discriminant analysis using cross validation was performed for each correction method in order to assess and compare classification accuracy for the different correction methods. A comparison of the kappa statistic for the different classification results is expected to indicate if there are significant differences in classification accuracy due to atmospheric correction.

In addition to atmospheric distortion of reflectance values, slope and aspect of the terrain need to be taken into consideration since they contribute to that distortion. Topographic correction methods to neutralize anisotropic effects (Smith et al. 1980, Colby 1991, 1998) depend on a digital elevation model (DEM) that is compatible with the resolution of the remotely sensed data to be corrected.

The DEM generated from digitized 1:100,000 National maps of Peru did not qualify for topographic normalization of the images due to scale incompatibility and moderate accuracy of the maps themselves. Consequently it was necessary to exclude pixels that most likely showed inconsistencies of illumination due to topographic effects. The coarse DEM and its derivative slope were used to identify areas above 500 masl and areas that have a slope greater than 5 degrees. Areas for which no data existed on the national maps (gaps in contour lines) were also excluded.

Data preparation

In order to cover the final study area two Landsat 7 images (Path 6 Row 67 (6/67) from July 29th, 1999; Path 7 Row 67 (7/67) from August 05th, 1999) were acquired. The two images were geometrically corrected and georeferenced to each other using GPS ground control points collected during the first field campaign. The rectification method used was a second order polynomial transformation with a root mean square error (RMSE) of less than half a pixel. The re-sampling method employed was a rigorous transformation using the nearest neighbor. In a first unsupervised classification using the cluster busting method (Jensen 1996) clouds were eliminated from both images. The final study area was delineated applying a 20 km buffer around major rivers and roads (digitized from the images) and eliminating elevations above 500 masl, slopes greater than 5 degrees and areas for which no elevation data was available (Fig.6).

Radiometric Correction

Equations (9) and (10) were combined in a model and the atmospheric condition dependent variables L_P , T_{Θ_v} , T_{Θ_z} and E_{d_λ} were adjusted according to each model.

$$R = \frac{\pi(L_S - L_P)}{T_{\Theta_v} (E_{0_\lambda} \cos \Theta_z T_{\Theta_z} + E_{d_\lambda})} \quad (9)$$

where

$$L_S = L_{\min_\lambda} + \left(\frac{L_{\max_\lambda} - L_{\min_\lambda}}{DN_{\max} - DN_{\min}} \right) (DN - DN_{\min}) \quad (10)$$

In the no-correction or apparent reflectance model radiance at-sensor was converted to reflectance without correction for any atmospheric conditions using the following settings for equation (9).

$$L_P = 0.0$$

$$T_{\Theta_v} = 1.0, \text{ for nadir viewing Landsat 7}$$

$$T_{\Theta_z} = 1.0, \text{ assuming no transmittance effects (absence of atmosphere)}$$

$$E_{d_\lambda} = 0.0$$

In the case of absolute radiometric correction using the COST method radiance at-sensor was converted to reflectance with correction for any atmospheric conditions using the following settings in equation (9)

$$L_P = \text{Value derived using dark-object criteria (Chavez 1988)}$$

$$T_{\Theta_v} = 1.0 \text{ ,for nadir viewing Landsat 7}$$

$$T_{\Theta_\lambda} = \cos \Theta_z \text{ (Chavez 1996)}$$

$$E_{d_\lambda} = 0.0$$

For relative correction of the reflectance values the DN images were converted to reflectance images using a model based on Equation (10). Overlap areas of these uncorrected images were used in a linear regression model (Equation 11). Overlap areas of the images were smoothed with a 5x5 low pass filter to sample 14 points within pseudo invariant features. The reason for this smoothing is the fact that the area covered on the ground by the same pixel differs from scene to scene slightly. In addition the georeferencing and image rectification process was within an RMSE of half a pixel. Results of the regression (Appendix A) are summarized in Table 1 and the scatterplots and regression line are shown in Figure 7.

Linear regression equation

$$Y = aX + b \tag{11}$$

where

Y reflectance value of the reference image (7/67)

X reflectance value of the target image (6/67)

a,b regression coefficients

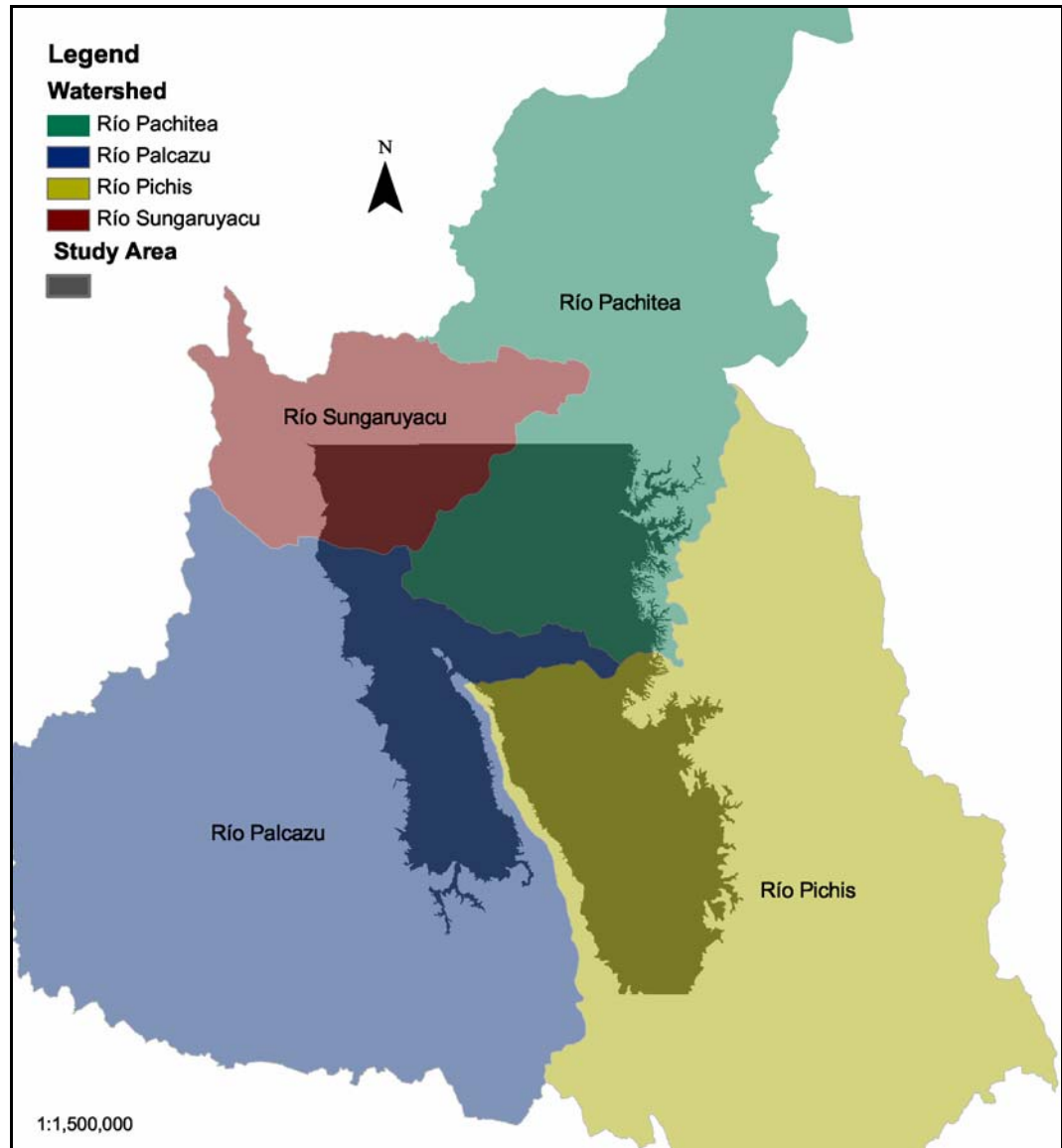


Figure 6. The study area and the four major watersheds show to what extent each watershed is covered. A 20 km proximity to major rivers and roads as well as elevations of less than 500 masl, and slopes of less than 5 degrees limited the study area. Missing reliable data in the elevation contours of the national maps further restricted the study area to the North and the South.

Table 1. Regression coefficients for bands 1–5 and 7 for 5x5 low-pass smoothed overlap areas and the corresponding r^2 .

Regression Coefficients and r^2						
	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
a	0.833	0.871	0.879	0.855	0.998	0.982
b	0.019	0.011	0.008	0.026	-0.018	-0.001
r^2	0.995	0.996	0.996	0.989	0.983	0.989

The regression coefficients were applied to the 6/67 uncorrected reflectance image, which resulted in a relative corrected image.

Test for Reflectance Differences

The conversion from radiance to reflectance without atmospheric correction as well as the COST correction model, each yielded two new images; the regression model generated a one image adjusting the 6/67 to the 7/67 uncorrected reflectance image. A matched-pair t-test for invariant pixels within the area of overlap for each of the three image pairs is expected to indicate how well each method adjusted the radiance to reflectance values of both images and therefore excluded possible problems for across-image signature extension during the classification process.

Overlap areas of all images were smoothed using a 5x5 low-pass filter for the same reasons mentioned above. 40 random samples were collected within the overlap area using pseudo-invariant feature polygons not using samples that were involved in the regression model. 25 of these samples were collected within 4 polygons of remote closed-canopy forest and 15 within 5 polygons on

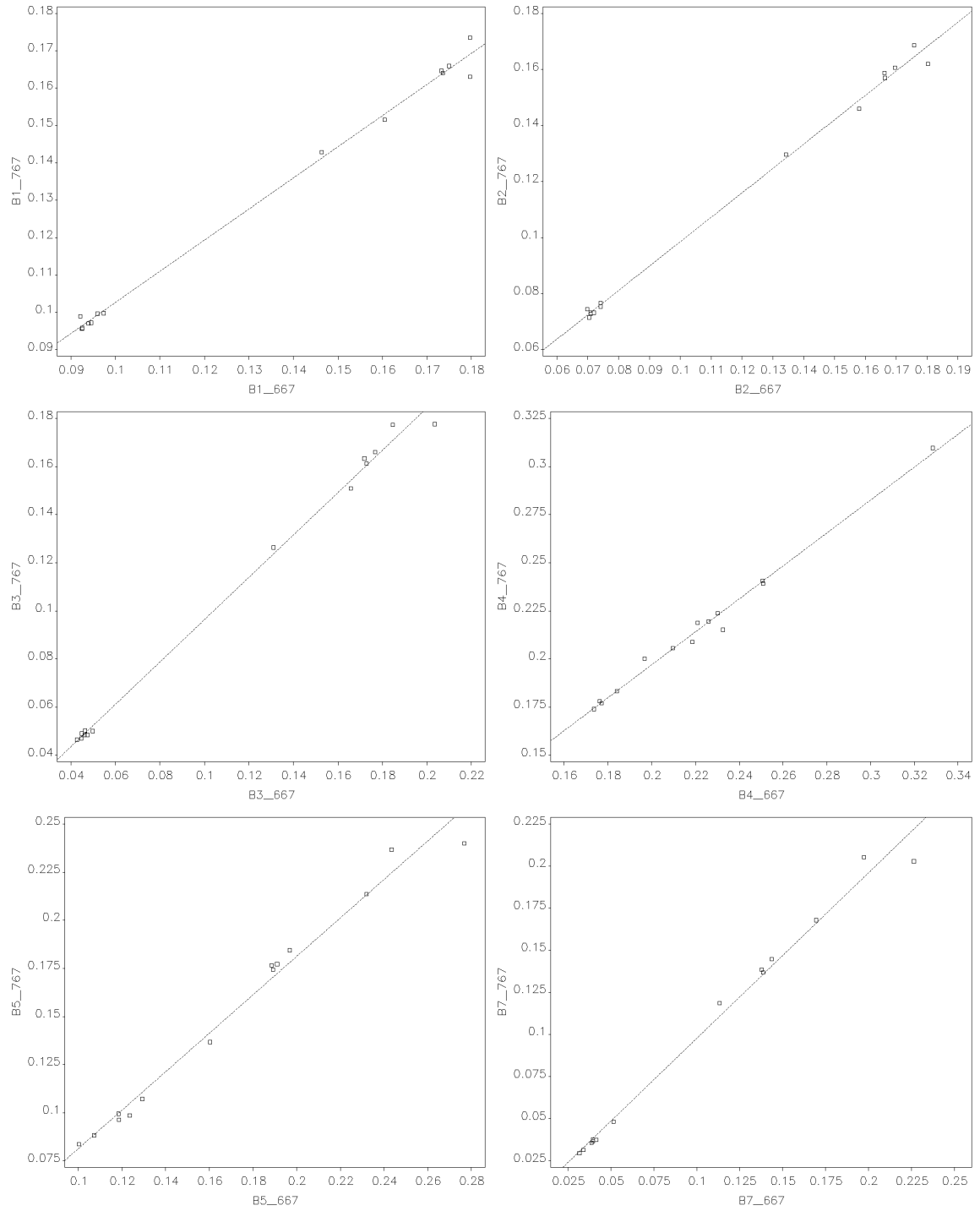


Figure 7. The scatterplots show reflectance values of image 6/67 (667) versus 7/67 (767) by band (B1 – B5, B7) and the corresponding regression lines. The samples were taken from pseudo-invariant features and were used to calculate the regression coefficients (Table 1).

riverbanks. For these 40 random samples reflectance values were extracted from all 5 low-pass filtered images. The hypothesis that was tested for each pair of images (COST 6/67 and 7/67, uncorrected reflectance 6/67 and 7/67, and the regression corrected 6/67 and uncorrected reflectance 6/67) is the following:

H_0 : There is no significant difference between the two images

H_1 : There is a significant difference between the two images

The chosen test statistic was a matched-pair t-test with a degree of freedom of $N-1$ ($40 - 1 = 39$). H_0 is rejected if the p-Value is less than 0.05 ($\alpha = .05$). The results summarized in Table 2 indicate that H_0 needs to be rejected ($\alpha = .05$) for the Cost and uncorrected images. There is not enough evidence to rule out that the uncorrected as well as COST method corrected images are not significantly different for bands 4,5 and 7. There is enough evidence though not to reject H_0 for bands 1,2 and 3 ($\alpha = .05$). It is also evident that H_0 cannot be rejected for any band of the regression corrected images ($\alpha = .05$). The results of the t-tests (see Appendix B for full report) lead to the conclusion that the only correction method that adjusted the reflectance values sufficiently enough to consider both images taken from the same population is the regression method. A visual comparison of the mosaics of the images supports the t-test results (data values of the reference image (7/67) were used for the overlap area). A difference between the pair of uncorrected images (Fig.8a) and also for the set of COST corrected images (Fig.8b) is obvious. Only

in the mosaic of the regression corrected images did the seam between the two images disappear (Fig.8c).

Table 2. Results of the matched-pair t-tests for the 5x5 low-pass filtered overlap areas for the three different correction methods. Regression was performed on 5x5 low-pass filtered images. Reported are the t-values and p-values. The grey fields indicate significance at $\alpha = .05$.

	Results for t-test (sampling 5x5)					
	Uncorrected		COST corrected		Regression	
	t-value	p-value	t-value	p-value	t-value	p-value
Band 1	1.457	0.153	-1.834	0.074	-1.059	0.296
Band 2	-0.210	0.835	-1.870	0.069	1.380	0.175
Band 3	-0.193	0.848	-0.984	0.331	-1.770	0.085
Band 4	-12.101	0.000	-15.079	0.000	1.884	0.067
Band 5	-32.370	0.000	-20.964	0.000	0.895	0.376
Band 7	-7.333	0.000	-7.540	0.000	-1.205	0.236

This means that spatial extension of signatures across images should perform best using the mosaic of the regression corrected images. The use of COST corrected and uncorrected images might lead to unsatisfying results in classification results when including bands 4,5 and 7. In order to further investigate if there are in fact differences in classification accuracy for the different images all three mosaics were used in a calibration and validation procedure.

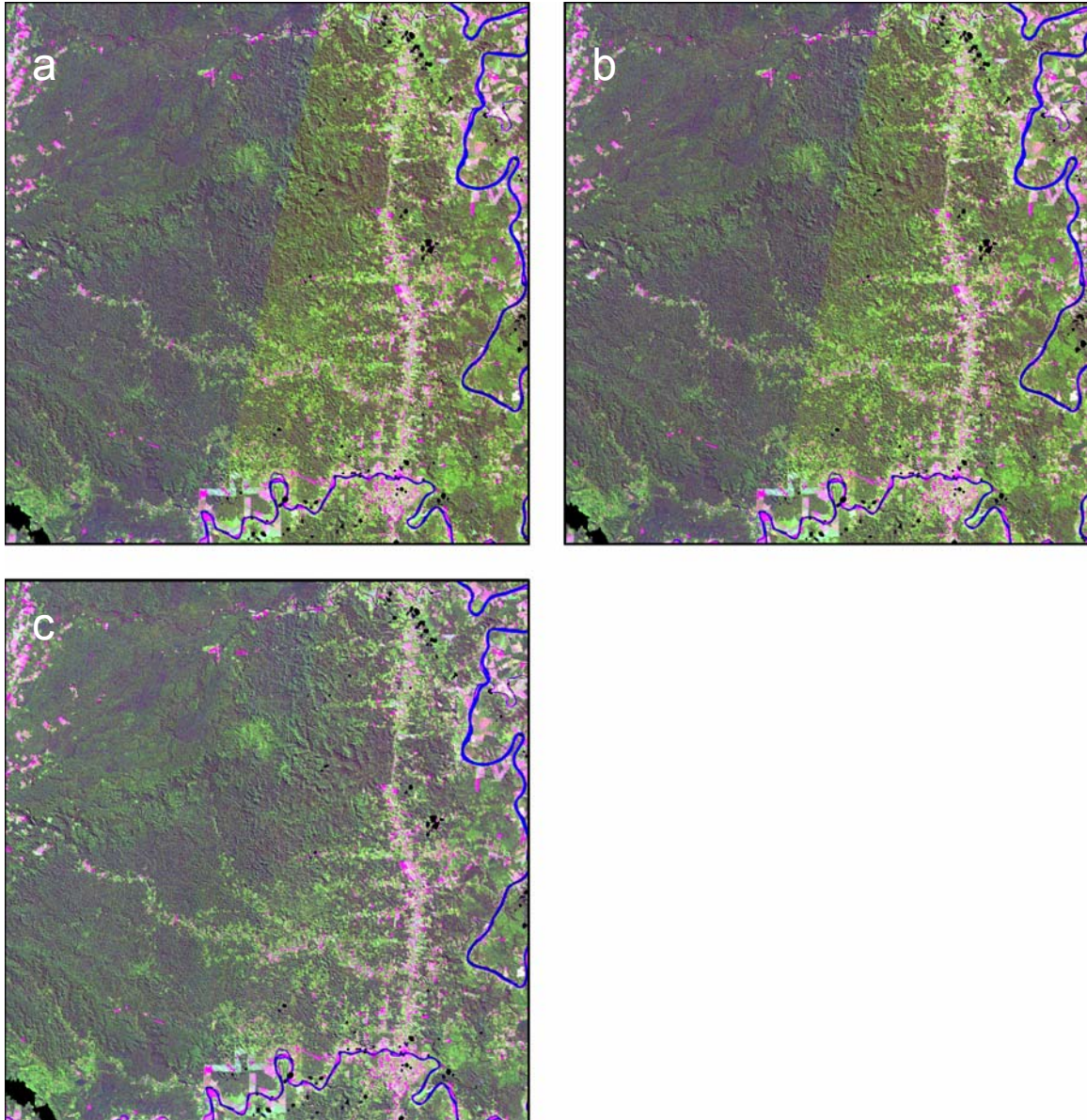


Figure 8 Mosaics of the three differently corrected images using bands 7,4 and 3 (RGB) at a scale of 1:500,000. a: uncorrected reflectance images, b: COST method corrected, c: regression corrected. Only in mosaic c is the seam between the two images not noticeable.

Supervised Classification and Accuracy Assessment

The conventional approach to supervised classification of remotely sensed data requires a calibration or training dataset, a method to discriminate classes, a classifier and a validation dataset for accuracy assessment. A set of reference points for each class is used to generate training statistics including the mean vector and the variance-covariance matrix of the spectral characteristics of each class. These training statistics are then used to evaluate the separability of the classes and to select the bands that best discriminate the classes. After classifying each pixel in the entire image using a classifier, an accuracy assessment is performed using a validation data set.

The most commonly used methods to select reference points for validation are post-classification stratified random sampling by class, systematic sampling or stratified systematic unaligned sampling (Congalton and Green 1957). For all three methods all mapped units have to have an equal and independent chance of being selected. A suggested minimum sampling size for each class can be calculated according to a multinomial distribution (Congalton and Green 1957). Hence this approach is not feasible for regions where the majority of the mapped areas are either inaccessible or only reachable under prohibitively high expenses in terms of time and money. For a big part of the mapped area in the Pachitea basin this is the case, and consequently the requirement for each unit to be selected at equal chance would be violated. Another method needed to be employed to estimate the accuracy, a method that would allow using one data

set for calibration and validation. Such approach would diminish expensive time in the field and provide an honest estimation of the expected classification accuracy. Additionally, one goal of this project was to determine if the radiometric correction of images is necessary and if it matters where within the study area reference data samples are collected. Of special interest was if calibration samples taken in one image could be successfully used in classifying pixels of another image. It is crucial to understand how sampling location impacts the accuracy of a classification in order to develop sampling designs that are efficient and accurate.

It is expected that for any class spectral reflectance values of pixels that are within one contiguous patch are more similar to each other than those from different patches. The same is true for patches that are closer to each other and for pixels are from within an image versus those of another image. Spatial independence of samples within the same patch is not assured and can therefore bias the training statistic. The result could be high classification accuracy within the patch, polygon or image where the sample was taken and poor accuracy for more distant pixels of other patches or images. In order to get a good estimate of classification performance for different calibration datasets without an expensive post-classification accuracy assessment calibration and validation needs to take place at these different sampling levels.

For this purpose a crossvalidation was designed that would test for all the above concerns and requirements. Crossvalidation is a process where one

dataset of known class samples is split into a calibration and validation data set. The calibration process applying a linear discriminant analysis to the calibration dataset yields class characteristic statistics, which then are used to predict the samples in the validation dataset. The splitting of the known samples in calibration and validation sets can take place based on different criteria that are of interest.

The crossvalidation was performed at the pixel, polygon, region and image level (Fig.9) to see if it matters what pixels are used for the calibration process in terms of spatial distance. For this purpose the study area was divided into three regions. Two regions were created in one image (7/67) and one region in a second adjacent image (6/67). Within each of the three regions ten polygons were selected for each of the seven classes, which yielded a total number of 210 polygons. A stratified random sample of 15 Pixels drawn from each of the 210 polygons provided the final data set of 3150 coordinates of pixels.

The reference data for calibration and validation was collected in the field. Polygons for each class were located and identified using GPS receivers. Only Polygons with a core area of at least 30 adjacent satellite pixels were considered in order to avoid mixed pixels in the calibration process. Polygon boundaries were drawn onto resolution-enhanced, brovey-transformed (bands 5,4,3 and 8) satellite images that had been printed at a scale of 1:20,000.

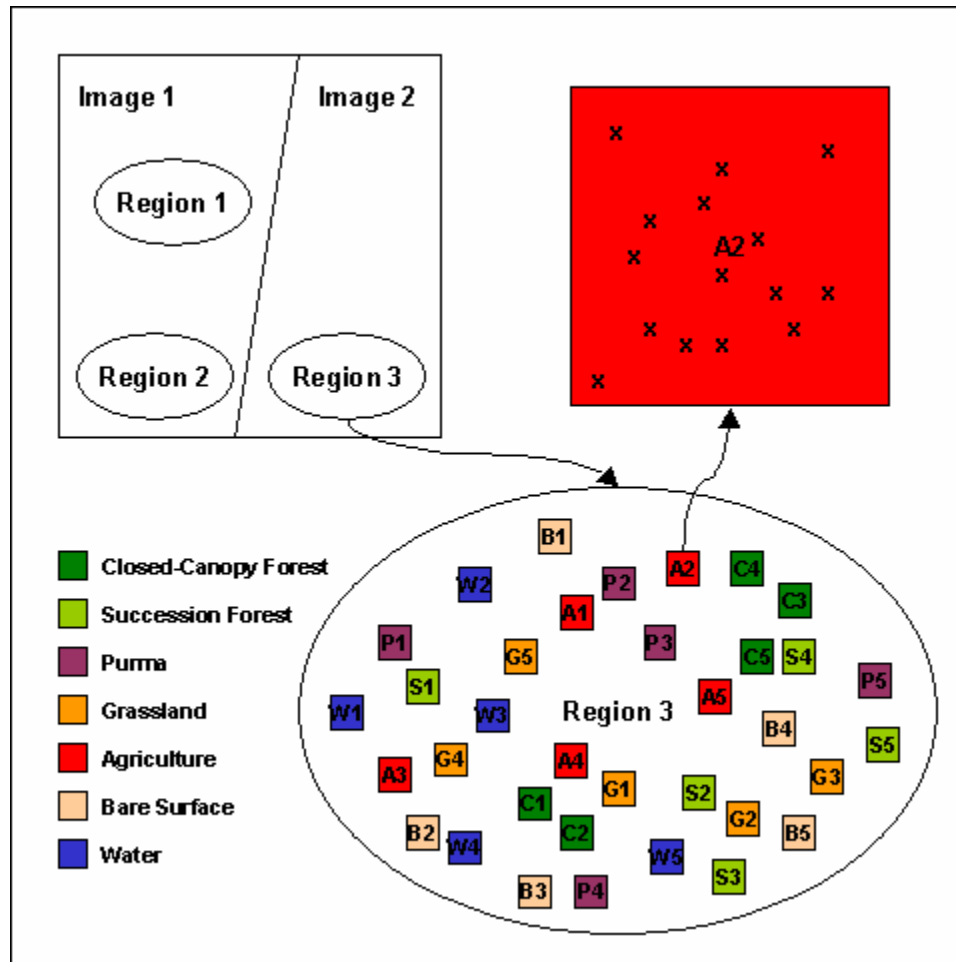


Figure 9. Crossvalidation concept for different levels of calibration. The crossvalidation was performed at four levels, image, region, polygon and pixel.

For each mosaic (uncorrected, regression corrected and COST corrected images) reflectance values of all 6 bands were converted to 8-bit data with values ranging from 0-255. Spectral values were extracted for all 3150 random sampling points and training statistics including mean vector and covariance matrixes were generated (Appendix C). Before the crossvalidation procedure was performed an evaluation of the separability between classes for different combinations of bands helped in selecting those bands that allow for the best

discrimination between classes (feature selection) (Jensen 1996). This process also identifies overlap between classes and consequently potential misclassification. One method used for separability analysis is divergence (Equation 12)(Swain and Davis 1978). Since this equation allows calculations only for two classes at a time, the average of the divergences is calculated (Equation 13) and the subset of bands with the highest average divergence is chosen.

The transformed divergence takes care of outlying easily separable classes and is calculated according to Equation (15)(Jensen 1996). Values of the transformed divergence range from 0 to 2000 where a score of 2000 indicates 'excellent' between-class separability, values above 1900 suggest 'good' and below 1700 'poor' separability (Jensen 1996).

$$D_{cd} = \frac{1}{2} \text{tr}[(V_c - V_d)(V_{d^{-1}} - V_{c^{-1}})] + \frac{1}{2} \text{tr}[(V_{c^{-1}} - V_{d^{-1}})(M_c - M_d)(M_c - M_d)^T] \quad (12)$$

where

D_{cd} Divergence for class combination c and d

$\text{tr}[\cdot]$ trace of a matrix

V_c, V_d covariance matrices for classes c and d

M_c, M_d mean vectors for classes c and d

$$D_{avg} = \frac{\sum_{c=1}^{m-1} \sum_{d=c+1}^m D_{cd}}{C} \quad (13)$$

where

D_{avg} Average divergence

m number of class

with

$$C\left(\frac{n}{q}\right) = \frac{n!}{q!(n-q)} \quad (14)$$

where

C number of combinations of n bands taken q at a time

further

$$TD_{cd} = 2000 \left[1 - \exp\left(\frac{-D_{cd}}{8}\right) \right] \quad (15)$$

where

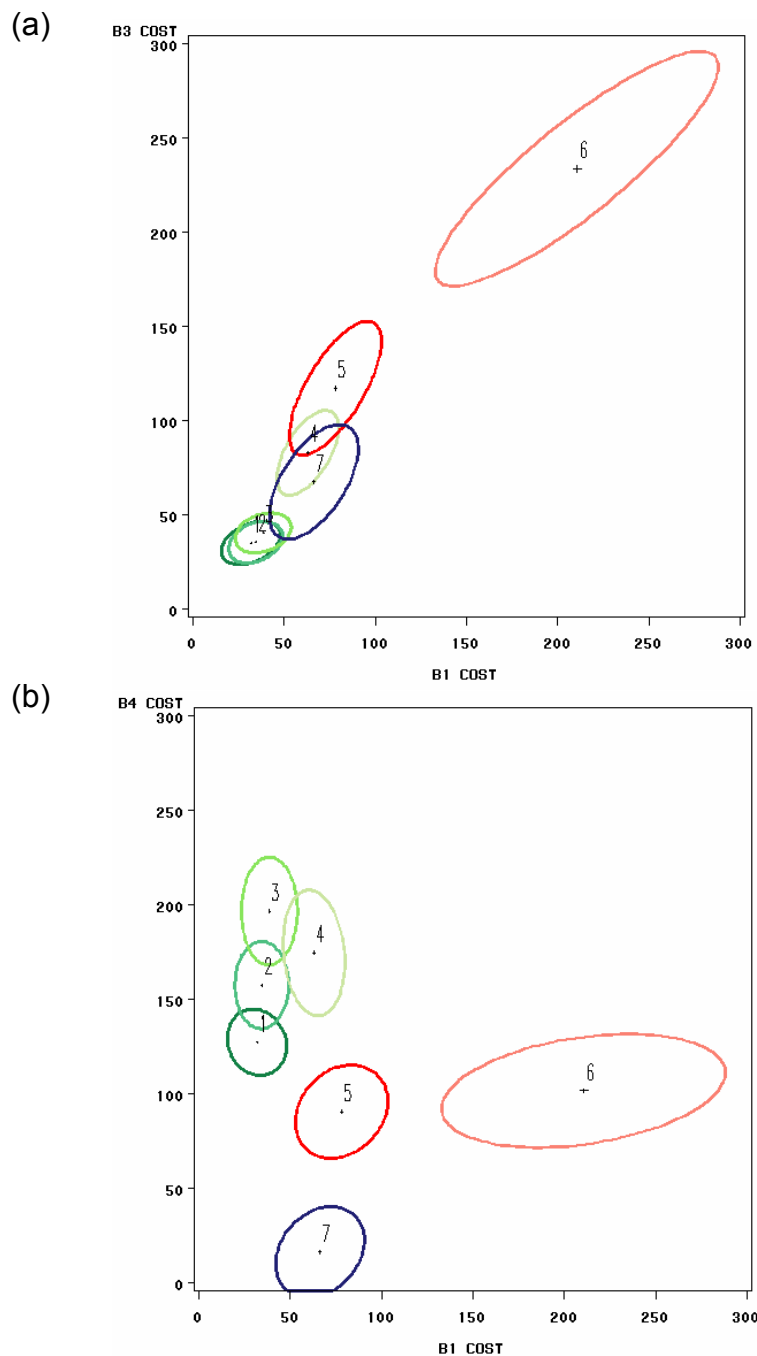
TD_{cd} Transformed Divergence for class combination c and d

Results of the separability analysis using transformed divergence (Appendix D) suggest that bands 1,3,4 and 7 are the most suitable bands to discriminate the classes, and this is true for all three mosaics. Using these 4 bands all except two classes were indeed spectrally separable using the

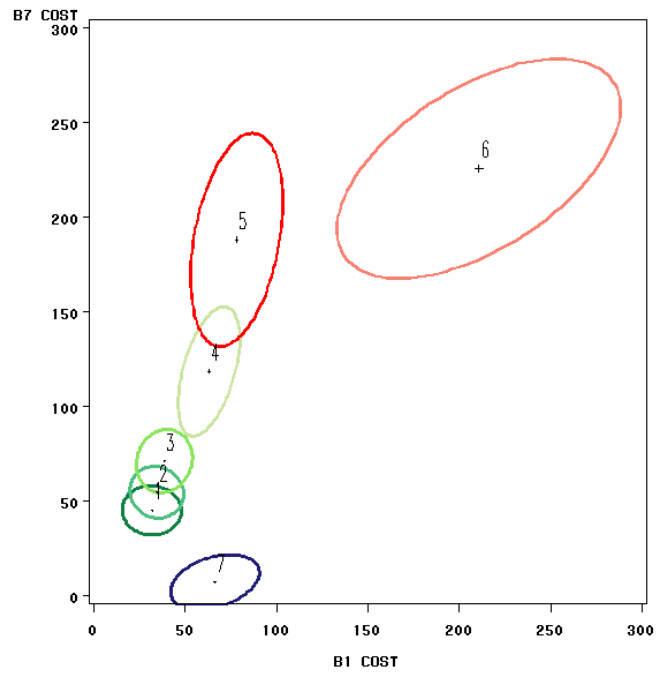
extracted signatures with minimum divergence values of 1873, 1903, and 1902 and average divergence values of 1982, 1984, and 1985 for uncorrected, regression corrected and COST corrected mosaics respectively. The only two classes that showed low separability were closed-canopy forest and succession forest with minimum values of 1635, 1661, and 1691. Acceptable separability is expected for succession forest and purma with values between 1800 and 1900. All other combinations of classes reach excellent separability (2000) with the the combination grassland and agriculture falls short of 2000 by 1 (1999) for regression and COST corrected and 2 (1998) for uncorrected. Graphs of the mean and probability ellipsoids of the training sets for all six band combinations (Fig.10 – Fig.12) and signature plots (Fig.13 – Fig.15) illustrate graphically that result.

For all 3150 samples identification variables for image, region, polygon, pixel number and class were extracted along with the spectral values for all three mosaics. The crossvalidation performed at pixel level uses all pixels except for one in the calibration process (3149). The discriminant procedure applied to the calibration dataset yields class specific statistics, which are then used to predict the validation dataset of one pixel. The class assignment of the pixel is recorded and the procedure is repeated until each pixel is predicted based on the calibration of all other pixels.

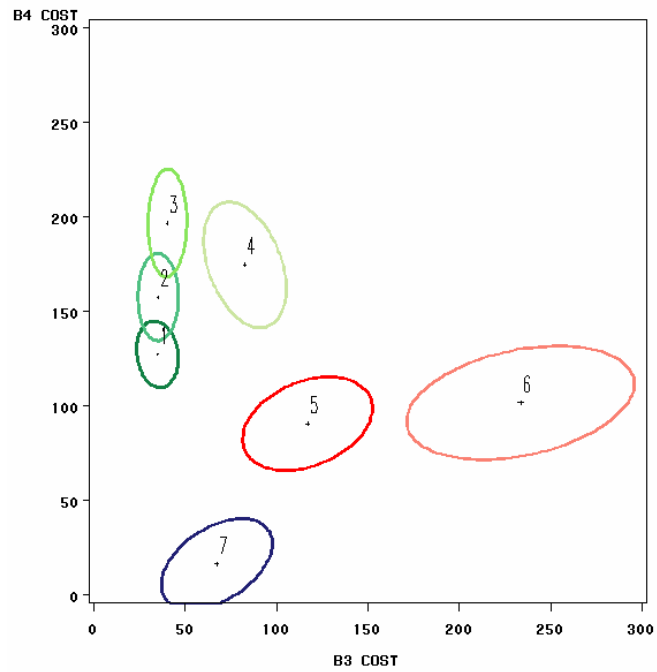
Figure 10. Graphs a – f: The graphs show mean, standard error of mean and ellipsoids for $p=.75$ for all band combinations of bands 1,3,4 and 7 of the COST image. Class assignment: 1 - Closed-Canopy Forest, 2 - Succession Forest, 3 - Purma, 4 - Grassland, 5 - Agriculture, 6 - Bare Surface, 7- Water.



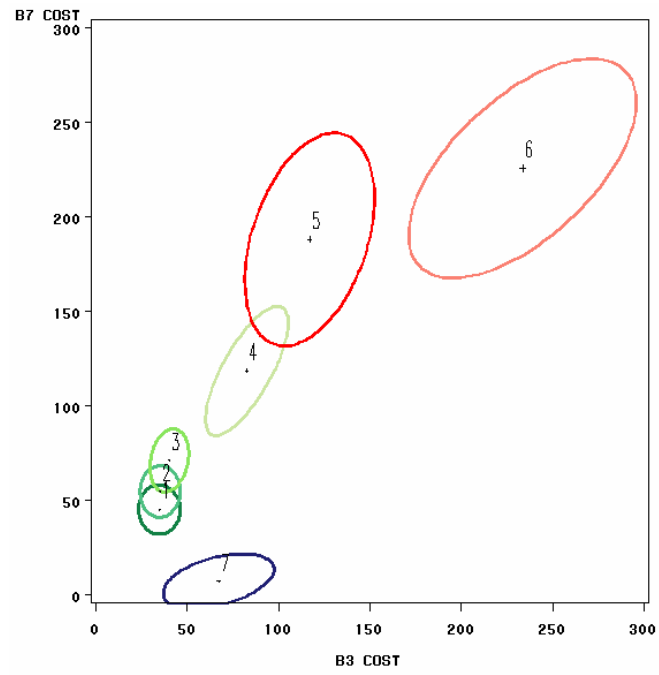
(c)



(d)



(e)



(f)

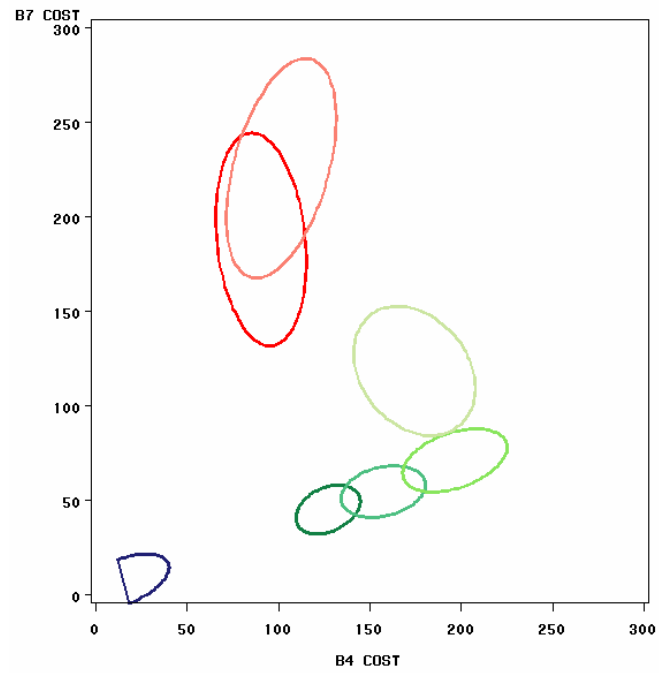
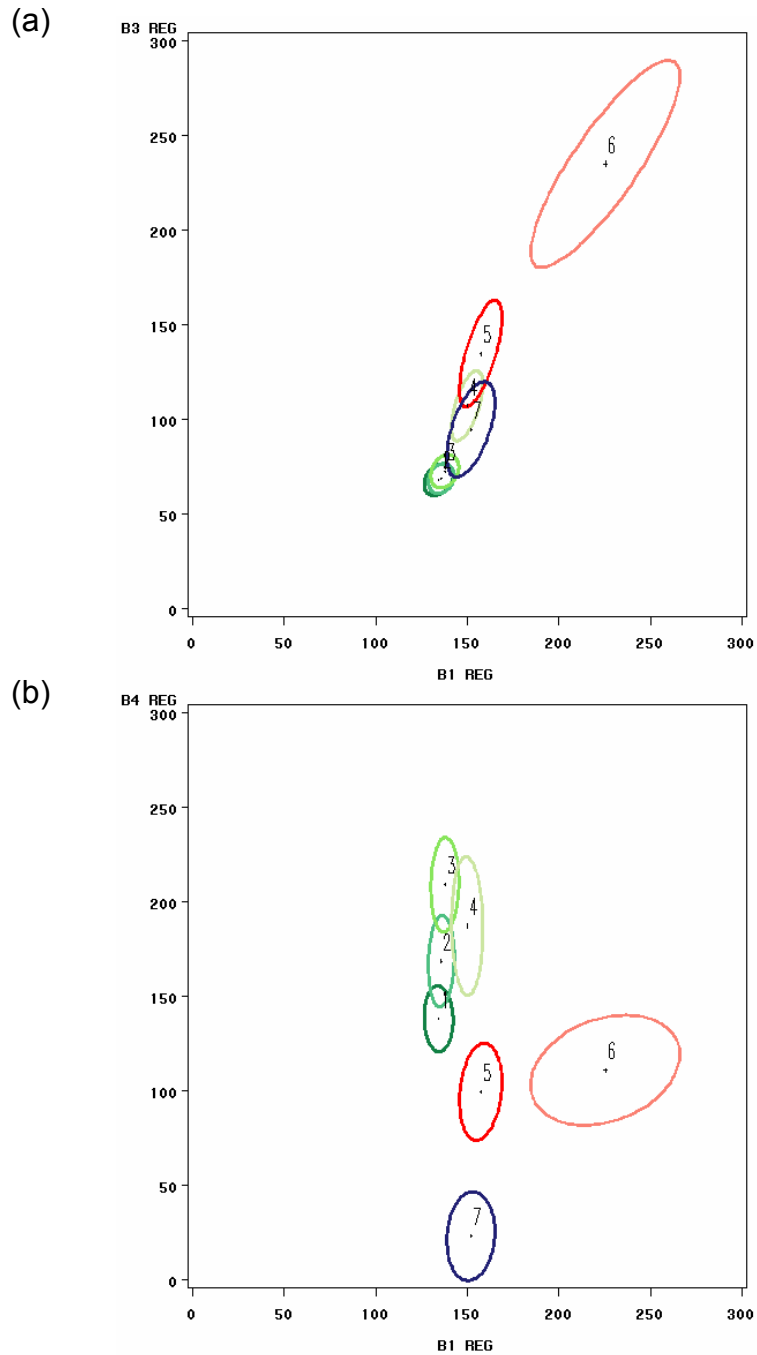
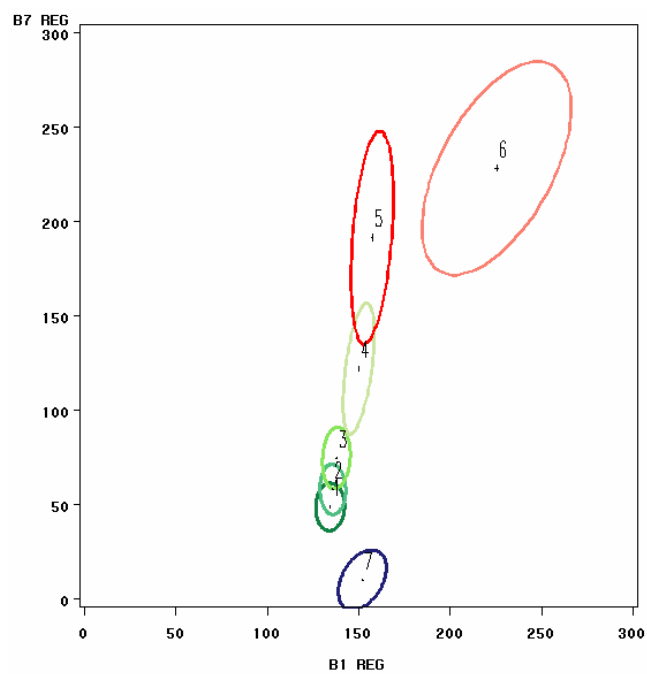


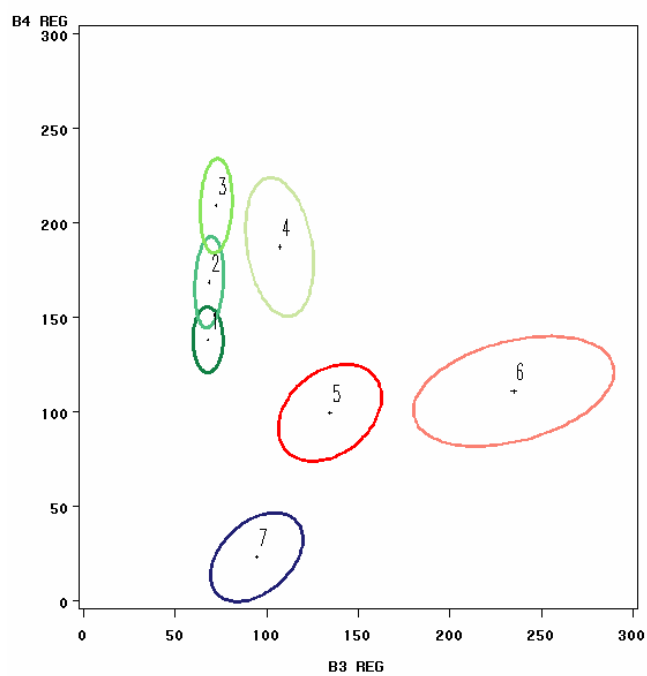
Figure 11. Graphs a – f: The graphs show mean, standard error of mean and ellipsoids for $p=.75$ for all band combinations of bands 1,3,4 and 7 of the regression corrected image. Class assignment: 1 - Closed-Canopy Forest, 2 - Succession Forest, 3 - Purma, 4 - Grassland, 5 - Agriculture, 6 - Bare Surface, 7- Water.



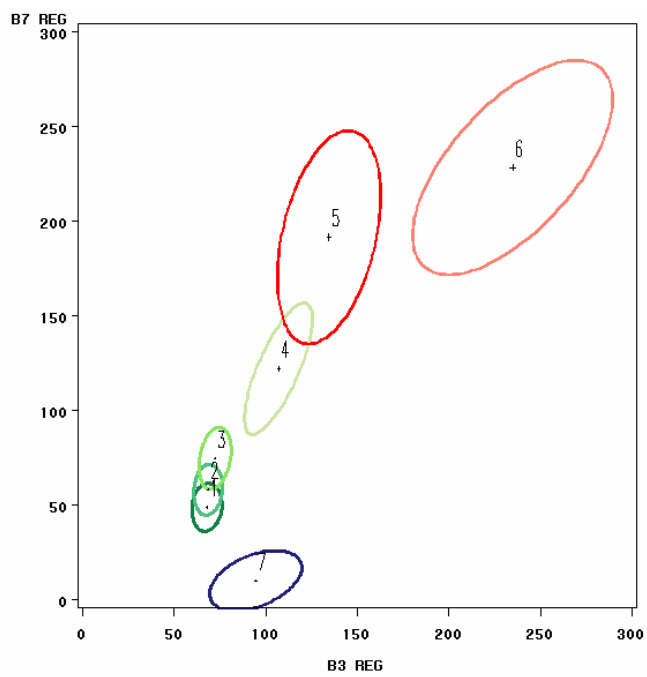
(c)



(d)



(e)



(f)

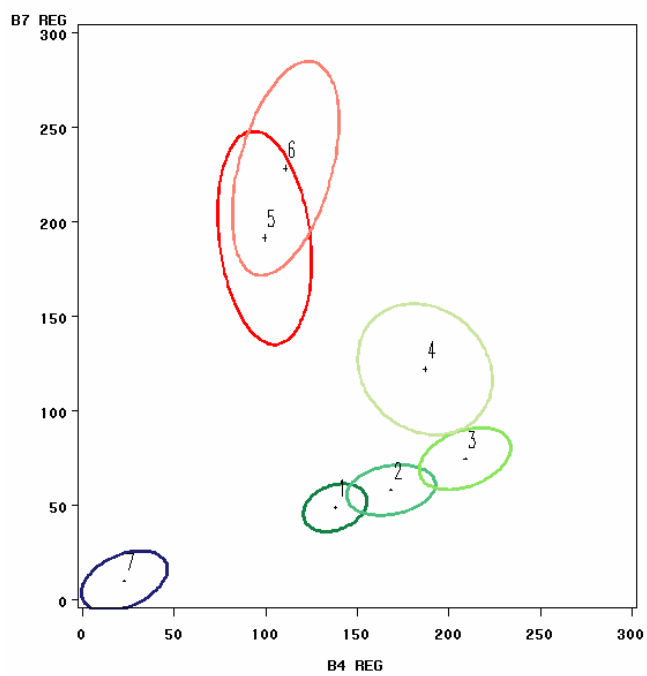
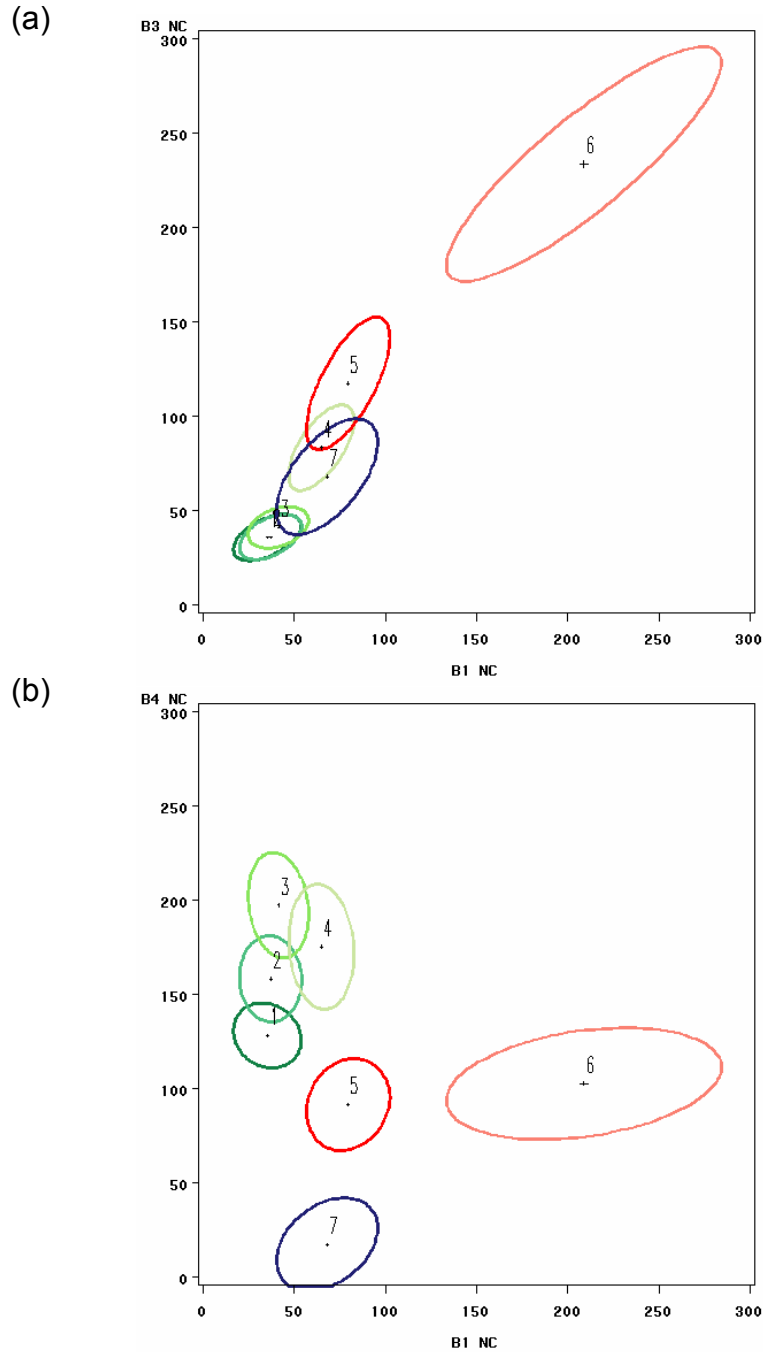
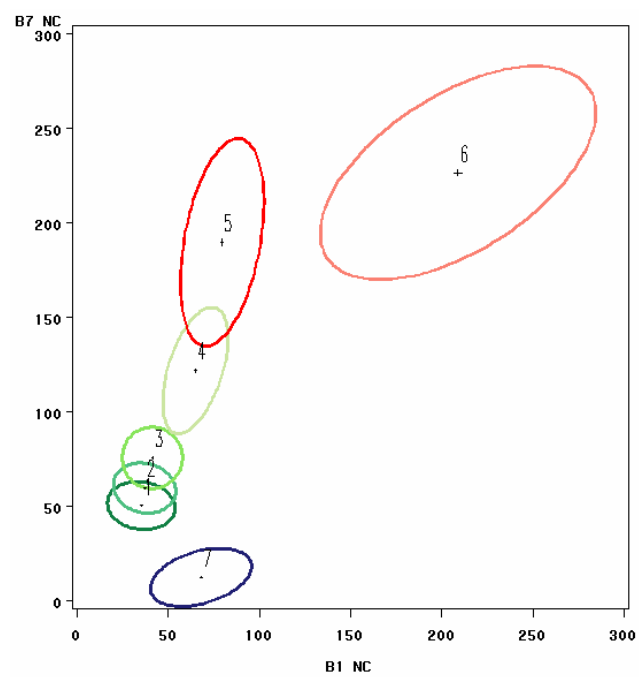


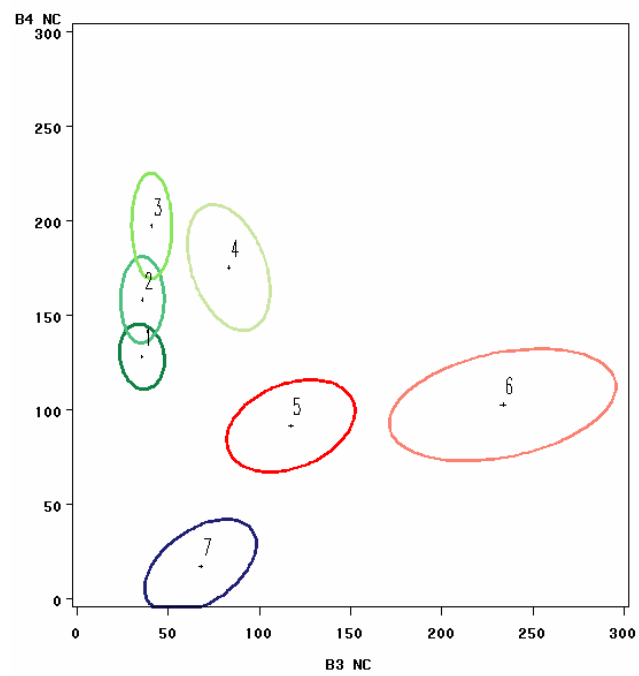
Figure 12. Graphs a – f: The graphs show mean, standard error of mean and ellipsoids for $p=.75$ for all band combinations of bands 1,3,4 and 7 of the uncorrected image. Class assignment: 1 - Closed-Canopy Forest, 2 - Succession Forest, 3 - Purma, 4 - Grassland, 5 - Agriculture, 6 - Bare Surface, 7- Water.



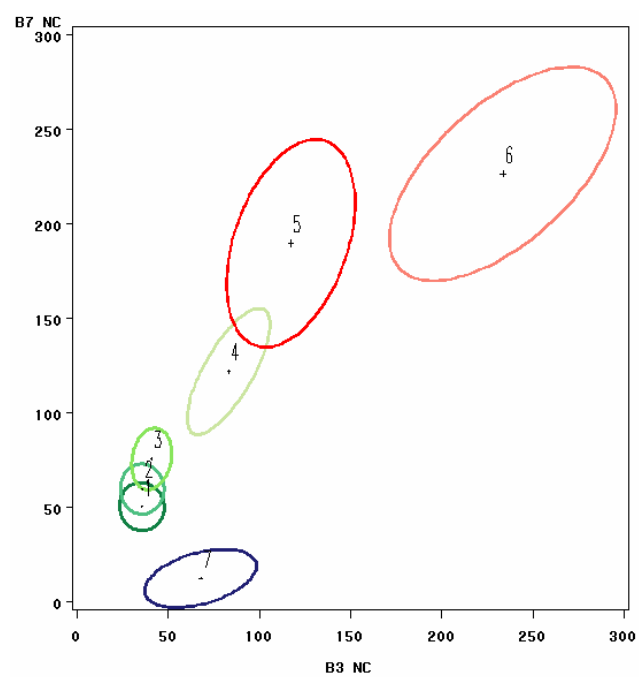
(c)



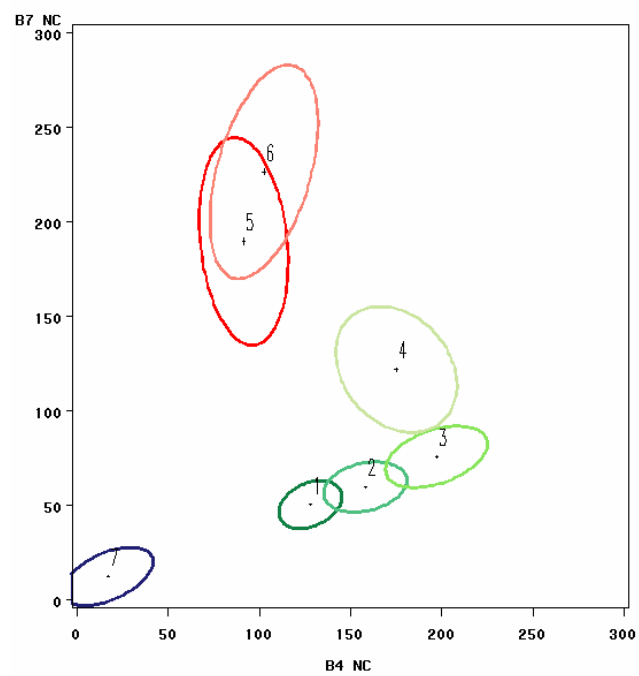
(d)



(e)



(f)



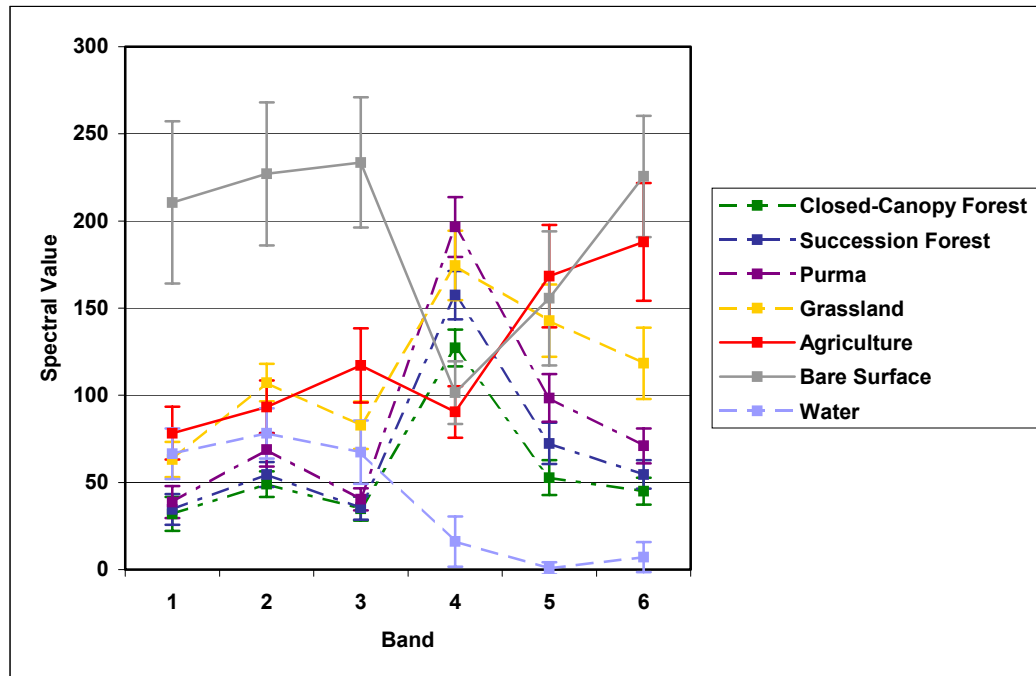


Figure 13. Spectral signatures for COST corrected images.

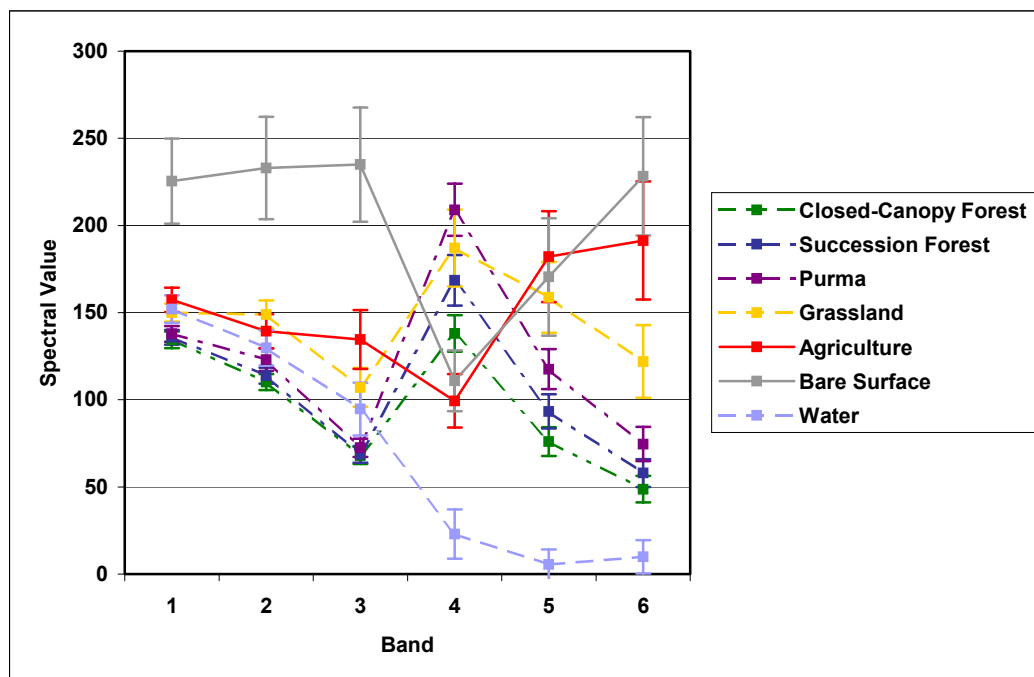


Figure 14. Spectral signature for regression corrected images.

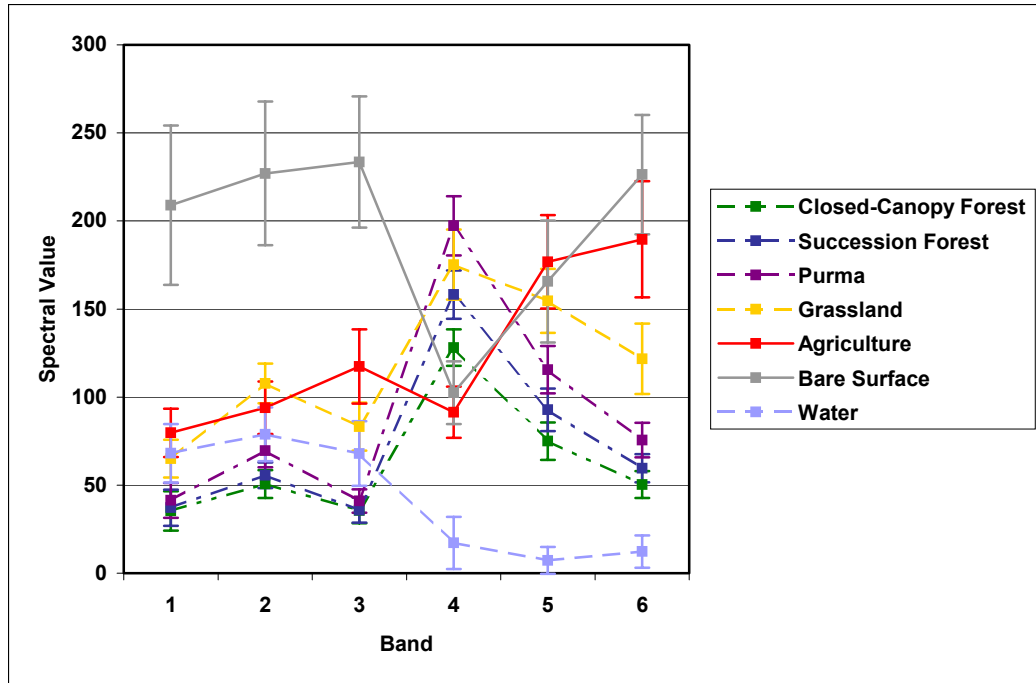


Figure 15. Spectral signatures for uncorrected images.

This leads to 3150 cycles of data splitting, calibration and validation. At the polygon level the 14 pixels that are within the same polygon as the pixel that is to be predicted are excluded from the calibration process. The calibration at region level excludes all pixels of the same region in the calibration and at image level all pixels of one image serve as calibration set to predict all pixels in the other image until each of the 3150 pixels are predicted. The results of all class assignments recorded at each validation process can be summarized in a contingency table also called confusion matrix. This crosstabulation of number of pixels in terms of reference class by predicted class allows for identification of misclassification between classes and also the expected accuracy of a classification in percent (Fig.16).

Table 3. Contingency table for uncorrected images crossvalidated at the pixel level. CCF–Closed-Canopy Forest, SF–Succession Forest, P–Purma, G–Grassland, A–Agriculture, BS–Bare Surface, W–Water.

Table of Reference by Classified								
Reference	Classified							Total
Frequency Row Pct	CCF	SF	P	G	A	BS	W	
CCF	418 92.89	32 7.11	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	450
SF	61 13.56	370 82.22	19 4.22	0 0.00	0 0.00	0 0.00	0 0.00	450
P	0 0.00	43 9.56	405 90.00	2 0.44	0 0.00	0 0.00	0 0.00	450
G	2 0.44	2 0.44	49 10.89	391 86.89	6 1.33	0 0.00	0 0.00	450
A	4 0.89	0 0.00	0 0.00	1 0.22	445 98.89	0 0.00	0 0.00	450
BS	0 0.00	0 0.00	0 0.00	3 0.67	28 6.22	419 93.11	0 0.00	450
W	2 0.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	448 99.56	450
Total	487	447	473	397	479	419	448	3150

Results of the crossvalidation at different levels for the different images that were extracted from the 12 contingency tables (see Appendix E for contingency matrixes and complete statistical output) are summarized in Table 4. The values represent percentages of pixels correctly classified, which are found on the major diagonal of the contingency matrixes. Since the number of observations for each class is 450 the values are naturally normalized.

Table 4. Accuracy in percent for all classes at the four different levels of crossvalidation for the different correction methods. UC stands for uncorrected, REG for regression corrected and COST for COST corrected images, with CCF–Closed-Canopy Forest, SF–Succession Forest, P–Purma, G–Grassland, A–Agriculture, BS–Bare Surface, W–Water. Adding all correctly classified pixels in the diagonal and dividing them by the sum of all pixels yields overall accuracy reported at the bottom of the table. Highest accuracy values for each class at each level of crossvalidation are marked in grey.

Class	Pixel			Polygon			Region			Image		
	UC	REG	COST	UC	REG	COST	UC	REG	COST	UC	REG	COST
CCF	92.89	93.78	91.56	91.56	93.56	91.56	90.67	93.11	89.78	91.11	89.11	89.56
SF	82.22	79.11	82.22	81.11	78.00	81.56	81.33	78.00	75.78	78.00	76.67	80.22
P	90.00	93.11	90.00	89.11	92.67	88.22	86.44	92.67	80.44	80.89	92.89	86.22
G	86.89	85.56	87.11	86.67	84.89	86.67	86.67	84.44	86.22	88.22	81.78	86.89
A	98.89	99.11	99.11	98.89	99.11	99.11	99.11	99.11	98.89	99.33	99.11	99.11
BS	93.11	93.11	92.44	92.89	92.89	92.44	92.89	92.67	91.11	91.78	92.89	92.89
W	99.56	99.56	99.56	99.56	99.56	99.56	99.56	99.56	99.56	99.56	99.56	99.56
Overall	91.94	91.90	91.71	91.39	91.52	91.30	90.95	91.37	88.83	89.84	90.29	90.63

The overall accuracy is defined by the sum of the diagonal in the contingency matrix divided by the total number of observations. The reporting of overall accuracy neglects errors of omission and commission by not taking into account the elements in the contingency matrix, which are not elements of the major diagonal. A statistic that includes these errors is the KAPPA statistic. The K_{hat} statistic is calculated according to Equation (16) (Jensen 1996).

$$K_{hat} = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} x_{+i})} \quad (16)$$

where

N total number of observations

r number of rows

x_{ij} number of observations in row i and column j

x_{i+} marginal total for row i

x_{+j} marginal total for column j

The Kappas for the classifications at different levels indicate that at polygon, and region levels the regression corrected images have the highest percentage of agreement between reference and classified data except for image level where the COST method has a higher value. At Pixel level there is no difference between uncorrected and regression corrected images (Table 5).

Table 5. Kappas for the different correction methods at different levels. The Kappa values are reported in percent. UC being uncorrected, REG regression corrected and COST corrected.

	Pixel			Polygon			Region			Image		
	UC	REG	COST	UC	REG	COST	UC	REG	COST	UC	REG	COST
PF	92.89	93.78	91.56	91.56	93.56	91.56	90.67	93.11	89.78	91.11	89.11	89.56
SF	82.22	79.11	82.22	81.11	78.00	81.56	81.33	78.00	75.78	78.00	76.67	80.22
P	90.00	93.11	90.00	89.11	92.67	88.22	86.44	92.67	80.44	80.89	92.89	86.22
G	86.89	85.56	87.11	86.67	84.89	86.67	86.67	84.44	86.22	88.22	81.78	86.89
A	98.89	99.11	99.11	98.89	99.11	99.11	99.11	99.11	98.89	99.33	99.11	99.11
BS	93.11	93.11	92.44	92.89	92.89	92.44	92.89	92.67	91.11	91.78	92.89	92.89
W	99.56	99.56	99.56	99.56	99.56	99.56	99.56	99.56	99.56	99.56	99.56	99.56
Overall	91.94	91.90	91.71	91.39	91.52	91.30	90.95	91.37	88.83	89.84	90.29	90.63
Kappa	90.60	90.60	90.30	90.00	90.10	89.90	89.40	89.90	87.00	88.20	88.70	89.10

In a final step the Kappas were compared to each other using a Chi-Square test (see Appendix E for complete report) to determine if there are significant differences between the classification accuracies for the different

correction types. The comparison of Kappas indicates that there are no significant differences in the accuracy of classifications at $\alpha = .05$ (Table 5).

Table 6. p-values of the Chi-Square for comparison of pairs of correction methods at all levels.

	Pixel	Polygon	Region	Image
Uncorrected vs. Regression	0.9631	0.8568	0.5638	0.5553
Uncorrected vs. COST	0.7475	0.8930	0.6630	0.1915
Regression vs COST	0.7827	0.7528	0.3111	0.0579

A final land-cover map for the study area was created by classifying the regression adjusted mosaic with a maximum likelihood classifier. All 3150 reference samples were used in the calibration or training process, since the highest overall accuracy of 91.94 percent (Table 4) was achieved at that level. A refinement of the maximum likelihood classification was necessary to eliminate noise.

Some parts of the classified map have a high noise level due to high heterogeneity over small spaces leading to many mixed pixels that could be classified in several categories. In order to eliminate as much noise as possible from the map, a fuzzy convolution filter was applied utilizing the distance files. The filter used was a weighted 3x3 matrix with equal weights of .75 for the non-center cells. Since the minimum mapping unit was set to be 0.4 hectares clumps of pixels smaller than 5 pixels were eliminated using a majority filter at each pixel of the small clumps. The final map is shown in Figure 16.

Conclusion

Even though a t-test test for evaluation of the radiometrically rectified images indicated significant differences for bands 4,5 and 7 between the pairs of uncorrected and COST corrected images, the discrimination of classes at all levels of calibration can be performed with no significant difference in classification accuracy. For the regression corrected method the t-test indicates that the images are not different for any band.

Confusion between classes occurred where expected between closed-canopy forest and succession forest, succession forest and purma, between purma and grassland as well as purma and agriculture. Agriculture also can be confused with grassland and bare surface in cases where soil is exposed or the crop is at an early stage of growth. Closed-canopy forest and succession forest are very similar in terms of reflectance throughout the entire spectrum (Fig.10 – Fig.12), therefore the confusion is expected to be highest between those classes which is indicated by the low transformed divergence values below 1700.

Overall the classification scheme is very coarse and classes are well discriminated, therefore it seems not to be necessary to radiometrically adjust the images for the classification scheme used. However with more complex and detailed schemes attempting to distinguish between for instance crop types or forest types classification accuracy might differ for the different correction methods.

Very encouraging is the fact that classification accuracy does not drop significantly when using pixels from one image in the calibration process and applying those signatures to another image. This means that for the given classification scheme the collected samples might be good for classification of images included in the future in an endeavor to monitor ongoing changes. Even if more reference samples are collected in the field they could be incorporated in the classification process as long as the data source provides consistent data sets over time.

To what degree extension of class signatures across space outside the boundaries of the study area is valid needs to be evaluated. This is especially of interest for regions where the naturally occurring vegetation cover differs drastically from that within the study area. Extension across elevation gradients for instance may not yield good classification results due to different species composition and therefore spectral reflectance of the different forest categories. Another difficulty could be encountered when it comes to distinguishing between naturally occurring grasslands in the higher part of the basin and cultivated grassland for the purpose of detecting human intervention.

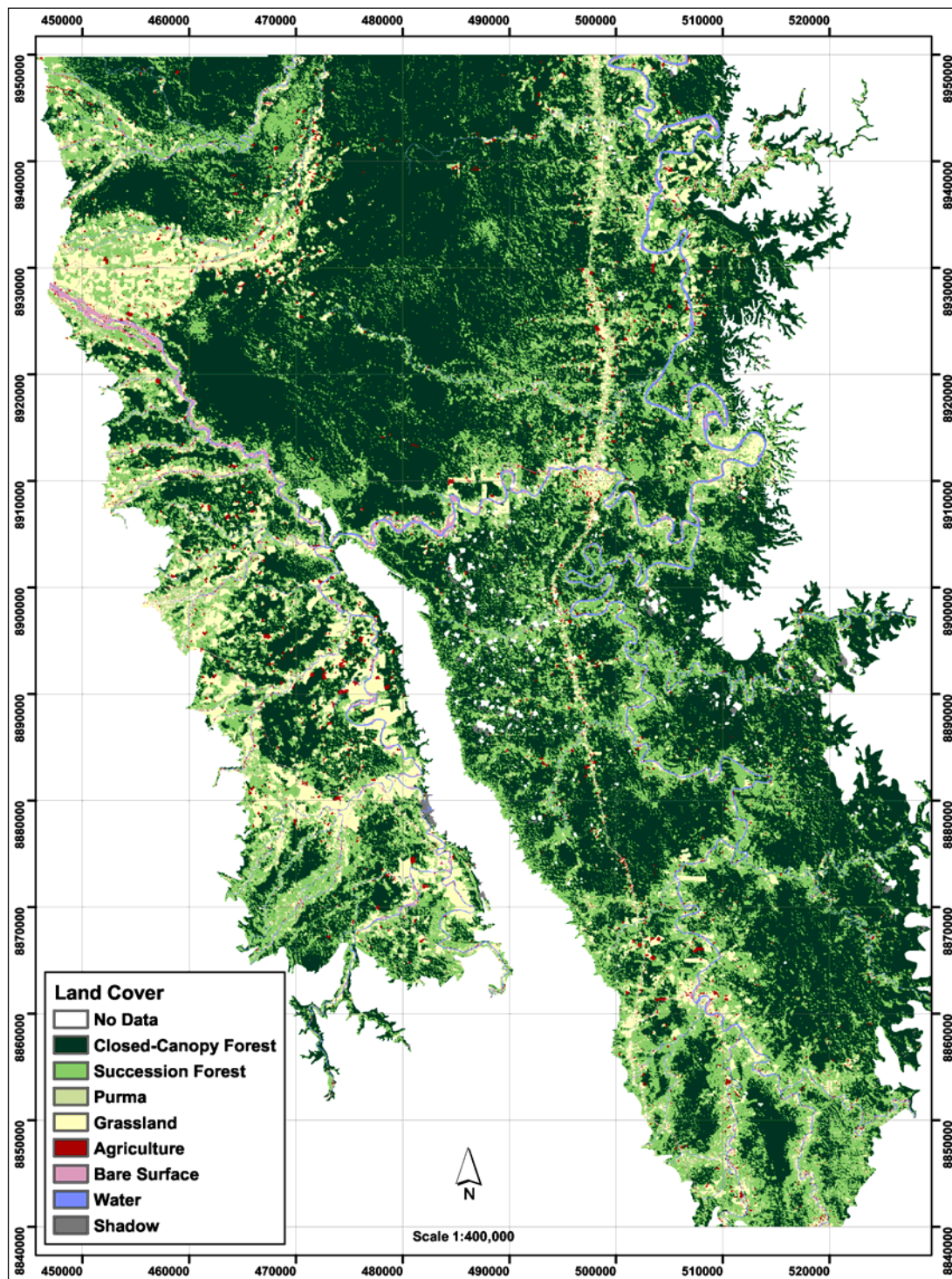


Figure 16. Land-cover map of the study area. The study area is limited by elevations below 500 masl and slopes of the terrain less than 5 degrees. The 'No Data' class also includes areas where clouds have been masked. The minimum mapping unit of the map is 4,061.25 square meters or about 0.4 hectares.

Map Projection: Universal Transverse Mercator Zone 18 South
False Easting = 500,000 Meters False Northing = 1,000,000 Meters

Datum: WGS 1984

CHAPTER 2

ASSOCIATIONS BETWEEN LANDSCAPE STRUCTURE AND NATURAL AND ANTHROPOGENIC FACTORS

The natural land-cover of the eastern slopes of the Andes is mainly a function of elevation and climate. For the study area in the Pachitea Basin the natural land cover corresponds to five life zones according to the classification system by Holdridge (Holdridge 1967), which are tropical moist forest, pre-mountain tropical wet forest, tropical wet forest, pre-mountain tropical rain forest, and low-mountain tropical rain forest (Fig.17). All these categories are forest categories, which indicates that the natural land cover of the entire area is some type of forest and whenever the current land cover is grassland or agriculture it is caused by human alteration. The degree to which each land use is represented in a region depends on various natural and cultural factors and their interactions. Natural factors that facilitate or impede certain land uses over others are climate, geology and relief, soil and water conditions and the natural vegetation (Vink 1975). The cultural factors that affect the land use are origin of people using the land and their traditional way of living, including their diet and life style as well as political decisions at different political levels. The factors that are anticipated to affect decisions on land use and therefore land-cover changes can be divided into the three categories; bio-geophysical factors of the terrain, socioeconomic or demographic factors, and market accessibility factors and their interactions. Each of these categories holds innumerable variables and each can be analyzed at different scales and at different aggregation levels.

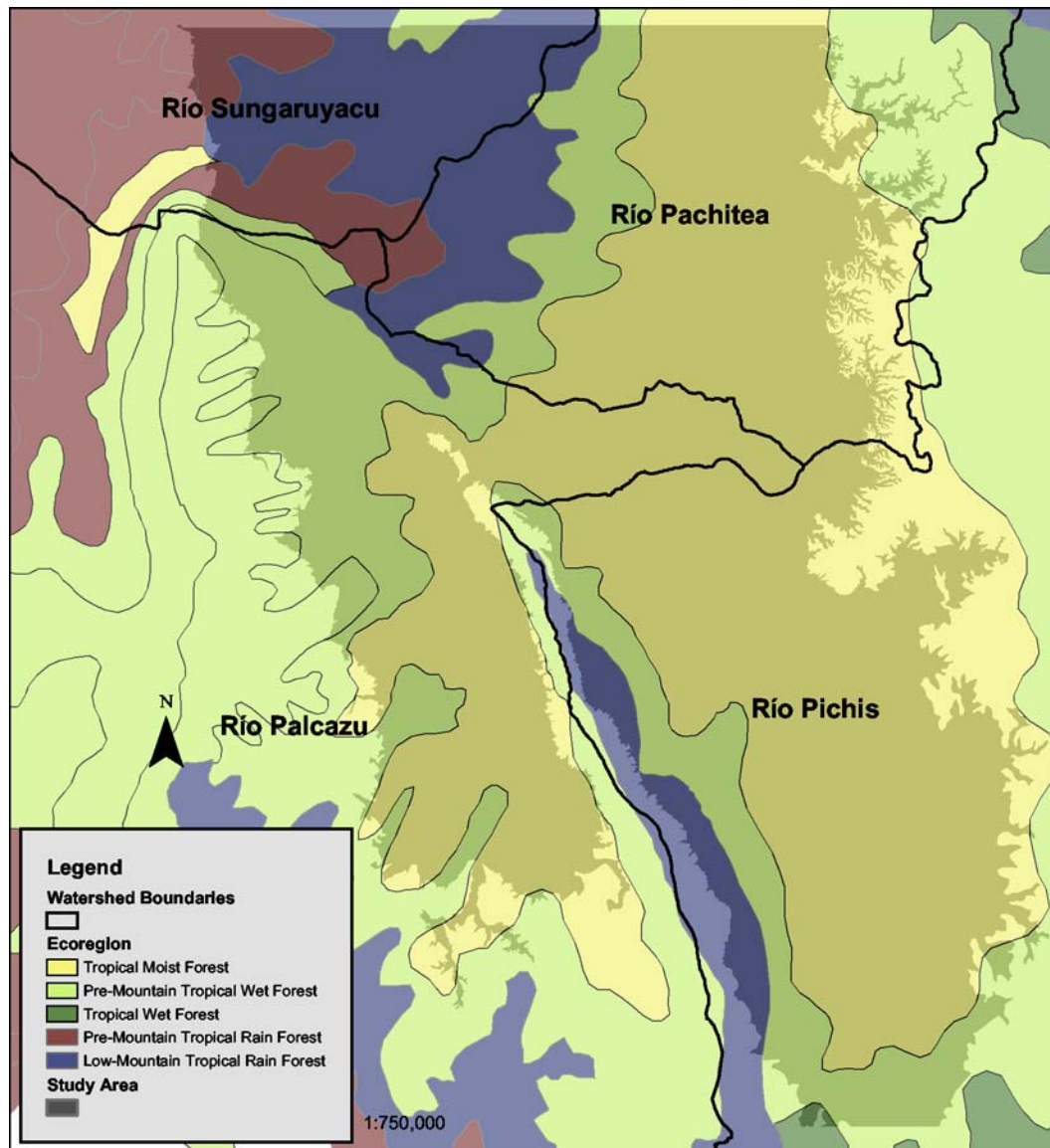


Figure 17: Five life zones according to the classification system by Holdridge are present within the study area of the Pachitea Basin; all of them being forest categories (Tropical Moist Forest; Pre-Mountain Tropical Wet Forest; Tropical Wet Forest; Pre-Mountain Tropical Rain Forest and Low Mountain Tropical Rain Forest).

A review of each category and its application in different models as well as a historic synopsis of the colonization of the basin will help to select variables

that explain variability of land-cover composition and configuration within the study area.

Landscape Composition and Configuration

Understanding the dynamics of land-cover changes, the forces that are driving them and the impact they have on ecosystems depends on the ability to describe the landscape and to identify the sources that promote change and those that maintain stability. The study of landscape function is based on the ability to quantify landscape structure and to measure the changes in structure. The definition of a landscape in terms of extent and grain size depends on the phenomenon under consideration. Since the patterns detected in a landscape mosaic are scale dependent the scale of investigation or analysis needs to be defined first. All patterns that are detected are then described relative to that scale (Forman 1995).

The landscape can be described either in terms of composition or in terms of configuration. Landscape composition describes the landscape in terms of presence and abundance of each cover type using fractions of the landscape occupied by each category, or total areas, without being spatially explicit. Landscape configuration, on the other hand, refers to the spatial characteristics and distribution of land cover within a specific landscape mosaic (Forman 1995). Considerable effort has been undertaken to develop landscape metrics or indices (O'Neill et al. 1988). Landscape configuration can be expressed by indices

derived from an analysis of size, shape, and spatial distribution or arrangement of the elements that are forming the landscape mosaic. Indices to describe the landscape as a whole are measures such as diversity, dominance, contagion, and the amount of edge between each land-cover category. At the patch level spatial attributes can be described in terms of shape indices, core areas, nearest neighbor probabilities and the complexity of patch perimeters using fractal dimensions (Johnson 1990, Turner 1990, Pastor 1992). Markov chain models have been used to describe the transition probability of individual cells from one class to another as a function of neighborhood cells (Pastor 1992, Lambin et al. 1999). These indices are highly scale dependent and therefore must be used with caution.

For the study area in the Pachitea, landscape composition is described in terms of abundance of grassland and agriculture, closed canopy forest and the two stages of succession forest. The landscape configuration is considered in a spatially more explicit framework of land-cover distributions as a function of distance from roads and rivers.

Theoretic Concept of Land-Cover Variability

The current land-cover pattern in the Pachitea is a reflection of the natural characteristics of the region, and the way people use the land, which in turn is a result of bio-geophysical characteristics of the landscape as well as personal and governmental decisions based on socioeconomic and political developments.

The dynamics of land use and land-use change in the case of the Pachitea can be traced back to the 19th Century. The goal of this research is to describe and explain the configuration of the landscape and to identify factors that can be associated with the variability of different patterns within the study area based on bio-geophysical parameters and in view of the socioeconomic and political developments during the last 150 years.

It has been shown that factors that are significant at a local scale might be irrelevant at a coarser scale and vice versa (Verburg et al. 1999). For that reason there is a general consensus that the significance of a variable depends on the spatial and temporal scale of the analysis (Geoghegan 1998, de Koning et al. 1999, Lambin et al. 1999, Verburg et al. 1999). The choice of variables at different scales will determine how realistically a model describes local, regional or global dynamics. Therefore the consideration of variables at a local scale needs to take place in the context of developments at the regional and national scale. The scale of research that has been chosen for different studies that include socioeconomic and demographic data varies remarkably. Some models that attempt to establish linkages between socioeconomic or demographic variables and land use are at a global scale (Bilsborrow and Okoth Ogendo 1992) others are at a very local scale (Kammerbauer and Ardon 1999). Several models have been developed to identify the causes of human induced land-cover changes that are correlated to deforestation in the Brazilian Amazon at the State level (Wood and Skole 1998) at the plot and county level (Pfaff 1999) and at the *Parroquia* and *finca* level in the Ecuadorian Amazon (Mena 2001). Work that

specifically concentrated on the Selva Central in Peru focused on agricultural intensification, deeding of land title, cash crop revenues at the plot level and ethnicity (Bedoya 1995, 1996).

Since the spatial and temporal resolution of remotely sensed data and data provided by the social sciences rarely correspond to each other, the synthesis of those different data sets may provoke major difficulties. To be able to link the data sets, they need to be transformed to a common level (Rindfuss and Stern 1998, Lambin et al. 1999). This can be achieved by setting the unit of analysis according to the data set with the most restricting aggregation level (Wood and Skole 1998). The most restricting data sets tend to be those that include demographic data. The drawback of adjustment at the coarsest level is that the possible finer resolution of the remotely sensed data sets is sacrificed. Therefore socioeconomic data, which only exists at district level for the study area, is not very useful.

A synopsis of the historical development of the Selva Central region as a larger framework, and especially the Pachitea Basin as a part of it will suggest parameters that are expected to contribute to the variability of land-cover patterns within the local study area.

The Selva Central

The historical and political developments that can be linked to the colonization and deforestation in the Selva Central are mainly governmental

policies that have been instituted within the last 150 years. These policies include immigration laws, economic policies, interior and foreign politics, agricultural development and infrastructure projects (Moore 1989, Bedoya 1995).

For several reasons the Selva Central, encompassing the current provinces Oxapampa, Chanchamayo and Satipo, was left untouched during three time periods when other parts of the Peruvian Amazon were invaded. In pre-Inca time the lack of important population centers in the neighboring Andean highlands prevented migration pressure. Later during the epoch of the Inca Empire the Inca town of Huánuco expanded mainly towards the northern regions, and during the destruction of the Inca Empire and the following period of Spanish settlement the main targets of the Amazon regions were those that were famous for their rich gold deposits. Things changed with the opening of the silver mines around Cerro de Pasco in 1632 to 1642, which led to a major population influx into the sparsely populated area of the Andean highlands west of the Selva Central (Santos-Granero and Barclay 1998).

In the 18th century the number of missions increased and the missionaries from the Huánuco conversion, in the North, succeeded in occupying parts of the Selva Central. The missionaries and Andean peasants that came with them established the first large and medium-size *haciendas*. The missions mainly produced coca and sugarcane for the production of *aguardiente*. An uprising of the indigenous population triggered a retraction of the colonization frontier from 1742 to 1847 (Santos-Granero and Barclay 1998).

With the stagnation of silver production in the Cerro de Pasco area and a rise in coffee prices at the end of the 19th and beginning of the 20th Centuries a rapid expansion of coffee plantations occurred in the Selva Central, substituting the sugarcane economy. This development brought the opportunity for the people in the Selva Central to expand their range of trade from the *sierra* to the coast, supplying Lima with export goods. With the establishment of this export economy, interest in the Selva Central increased at a national level, as reflected in the number of people entering the region as work labor (Santos-Granero and Barclay 1998).

Studies conducted in the region of the Selva Central indicate that land-use patterns are not homogeneous for the entire region. Land-tenure systems, agricultural intensification, availability of land, time of colonization, connectedness to markets, and increasing population density seem to be interrelated and can help to describe the land-use pattern in the region (Santos-Granero and Barclay 1998). The transformation from forest to an economically viable agricultural system is generally described as a progression where the logging industry is identified as the spearhead of the market economy and the subsequent agricultural activities, especially perennial crops such as coffee and fruits, as the economic stabilizer (Santos-Granero and Barclay 1998).

Santos Granero and Barclay distinguish between three stages of development in that progression. Centers of economic activity are defined as areas that show a high population density and that are primarily dedicated to fruit

and coffee production. These areas are linked to the national and export market. The high colonization pressure in the second half of the 20th century led to a subdividing of the large haciendas into mid-size and small-size land holdings in these regions. In addition to the colonization pressure the agrarian reform of the Velasco Regime (1968-1975) pursued this development ending the vast domain of large landholders (Moore 1989). The result is smaller average land holdings, which can be also described as higher land fragmentation (Bedoya 1995). This extreme property fragmentation in return caused land owners to intensify the use of their land keeping a very high percentage of their land at all times under production. This means not enough land left fallow for too short periods of time that are not sufficient for soil recovery (Santos-Granero and Barclay 1998).

The peripheral areas on the other hand are areas that have a short history of connectedness to the centers of economic activity. The low-density areas that are only poorly connected to any of the other two and do not have any urban centers form the economic frontier where logging is the predominant source of economic revenues, agriculture takes place almost exclusively as subsistence (Santos-Granero and Barclay 1998). The distinction of these different stages of economic progression is useful when it comes to establishing a link between the different land-use patterns and factors that promote certain land-uses over others. Some factors though that seem to be obviously linked to land-use development are not as straightforward when seen in the context of other factors. It seems that time of colonization for example is not necessarily a very important variable when looking at intensification of agriculture and land-tenure system,

since areas in watersheds that were more recently colonized than the Pachitea saw a higher rate of intensification and subdivision of agricultural units which led to a decrease of land holdings. The explanation provided is that value of the land is not only determined by the suitability of the land for agricultural activities but rather the connectedness to markets, which triggers an influx of people (Santos-Granero and Barclay 1998). But even if the peripheral or expansion areas are connected to market centers there are at least two factors that can inhibit further development. One is the fact that coffee production is only viable in higher altitudes and the production of fruits for distant markets poses a logistic challenge. Therefore lowland agricultural activities in the peripheral and frontier areas that could serve as stabilizers still need to be found (Santos-Granero and Barclay 1998). Many of the peripheral areas in the lower parts of the Amazon basin where production of coffee is impossible therefore saw an increase in alternative perennial cash crops that do not have the logistic problem of fresh fruits such as coca and in recent years hearts of palms as practiced in the Pichis valley. Coca of course meets the challenge of being an illegal crop and the viability of the hearts of palm, as any other crop, still needs to be demonstrated. This could be one of the reasons why those peripheral and expansion areas are dedicated mainly to cattle ranching whereas agriculture serves primarily for subsistence.

Within the Pachitea all three categories of economic development are represented but only the area around Oxapampa (Fig.21) qualifies as a small center of economic activity directly connected to Lima. The Palcazu and Pichis

for the major part are peripheral areas that were most recently connected to Oxapampa and Villa Rica. The lower Pozuzo which is connected to other parts of the region only by river and narrow trails as well as large parts of the Pachitea are those that can be labeled economic expansion frontier (Fig.21). Variability in land-use patterns at a local scale taking into consideration only peripheral and extension areas of the basin still needs to be demonstrated. Those patterns are expected not to be random in nature but can rather be associated to explicit identifiable factors.

Variability in the Peripheral and Expansion Areas of the Pachitea

A review of the bio-geophysical landscape characteristics of the landscape and the historic development within the borders of the study area will help to identify factors that could be associated with variability in land-use patterns at the local scale. Specific attention is brought to the lowland areas of the lower Pozuzo valley as part of the Palcazu watershed and the valley of the Palcazu River itself, the watersheds of the Pichis, the Sungaruyacu a tributary to the Pachitea and the southern part of the Pachitea (Fig.6).

Bio-Geophysical Factors

The incorporation of biophysical and geophysical components into an analysis of land-cover alterations helps to gain a better understanding of the

natural attributes of a location that will increase or decrease the probability of changes to occur. Differences in altitude, climate, soil characteristics and original vegetation might have an influence on the choice of land utilization (Vink 1975).

By defining the study area below 500 masl and with slopes of less than 5 degrees, aspects of altitude and slopes as possible factors for land-use variability have been eliminated. The suitability of land for different uses can be summarized in best land use maps. These maps suggest use of the land based on the slope of the land, the type of soil and precipitation patterns.

The optimal condition for the distribution of land cover would be to closely match recommended best land use. This means that in areas for which the best land use is suggested to be "Protection", agriculture and grassland would be absent and purma and secondary forest would occur only sporadically closing naturally occurring tree-fall gaps. In the categories that allow forestry we should not encounter agriculture or grassland but we could expect moderate levels of secondary forest and purma. For the categories agriculture and grassland we expect to see a mixture of all types of land covers where primary forest and secondary forest indicate potential room for "suggested" conversion to agriculture and pasture. Discrepancies between actual land cover and best land use poses a conflict in the attempt to prevent primary forest degradation with little long-term benefit for activities such as agriculture and cattle ranching. The inconsistency between forest degrading as a result of agriculture and ranching is called a conflict whereas the remainder of primary forest and secondary forest on land

suitable for agriculture and pasture is labeled underutilized (Fig.19). Clearly the conflict areas are those that need to be protected. Underutilization of potential farmland can only be encouraged as long as not other, protected areas or areas designated to forestry are converted into agricultural plots or pastures.

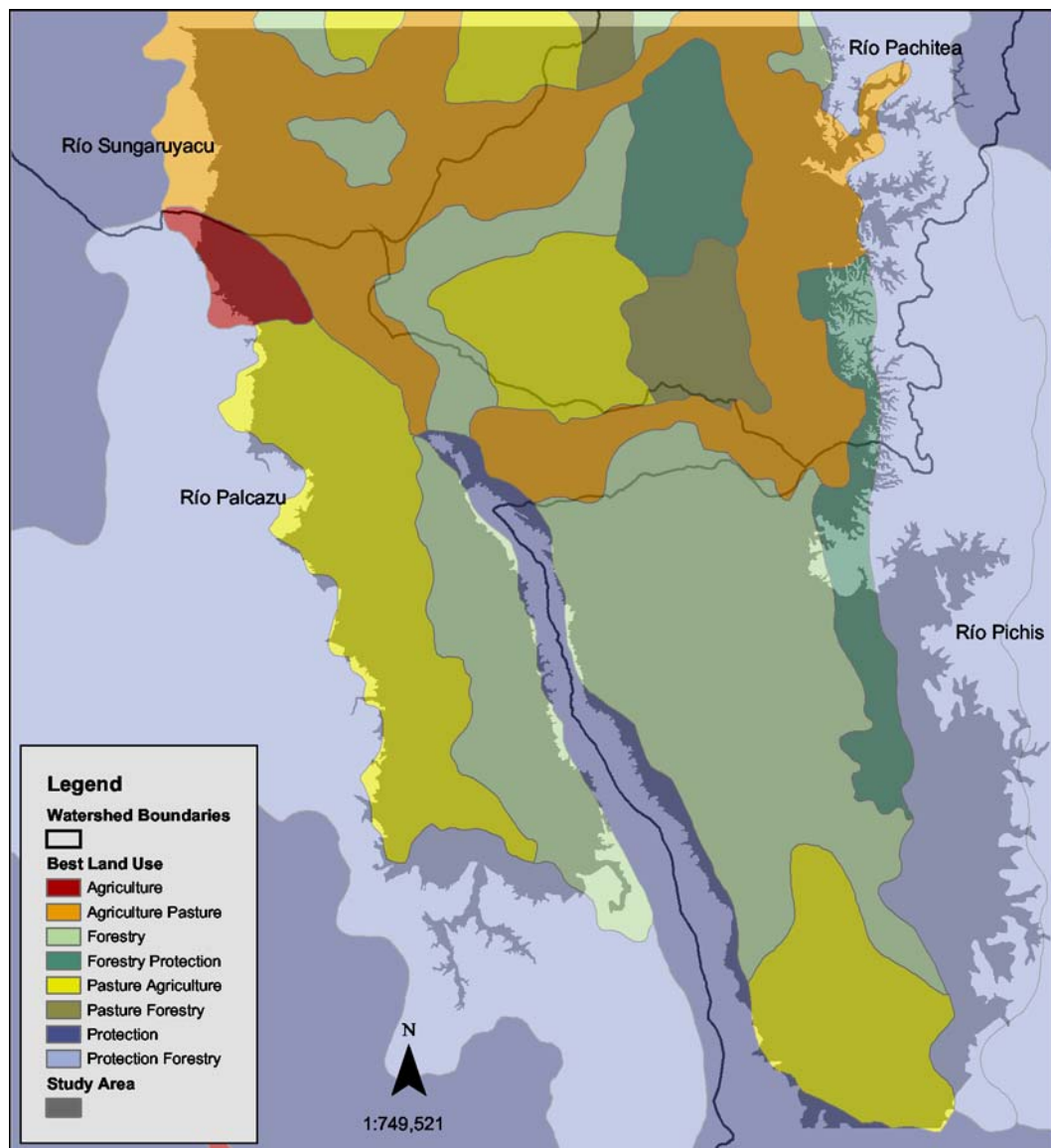


Figure 18: Best land use categories based on soil, slope and precipitation pattern (Data source: Map by ONERN at a scale of 1:100,000).

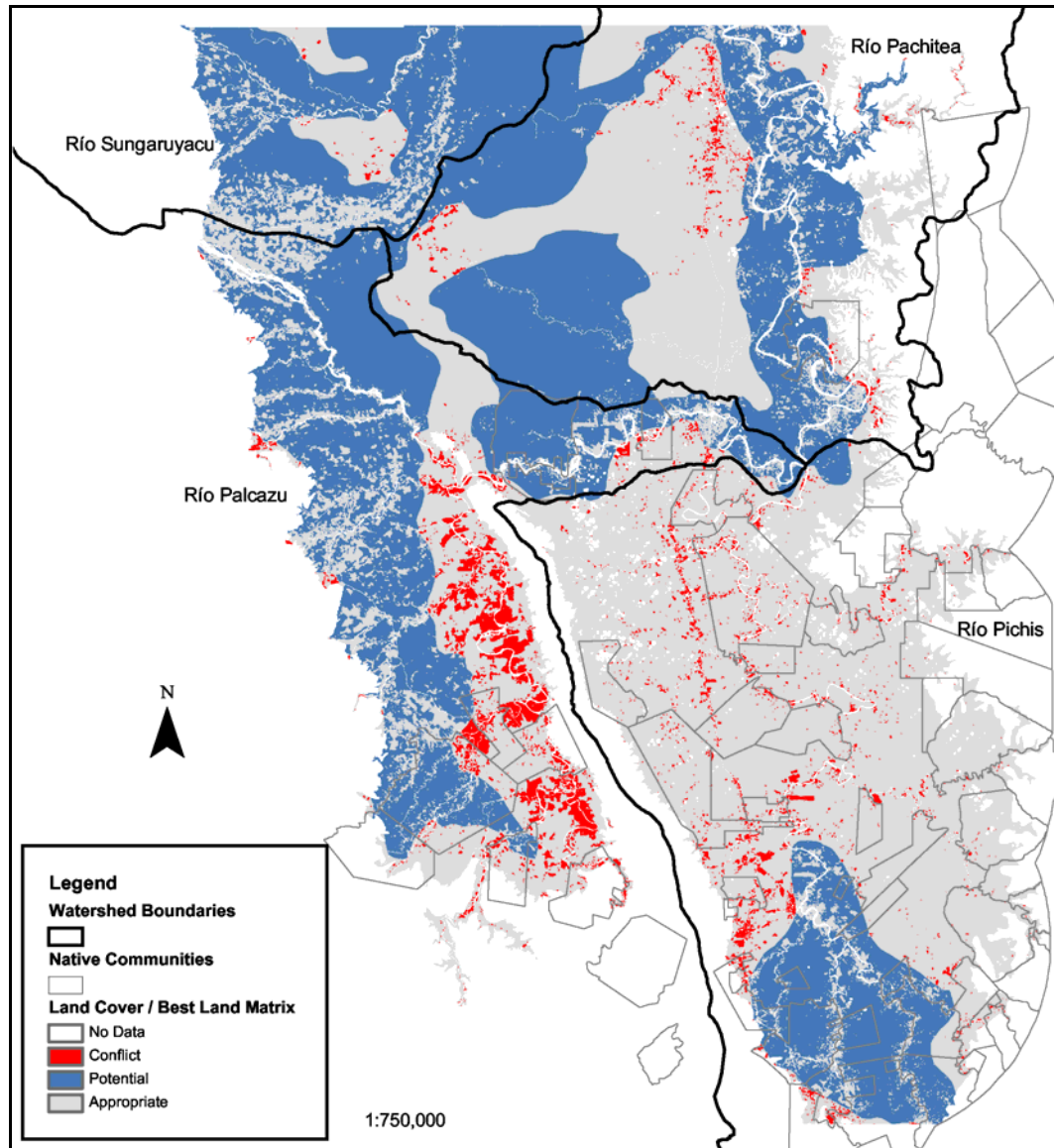


Figure 19: Matrix of current land cover and best land use practices. Map shows the areas that are used inappropriately in red; the ones that could be used for agriculture or pasture but are currently forested in blue and those that are currently used appropriately in gray. The matrix has been generated according to the categorization provided in Table 7.

A matrix of actual current land cover and best land use designation (Table 7) summarizes the areas of conflict and highlights areas that are

underutilized and have potential for future beneficial agriculture or cattle ranching.

Table 7. Land cover by best land use category indicating conflicts and underutilization. Prot – Protection; For – Forestry; Agri – Agriculture; Pas – Pasture. The Category mentioned first being the more suitable. The grey fields indicate not optimal land cover agreement with best land use. “C” stands for conflict at the cost of forest. “P” indicates potential suitable land for Pasture and Agriculture, clear indicates currently used appropriately.

Land cover	Prot	Prot - For	For - Prot	For	Agri - Pas	Pas - Agri	Agri
Closed-Canopy Forest					P	P	P
Secondary Forest					P	P	P
Purma	C						
Grassland	C	C	C	C			
Agriculture	C	C	C	C			

Ethnicity – Native Communities versus Colonists in the Pachitea

For a long time the regions east of the main range of the Andean mountains were remote areas touched only by the sustaining activities of native peoples. The first people known to settle within the borders of the Pachitea basin were probably peoples speaking the Arawakan language. These were the ancestors of the Yanesha (Amuesha) and Asháninka (Campa) Tribes that moved into the region before 2000 B.C.E. For about 3,700 years these tribes lived in total isolation without contact to the neighboring civilizations that came and left (Moore 1989).

The experiences with colonists from the very beginning influenced and impacted the traditional life of the indigenous people of the region. The production of iron tools for agriculture used by the missionaries were so well

established among the native population that even during the time span between 1742 and 1847 when they re-conquered their territory, the natives maintained the missionaries cattle and crop production and mission forges for tool production (Moore 1989). The first successful occupation of the Pachitea was a result of an immigration law to import foreign labor forces with the intention to colonize the jungle areas of Peru, which was passed by president Ramón Castilla in 1849. On July 25, 1859 a group of 296 German colonists from the Rhine Valley and Austrians from Tyrol settled at the Pozuzo River in the Province of Oxapampa of which only some 150 colonists managed to establish themselves (Moore 1989).

In contrast to other immigration groups that colonized the southern parts of the Selva Central establishing trading bonds with the city of Tarma, the colonists of the Pozuzo remained totally isolated from the rest of Peru only sporadically exporting their coca, tobacco and *aguardiente* to the highlands of Huánuco. The reason for this isolation was the difficulty to access the valley. The colony was totally self-sufficient, mainly raising livestock and to a lower extend growing crops for their own consumption (Moore 1989). Shortly after 1880 the first rubber tappers began to appear in the lower Palcazu Valley, interacting with the indigenous people of the region that were established in the Palcazu. During the rubber boom several thousand hectares of land were distributed to each of the big rubber companies, which forced the migration of Yanesha Indians into more remote areas of the Palcazu Valley, but their dependency on the colonists was already established. Colonist pressure on the Palcazu Valley increased tremendously in the 1880's when Franciscans opened several new missions and

again in 1891, as British bondholders were granted two million hectares of land for large-scale coffee production. The descendants of the rubber tappers and the colonists from Pozuzo who first expanded their agricultural activities to the area that is known as the Codo del Pozuzo and in the 1940's into the Palcazu Valley are among the largest landholders and cattle ranchers in the Palcazu Valley today (Moore 1989).

This process was intensified when the Cerro de Pasco Corporation started oil exploration in the 1960's. They were unsuccessful and when they left the valley they handed the land over to colonists from San Martín and Ucayali that had arrived with them. At about the same time several waves of poor colonists from the highlands and coastal areas of Peru entered the region hoping for new roads and agricultural development projects that were proposed by president Belaúnde (1963 – 1968 and 1980 – 1986) (Moore 1989). This means that the last part of the 20th Century the population of the Pachtea was made up of three large ethnic groups, European settlers, *mestizos* from the Andean highlands and indigenous people.

Although the allocation of land to indigenous collectives took place at the end of the 19th beginning 20th century, it was not until 1957 that the Peruvian State passed a decree law to protect and reserve land for the indigenous population. The colonization pressure on the indigenous population for almost one hundred years starting in 1849 displaced them several times and drove them to more and more remote areas in the Palcazu and Pichis Valleys (Santos-

Granero and Barclay 1998). The colonists throughout the entire Selva Central not only forced native people off their land; they were also abused as cheap laborers in rubber tapping, cattle grazing and coffee production. Even the new decree law was not very effective because only two reserves were granted between 1957 and 1968 in the whole region of the Selva Central. In 1968 with the Velazco regime (1968 – 1975) things changed and a new decree, the Law of Native Communities, was passed in 1974 (Santos-Granero and Barclay 1998).

In the 50s and 60s indigenous people began to live in bigger units, giving up their traditional pattern of dispersed settlement. This trend is even more obvious between 1974 and 1988 as the number of settlements with a population over 100 people increased by 21 percent (Santos-Granero and Barclay 1998). The native communities especially in the Pacazu valley do not practice anymore their traditional way of living; they rather represent modern societies highly involved in the local economy.

Organizing themselves in movements and organizations they started to fight for autonomy and integration into the regional economy, not only as labor workers but also as producers. As a result the majority of indigenous settlements got legally registered during the Bermúdez legislation (1975 – 1980) so that by 1989 the vast majority of indigenous groups had been registered and titled (Moore 1994, Santos-Granero and Barclay 1998).

The use of community land is basically determined by the surrounding colonist activities. This circumstance developed as indigenous people worked for

colonists and little by little built up their own stock in the case of cattle ranching or they started cultivating their own coffee. The size of land that is granted to each community depends, as in the case of colonist land holdings, on the stage of colonization of the area and the land availability.

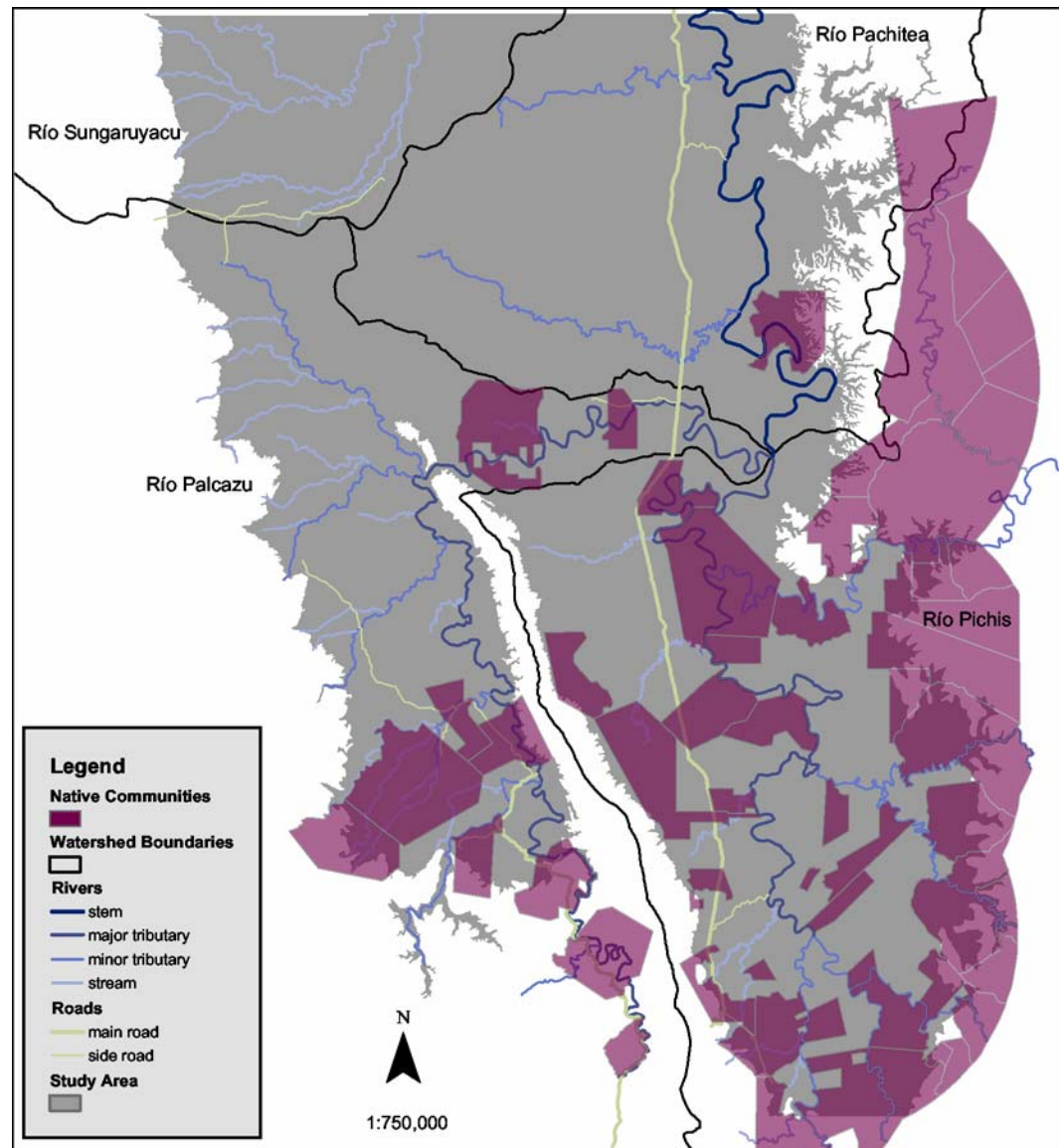


Figure 20: The legal boundaries of the Native Communities (Data Source: Courtesy of the Instituto del Bien Comun, Lima Peru) within the study area indicate that the Communities that are established in the Pichis Basin are more abundant and interconnected.

In rather remote, peripheral areas such as the Pichis Valley larger pieces of land were distributed (Fig.20) but generally the per capita land suitable for agriculture granted to indigenous communities is by far smaller than that given to colonists (Santos-Granero and Barclay 1998). This circumstance could force indigenous people to use inappropriate land for agriculture or pasture to an even higher extent than colonists. Studies show that the deforestation rates and consumption of natural resources on colonist managed land increases with abundance of land whereas in the case of indigenous people it seems that the more resources are available the more diversified are the uses (Bedoya 1996). The explanation for this difference is that colonists live almost exclusively on agriculture, cattle ranching, timber production and cash crops such as coffee and cacao for commercial purposes. All these land uses compete with the natural land cover. Native communities besides cultivating a great variety of vegetables in slash and burn farming supplement their diet with hunting, fishing and fruit collection (Ocaña 1990).

It is questionable if this observation can be made in the case of the peripheral areas of the Pachitea Basin for the given conditions. Nevertheless studies of the mid 70s indicate that all the native communities within the Palcazu basin participate in hunting and fishing activities and that more than 50% are also involved in gathering activities to supplement their diet (SINAMOS 1975 in (Santos-Granero and Barclay 1998)). A recent study in the Palcazu Valley on the other hand indicates that at least for that basin there are only minor differences between native communities and colonists when it comes to agricultural activities

and supplemental food supplies. Both groups participate in hunting and gathering as well as in subsistence agriculture and raising of livestock (Cossio-Solano 2001). It is therefore not expected to see a major difference in land-use patterns for the different ethnic backgrounds within a watershed. It is rather expected that the major differences depend on colonization pressure and land availability.

Land and Market Accessibility

Closely related to socioeconomic factors are infrastructure factors since infrastructure allows for participation in local regional or global markets. Roads and navigable rivers serve as transportation media for commodities; therefore, human economic activities are limited by the infrastructure of a region. Proximity to navigable rivers and roads has a significant impact on the ability to participate in trade. The closer a market opportunity the higher is the tendency to participate and increase agricultural output (Pfaff 1999). For this reason deforestation along roads and navigable rivers appears to be higher than in areas more distant from those features. Not only is it the accessibility to markets but also the over all accessibility of the land when considering the density of the forest. Penetration into dense forest is easier from cleared access points such as road edges and shores of rivers. A spatial proximity analysis can demonstrate to what extent infrastructure influences the quantity and spatial pattern of land-cover.

The Pachitea Basin saw an opening to the rest of the region by expanding the road system, which advanced from Southwest to Northeast. The Via Central,

a mule path that connected Puerto Tucker (Puerto Bermúdez) in the Pichis Valley with Tarma the main gate for trade with Lima, was established in 1898, but the remoteness of Puerto Tucker with long distances to markets prevented a colonist invasion into the Pichis Valley (Moore 1989).

Until the construction of roads that eventually connected the Pachitea with the rest of the Selva Central to the South, the whole basin was economically oriented towards the town of Pucallpa downstream to the North (Santos-Granero and Barclay 1998). The goods (mainly cattle) were transported by raft on the river, which was the only means of transportation at that time even though the southern part of the Selva Central was geographically a lot closer.

This situation changed however when the first road connecting Oxapampa to La Merced in the southern Selva Central was inaugurated in 1943. This road opened a faster export route of goods to the capital Lima (Fig.21). Further road construction in the Pozuzo Valley connected Huancabamba and Pozuzo in 1965 and 1975 respectively. When the road between Villa-Rica and Cacazú was opened in 1975, it was just a question of time before the Palcazu Valley would experience a major influx of colonization. It took another nine years to finish the road that connected Cacazú with Iscozacín. During the same time period the main road between Villa-Rica and Cacazú was extended to Puerto Bermúdez opening the third valley of the Pachitea basin the Pichis to economically viable export trade centers to the South in 1985 (Santos-Granero and Barclay 1998) (Fig.21).

It is expected that the progression of the colonization from West to East and the infrastructure expansion, which followed the same pattern, would be reflected in different land-use patterns for the different sub-basins of the Pachitea. With the accessibility of areas by road, large-scale timber production also became a lucrative business that first appeared in the area around Oxapampa, and then gradually expanded north along the Pozuzo following the advances of road construction. Oxapampa was for more than twenty years the major production center of tropical hardwood in Peru. Due to extensive logging timber extraction in the districts of Oxapampa and Villa-Rica has steadily declined since 1974 and 1980 respectively. Additional logging roads perpendicular to the main road encouraged agriculture and cattle ranching to enter newly opened space, which for several reasons is not adequate for those uses (Bedoya 1996).

This invasive development does not allow for secondary succession of the original land cover and even accelerates the clearing of virgin forest for pasture and agriculture. The same tendency then is expected for the other valleys of the Pachitea that were opened up recently. This discussion on the influence of infrastructure in general and in the case of the Pachitea leads to the assumption that the stage of deforestation in the Palcazu basin is assumed to be higher than that of the Pichis. The reason for this assumption is based on the long-term invasion of settlers from the North through the Pozuzo, who were highly involved in cattle ranching.

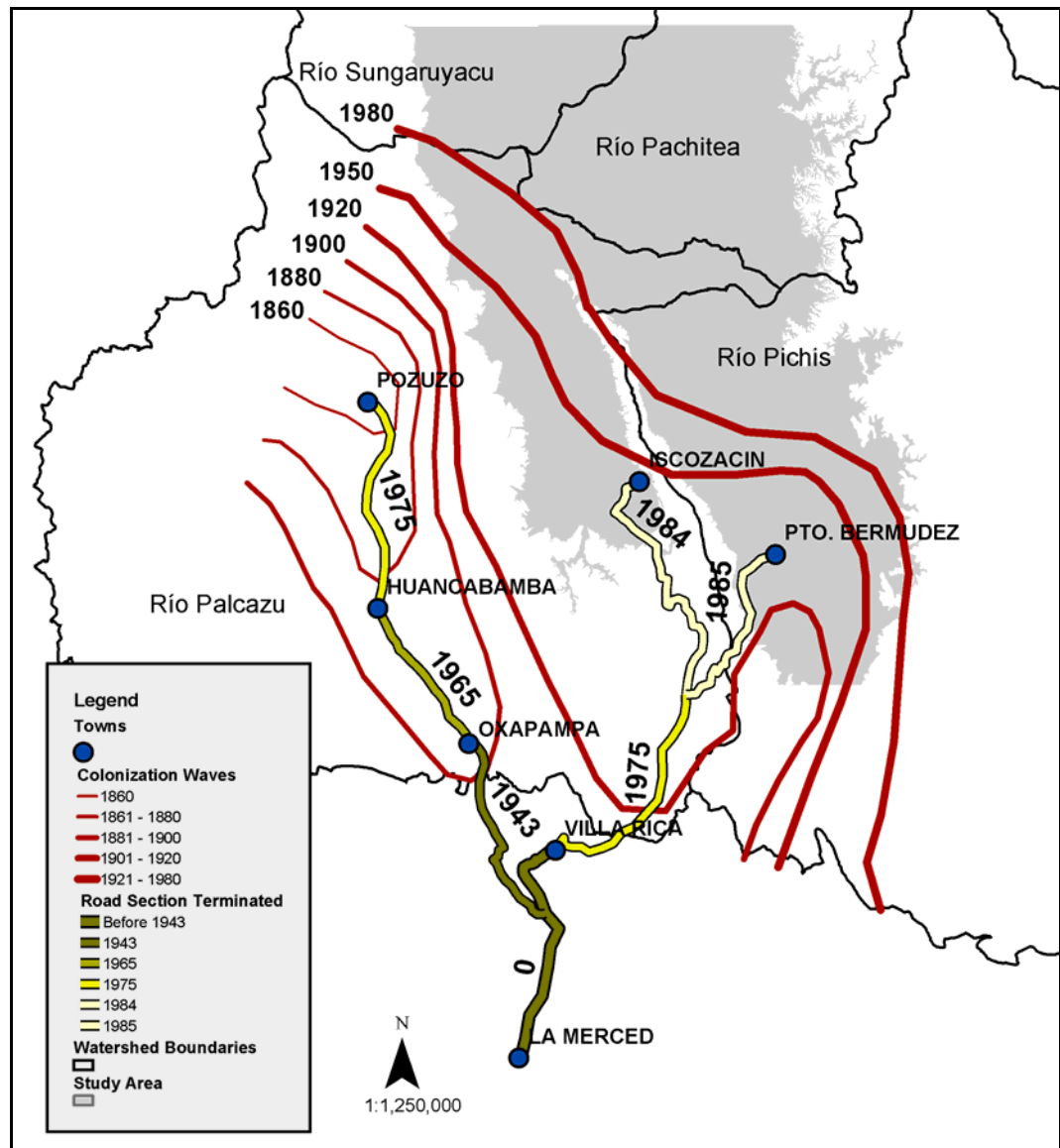


Figure 21: Colonization progression and infrastructure expansion in the Pachitea basin. The colonization advanced from the Northwest to the Southeast. The expansion of the road system entered the different basins from the South through the city of La Merced. Adapted from (Santos-Granero and Barclay 1998).

Summary of Research Questions

The historical development of land use and infrastructure as well as the present ethnic composition of the people in the Pachitea basin lead to the consideration of the following questions:

Land-Cover Abundance

What is the overall distribution of land cover and are there differences between watersheds?

What is the distribution of land use by watershed, by ethnicity, and by watershed and ethnicity?

What are the differences in the overall distribution of land cover for the different best land use categories?

Land Cover as a Function of Distance from Rivers and Roads

Are there differences in distribution of land cover as a function of distance from major roads and all roads, major rivers and all rivers and all rivers and roads?

Are abundance of land-cover type rather a function of suitability of the land or distance to infrastructure features?

Data and Methodology

Dependent Variables

The land-cover map in raster format (Chapter 2) served as the base for the analysis. The grid code for the final land-cover grid was:

No Data = 0; Closed-Canopy Forest = 1; Succession Forest = 2;

Purma = 3; Grassland = 4; Agriculture = 5; Bare Surface = 6; Water = 7;

Shadow = 8.

Independent Variables

The variables investigated for their usefulness in describing variability in land-use distribution and configuration within the study area are the different watersheds within the Pachitea Basin (as a coarse indicator for time of colonization and population pressure or availability of land), indigenous communities versus colonist managed land (as indicators for ethnic and cultural differences), bio-geophysical aspects (as indicators of suitability of the land for a variety of land-use activities and current patterns in terms of conflicting land uses), and distance from infrastructure features (as indicator of accessibility to land and markets).

Watershed Boundaries

The major watersheds that coincide with the study area are those of the Palcazu, Pichis, Pachitea, and Sungaruyacu rivers (Fig.3). Watersheds were

delineated from digitized elevation contours from national maps at a scale of 1:100,000 provided by the “Instituto Geográfico Nacional de Perú” (IGN).

Grid code for watersheds:

Palcazu = 1; Pichis = 2; Pachitea = 3; Sungaruyacu = 4.

Indigenous Communities

The legal boundaries for indigenous communities were provided by the Instituto del Bien Común, a non-governmental organization in Peru conducting surveys on boundaries of community land, population and land use within native communities (Fig.20). The vectors were converted to raster format using the following grid code:

No Data = 0; Colonists = 1; Native Communities = 2.

Bio-Geophysical entities

The best land use information was digitized from maps generated by “La Oficina Nacional de Evaluación de Recursos Naturales” (ONERN) at a scale of 1:1,000,000. A raster that includes areas of conflict, underutilization and appropriate use was generated, by applying the matrix given above (Table 7).to the best land use raster combined with the land-cover raster (Fig.19).

Grid codes for best land use:

No Data = 0; Forestry = 1; Pasture/Forestry = 2; Forestry/Protection = 3;

Agriculture/Pasture = 4; Protection/Forestry = 5; Protection = 6;

Pasture/Agriculture = 7; Agriculture = 8.

Grid code for conflict raster:

No Data = 0; Conflict = 1; Potential = 2; Appropriate = 3.

Infrastructure

Rivers and roads were taken into consideration in their contribution to land and market accessibility. River navigability was categorized into three categories. Rivers that carry enough water to allow for year round navigation consist of the Pachitea as the stem and the major tributaries, the Palcazu and Pichis Rivers. The minor tributaries are rivers that are only navigable by small boats with out-board engines and the third category includes small streams that can be navigated by canoes and rafts only. Roads were grouped into two categories. The Carretera Marginal de la Selva and the road connecting Iscozacín to Villa Rica are considered major roads; all other roads are minor roads (Fig.20).

Roads and Rivers were digitized from the geometrically corrected satellite images, converted to raster and combined with the output of the land-cover classification in order to get a continuous river system that includes edges of rivers. Distance grids to different combinations of river and road categories (roads only, rivers only, all roads and rivers, stem and major tributaries, main roads, and major rivers and main roads) were generated in a raster GIS. Mountain ranges pose natural limitations to the movement of people and goods therefore the calculation of distance to infrastructure features across mountains was prevented using a cost grid. The cost grid has a value of 1 for pixels on flat

terrain ($< 5\%$) and a value of 100 for pixels on slopes ($> 5\%$). The mask for the study area limitations was applied to the different distance grids (Fig.21)

For each pixel within the study area land-cover information was extracted along with distances to rivers and roads, best land use and conflict as indicator for the deviation from best management practices, sub-basins as parameter for time of colonization and population pressure, and native communities versus colonist land as management differences based on cultural differences. A comparison of land-cover distributions as a function of distance to infrastructure features helps to determine if infrastructure is an important factor in describing the spatial distribution of land-cover types for the study area. One way of comparing the distributions as function of distance is by the utilization of histograms. A comparison of histograms is not practical though because histograms depend very much on the width and end points of the bins and they are not very smooth.

Kernel Density Estimators (KDE) can be used instead in order to prevent end point binning artifacts and instead compare smooth curves to each other. A KDE is a statistical density estimator utilizing observed data to approximate a hypothesized density function. A KDE was used to estimate the density of number of pixels for distances to different infrastructure features and the land-cover density at those distances. The proportion of land-cover type to overall area at various distances can then be plotted in a smooth curve using a Spline

function to interpolate between the estimated values. Combining plots for different land-cover types then allows for direct comparison.

The method chosen to calculate the bandwidth of the KDE function is Silverman's Rule Of Thumb (SROT) (Equation 17)

$$h = 0.9 \min \left[\sigma, \frac{(Q_3 - Q_1)}{1.34} \right] n^{-\frac{1}{5}} \quad (17)$$

where

h is the bandwidth

σ is the sample standard deviation

Q_3, Q_1 the third and first sample quartiles, respectively.

For each distance raster the lower grid limit was set to 25 and the upper grid limit was calculated by rounding down to the next multiple of 50 and then subtracting 25 (grid size/ 2). The number of grid points was then calculated by dividing the upper grid limit by 50 (Table 8).

Table 8. Parameters for each distance raster. The grid limits are in meters.

Distance Raster	Lower Grid Limit	Upper Grid Limit	Number of Grid Points
All Rivers and Roads	25	27,225	545
Major Rivers and Roads	25	52,975	1,060
All Rivers	25	27,225	545
Major Rivers	25	54,625	1,093
All Roads	25	34,575	692
Major Roads	25	52,975	1,060

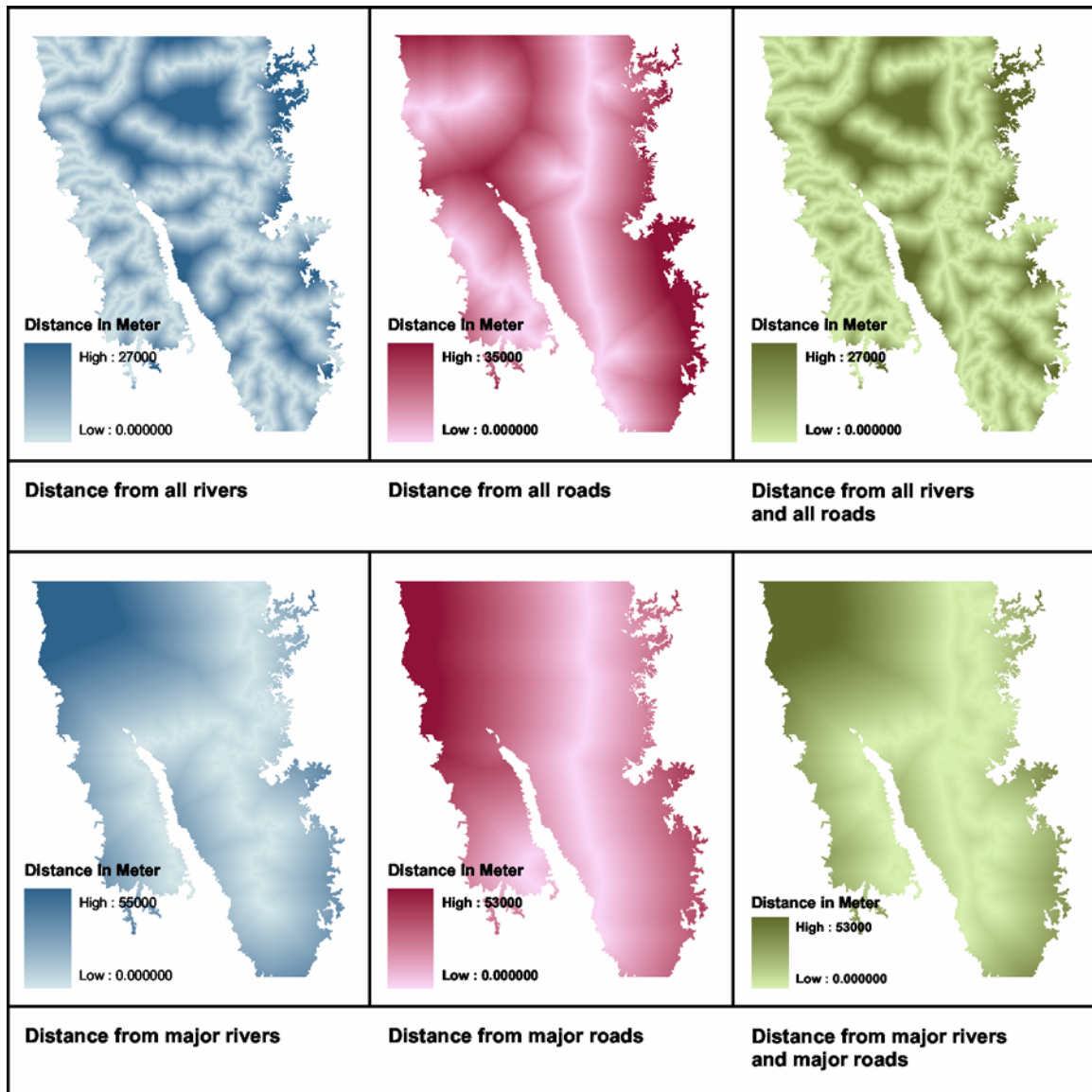


Figure 22. Distance raster for different combinations of infrastructure features. These grids were generated using a cost grid with values of 1 for slopes less than 5 degrees and 100 for slopes greater than 5 degrees.

Results and Discussion

The study area, which is defined by areas of elevation below 500 masl and slopes less than 5 degrees areas with less than 20 kilometers to major rivers

and roads, is a subset in the center of the Pachitea basin dissecting the major watersheds of the Palcazu, Pichis, Pachitea, and Sungaruyacu. The Pachitea basin has a size of approximately 29,027 square kilometers or 2,902,668 hectares of which 572,311 or about 20% are included in the study area. The area covered of each watershed amounts to between 16% of the Palcazu to 26% of the Sungaruyacu. (Fig.6, Table 9).

Table 9. Area covered by the major watersheds and the area of each watershed covered by the study area in hectares and percent

	Palcazu	Pichis	Pachitea	Sungaruyacu	TOTAL
Total area of the watershed	964,339	1,030,368	679,019	228,941	2,902,668
Area covered by study area	154,252	201,659	156,309	60,091	572,311
Percentage covered	16.00	19.57	23.02	26.25	19.72

Land-Cover Abundance within the Study Area

The land cover derived from the satellite images indicates that still about 55% or 316,880 hectares of the area remains as primary forest and together with secondary forest the forested areas amount to about 474,755 hectares or 83% (Table 10). As suggested by Santos Granero and Barclay the Selva Central does not follow the pattern of shifting agriculture, which is valid for a large part of the Amazon Basin. The typical shifting agriculture pattern shows a proportion of fallow land (purma) to that under actual cultivation of 3 to 1. The suggested ratio for the Selva Central is rather the inverse of 1 to 3 given an increased intensification (Santos-Granero and Barclay 1998).

Table 10. Land cover distribution within the study area in hectares and percent. No Data means that within the study area clouds or shadows covered these areas.

Land cover	Area	Percent
Closed-Canopy Forest	316,880	55.37
Succession Forest	157,876	27.59
Purma	22,828	3.99
Grassland	46,752	8.17
Agriculture	6,574	1.15
Bare Surface	4,127	0.72
Water	11,092	1.94
No Data	6,183	1.08
TOTAL	572,311	100.00

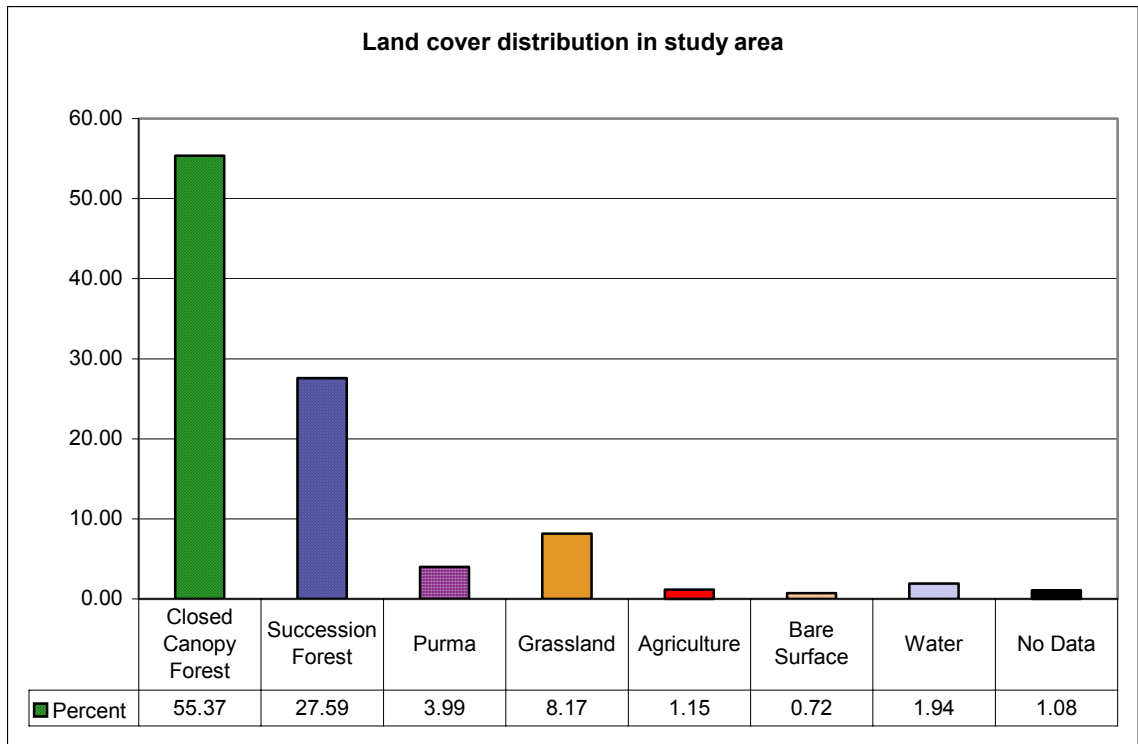


Figure 23. Land-cover distribution for the study area in percent. No Data indicates cloud or shadow cover on source data.

The absence of a regional or national market opportunity for agricultural products leads to the expectation that the majority of utilized land in the study area is pasture and to some degree purma and succession forest not particularly as a result of fallow agricultural even though included, but rather as an overgrowth of unused pasture or the result of some type of extraction forestry in the case of succession forest, which would open up the canopy.

Only a minor proportion of the used land is expected to be subsistence agriculture and the remainder of that is expected to generate commercial revenues. Therefore the ratios presented are not expected to hold true for the study area and indeed Purma is represented at 4% as an effect of transition from overgrowing grassland and agricultural plots. The areas used for agro-pastoral activities sum up to about 9.32% with grassland constituting the big part of 8.17% leaving 1.15% to farmland. Succession forest in an advanced stage is represented with 27.59%. Ratios of fallow to utilized land for the study area are 1 to 2.3 if considering only Purma. When including also succession forest the ratio changes to 3.4 to 1. This means that the actual ratio lies somewhere between 1 to 2.3 and 3.4 to 1.

Differences in Overall Distribution of Land Cover between Watersheds

I expected that the watershed of the Palcazu would have highest deforestation based on the early colonization pressure from the North, which increased when the infrastructure expanding from the south opened fast access

to other parts of the Selva Central and Lima first to this part of the eastern Pachitea.

The distribution of land cover types in the different watersheds indicates that as expected the Palcazu has the lowest percentage of closed canopy forest left (42%) compared to the Pichis and Pachitea (56% and 67% respectively) (Table 11, Fig. 24). The percentage of grassland in the Palcazu on the other hand is approximately 4 and 3 times higher than in the Pichis and Pachitea respectively. For agricultural land we can see a similar pattern with 1.63% in the Palcazu, 0.9% in the Pichis and only 0.88 in the Pachitea. The ratios of grassland to agriculture for the four watersheds are 9.4, 4.3, 6.1 and 8.0 for the Palcazu, Pichis, Pachitea and Sungaruyacu, respectively.

Table 11. Land cover by watershed within the study area in hectares and percent.

Land cover	Palcazu		Pichis		Pachitea		Sungaruyacu	
	Area	Percent	Area	Percent	Area	Percent	Area	Percent
Closed-Canopy Forest	65,208	42.27	113,871	56.47	105,177	67.29	32,624	54.29
Succession Forest	44,990	29.17	64,037	31.76	32,874	21.03	15,975	26.59
Purma	10,394	6.74	5,236	2.60	4,537	2.90	2,661	4.43
Grassland	23,687	15.36	7,763	3.85	8,378	5.36	6,923	11.52
Agriculture	2,522	1.63	1,817	0.90	1,372	0.88	863	1.44
Bare Surface	2,884	1.87	614	0.30	440	0.28	189	0.31
Water	3,709	2.40	3,803	1.89	2,731	1.75	849	1.41
No Data	858	0.56	4,518	2.24	801	0.51	6	0.01
TOTAL	154,252	100	201,659	100	156,310	100	60,090	100

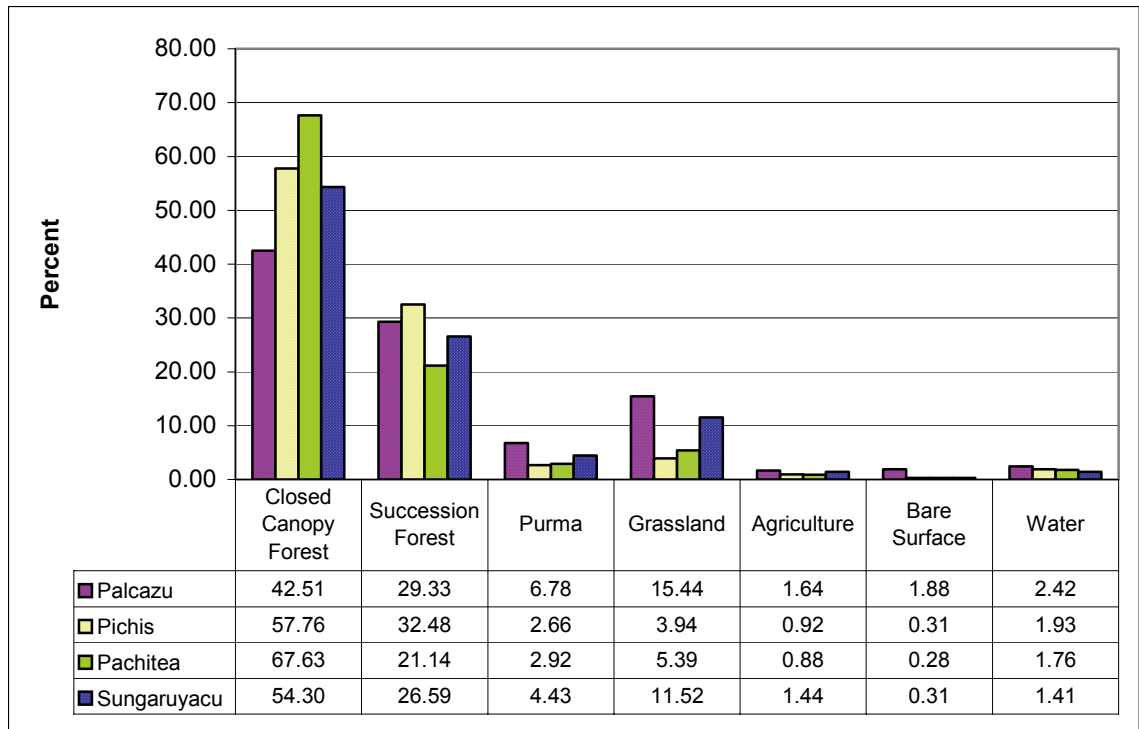


Figure 24. Land cover distribution by watershed in percent not including No Data values.

Differences in Overall Distribution of Land Cover between Native Communities and Colonists

I expected that the land use between colonists and native communities on one hand is expected to differ due to the given reasons of cultural background and diet. On the other hand the adaptation of natives to the changing circumstances brought about by the colonists might be an important factor that will weaken the former expectation. The summary in Table 12 indicates that closed-canopy forest remaining in the case of colonist as well as native communities is about 58% in either case. Differences can be seen in the categories succession forest 34% and 27%, Agriculture 0.78% and 1.3%, and

grassland with 4.1% and 9.61% respectively. The differences are obvious in the categories grassland and agriculture only. But these differences could be due to the different patterns in the watersheds.

Table 12. Land-cover distribution by type of management in hectares and percent.

Land Cover	Native		Colonist	
	Area	Percent	Area	Percent
Closed-Canopy Forest	64,868	58.12	252,011	57.37
Succession Forest	37,458	33.56	120,418	27.41
Purma	3,867	3.47	18,960	4.32
Grassland	4,547	4.07	42,204	9.61
Agriculture	868	0.78	5,706	1.30
TOTAL	111,609	100.00	439,300	100.00

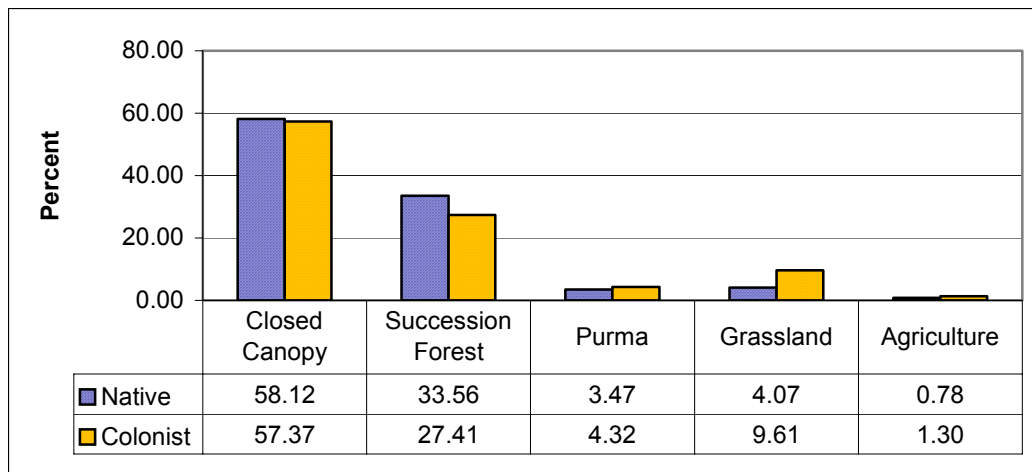


Figure 25. Land-cover distribution of native communities versus colonist managed land in percent.

The comparison of land cover distribution between native communities and colonists by watershed (Table13, Fig.26) in fact indicates that there is more

similarity between colonists and natives within a watershed than for natives between watersheds or colonists between watersheds.

Table 13. Land-cover distribution by watershed and native communities versus colonists in hectares and percent.

	Palcazu				Pichis			
	Native		Colonist		Native		Colonist	
	Area	%	Area	%	Area	%	Area	%
Closed-Canopy Forest	12,915	47.69	52,293	43.84	49,562	61.41	64,309	57.41
Succession Forest	10,105	37.31	34,885	29.25	26,212	32.48	37,824	33.77
Purma	1,607	5.94	8,787	7.37	2,153	2.67	3,083	2.75
Grassland	2,188	8.08	21,499	18.03	2,203	2.73	5,561	4.96
Agriculture	265	0.98	2,257	1.89	576	0.71	1,241	1.11
TOTAL	27,080	100.00	119,721	100.38	80,706	100.00	112,018	100.00

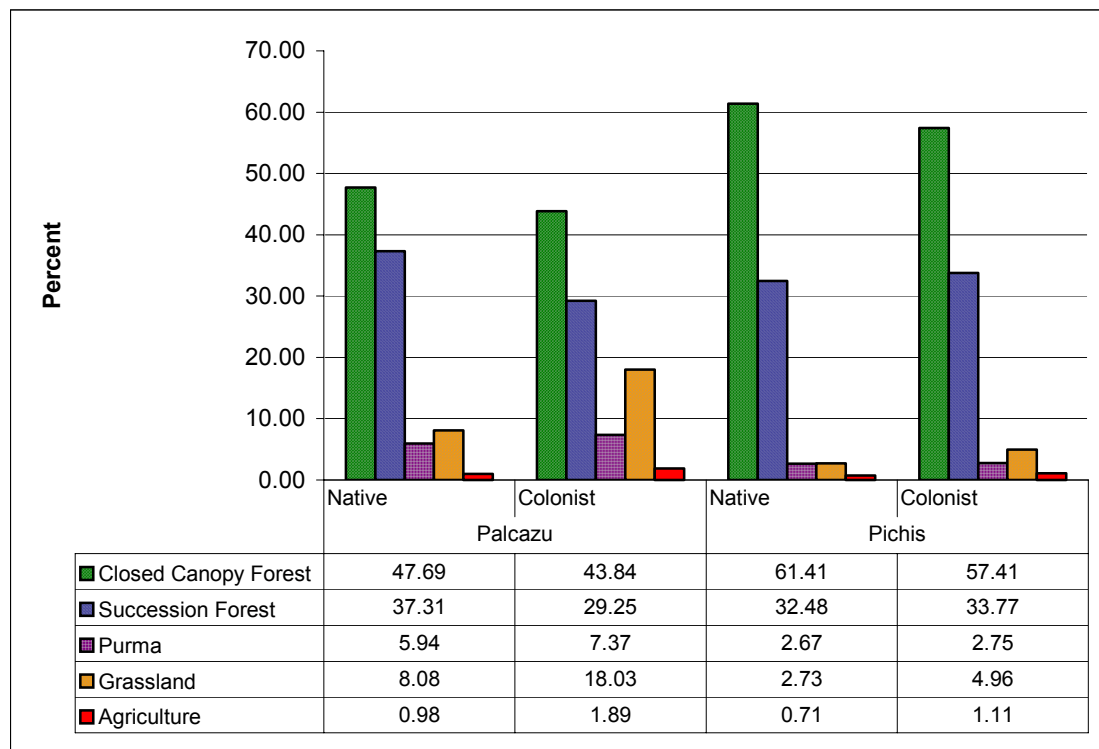


Figure 26. Comparison of land cover distribution between native communities and colonists by watershed in percent.

The differences between the watersheds seem to be more predominant rather than the differences between management type native versus colonist. This seems reasonable given the historical development of the current land tenure system and the vivid interaction between both groups. Since both groups participate in local market activities current pressure on available land and the time of colonization seem to be better predictors for land use abundance.

Distribution of Land Suitable for Agriculture, Pasture, Forestry and Protection by Watershed and Management Type

Is the land suitable for agricultural uses evenly distributed between watersheds and more important between native communities and colonists? Only about 1.71% of the study area requires absolute protection, whereas 51% and 47% are categorized for forestry and agro-pastoral uses respectively (Table 14). The watershed that allows for least agricultural pursuits and demands the highest percentage of protection is that of the Pichis followed by the Pachitea, Palcazu and Sungaruyacu.

Table 14. Best land use (BLU) suggestions for the entire study area by watershed. The categories were summarized according to the conflict matrix (Table 7).

BLU	Palcazu		Pichis		Pachitea		Sungaruyacu		TOTAL	
	Area	%	Area	%	Area	%	Area	%	Area	%
Protection	2,848	1.85	6,713	3.33	225	0.14	0	0.00	9,786	1.71
Forestry	55,262	35.83	159,020	78.82	71,302	45.62	7,403	12.34	292,987	51.19
Agro/Past	96,142	62.33	36,021	17.85	84,780	54.24	52,593	87.66	269,536	47.10
TOTAL	154,252	100.00	201,754	100.00	156,307	100.00	59,996	100.00	572,309	100.00

The percentage of land to be protected for native communities (3%) versus colonist-managed land (1.4%) is more than twice as high. Land suitable for agriculture and pasture in native communities accounts for about 29%, whereas for colonist land that category makes almost 52% (Table 15). This circumstance suggests that the land in native communities might be used inappropriately to a larger extent than outside the communities due to land scarcity.

Table 15. Best land use suggestions by management type (native within community boundaries versus colonist. The categories were summarized according to the conflict matrix (Table 7).

Best land use	Native		Colonist		TOTAL	
	Area	Percent	Area	Percent	Area	Percent
Protection	3,457	2.95	6,330	1.39	9,787	1.71
Forestry	79,977	68.21	213,010	46.81	292,987	51.19
Agro/Past	33,819	28.84	235,716	51.80	269,535	47.10
TOTAL	117,253	100.00	455,056	100.00	572,309	100.00

Pattern of Actual Land Cover versus Recommended Best Land Use Practices by Watershed and Management Type

For the entire study area Table 16 shows that within the category best suited for protection almost 10% of the land is used for agriculture and cattle ranching whereas in the categories suitable for agriculture and pasture, pasture and agriculture and agriculture only, still roughly 58%, 52% and 15% of the land

remains closed-canopy forest. Together with succession forest these numbers increase to 81%, 83% and 36% respectively.

When grouping the different conflict categories and underutilized categories according to Table 7, the distribution of the new categories “conflict”, “potential” and “appropriate” use is 3.5%, 38.7% and 57.8% respectively (Table 17; Fig.19).

Table 16. Actual land cover by best land use category in hectares and percent. Where Prot – Protection; For – Forestry; Agri – Agriculture; Pas – Pasture. The Category mentioned first being the more suitable. The grey fields in the upper part of the table indicate land cover conflicting with best land use suggestions, in the lower part they indicate potential for future appropriate use in agricultural and pastoral development.

Land cover	Prot		Prot - For		For - Prot		For	
	Area	Percent	Area	Percent	Area	Percent	Area	Percent
Closed-Canopy Forest	3,733	39.42	41,573	70.86	36,680	63.21	96,809	58.73
Succession Forest	4,806	50.75	14,125	24.07	15,825	27.27	48,292	29.30
Purma	464	4.90	1,258	2.14	2,224	3.83	4,664	2.83
Grassland	407	4.30	1379	2.35	2,725	4.70	13,160	7.98
Agriculture	60	0.63	336	0.57	579	1.00	1,910	1.16
TOTAL	9,470	100.00	58,671	100.00	58,033	100.00	164,835	100.00

Land cover	Agri - Pas		Pas - Agri		Agri	
	Area	Percent	Area	Percent	Area	Percent
Closed-Canopy Forest	70,133	55.26	66,885	52.75	1,067	17.22
Succession Forest	33,904	26.71	38,874	30.66	2,050	33.09
Purma	5,984	4.71	7,363	5.81	870	14.04
Grassland	15,014	11.83	12,067	9.52	2,000	32.28
Agriculture	1,883	1.48	1,597	1.26	208	3.36
TOTAL	126,918	100.00	126,786	100.00	6,195	100.00

Table 17. Distribution of land-use conflicts and potential for the entire study area in hectares and percent (grouped according to Table 7.

Category	Area	Percent
Conflict	19,479	3.54
Potential	212,906	38.65
Appropriate	318,507	57.82
TOTAL	550,892	100.00

Taking a look at the differences between watersheds we expect that, since in the Pichis valley such a small percent of the land is suitable for agriculture and pasture, the land is either used inappropriately or that in fact, if the best land use is a good predictor for land use, a high percentage of the suitable land is left for future potential use. For two reasons the Sungaruyacu is expected to have the smallest percentage for areas of conflict and high potential for further appropriate development open. First of all the area is categorized as frontier expansion, which means low population pressure and the major activity being forestry, and secondly no land falls in the category protection but 88% qualify for agriculture and pasture and the remaining 12% qualify for forestry (Table 14).

The actual pattern is presented in Table 18. In the Sungaruyacu as expected the areas of conflict are 0.3% and 70.3% are not used yet. The areas of conflict and potential in the Pichis amount to 3.2% and 14.9% respectively, which indicates that the land is used close to optimal with low conflict but also little room for future development. The Palcazu on the other hand shows a pattern that is unexpected. With 7.6% of the land being used inappropriately at the present, 47.1% remain underutilized. The Pachitea has room for future growth (48.3%) and only 1.4% are used unsuitably. Especially in the case of the Palcazu it is evident that the best land use suggestion is not necessarily a good predictor for land use.

Table 18. Distribution of land-use conflicts and potential by watershed in hectares and percent (grouped according to Table 7).

	Palcazu		Pichis		Pachitea		Sungaruyacu	
	Area	Percent	Area	Percent	Area	Percent	Area	Percent
Conflict	11,079	7.55	6,171	3.20	2,052	1.35	179	0.30
Potential	69,206	47.14	28,686	14.88	73,505	48.25	41,516	70.31
Appropriate	66,516	45.31	157,868	81.91	76,782	50.40	17,352	29.39
TOTAL	146,801	100.00	192,724	100.00	152,338	100.00	59,046	100.00

Further considering the differences between colonists and native communities reveals that unexpectedly the percentage of land used inappropriately in native communities is lower than that for colonist managed land, despite the fact that the percentage of land suitable for agro-pastoral uses is 29% for native communities and 52% for colonists (Table 19). The case of the colonists is similar to that of the Palcazu valley in that even though about 42% of the land is still underutilized, 3.7% are used inappropriately (Table 19).

Table 19. Distribution of land-use conflicts and potential by management type in hectares and percent (grouped according to Table 7).

	Native		Colonist	
	Area	Percent	Area	Percent
Conflict	3,305	2.96	16,175	3.68
Potential	28,503	25.54	184,409	41.98
Appropriate	79,801	71.50	238,715	54.34
TOTAL	111,609	100.00	439,300	100.00

The reason for this seemingly contradictory pattern can be further analyzed adding a factor that was left out so far; proximity to infrastructure features such as navigable rivers and roads.

Land-Cover Distribution in Relation to Rivers and Roads

Based on the assumption that proximity to rivers and roads determines accessibility to land and market opportunities and therefore has an impact on the land use, we expect that with increasing distance from rivers and roads the proportion of closed canopy forest cover increases and that of grassland and especially that of agriculture decreases. Comparing the distribution of land cover proportions at different distances to navigable rivers and roads gives an idea of how the distance to these infrastructure features influences the land use. If the variable under consideration does not have an effect on the distribution of land-cover we expect equal distributions of proportions at any distance to that feature. The proportion of a certain land-cover type is then expected to be the same as the overall proportion for that cover type for the entire study area (Fig.27). Proportion graphs were generated using KDEs for distances at 50 meter increments and the corresponding land-cover densities at the same distances.

The proportion of each land cover was calculated by dividing the overall density of pixels at a certain distance by the density of each land-cover type at that distance. A spline interpolation was used to interpolate values that were not estimated by the KDE. At each given point along the distance gradient all land-cover proportions together add up to 1 (Fig.28). For each distance, at which density estimations were generated, the area that is covered by a certain land-cover type was calculated as a percentage to the total area of that land-cover type in the study area.

Summing all percentages below each distance for all land-cover types lead to the cumulative percentage at each distance.

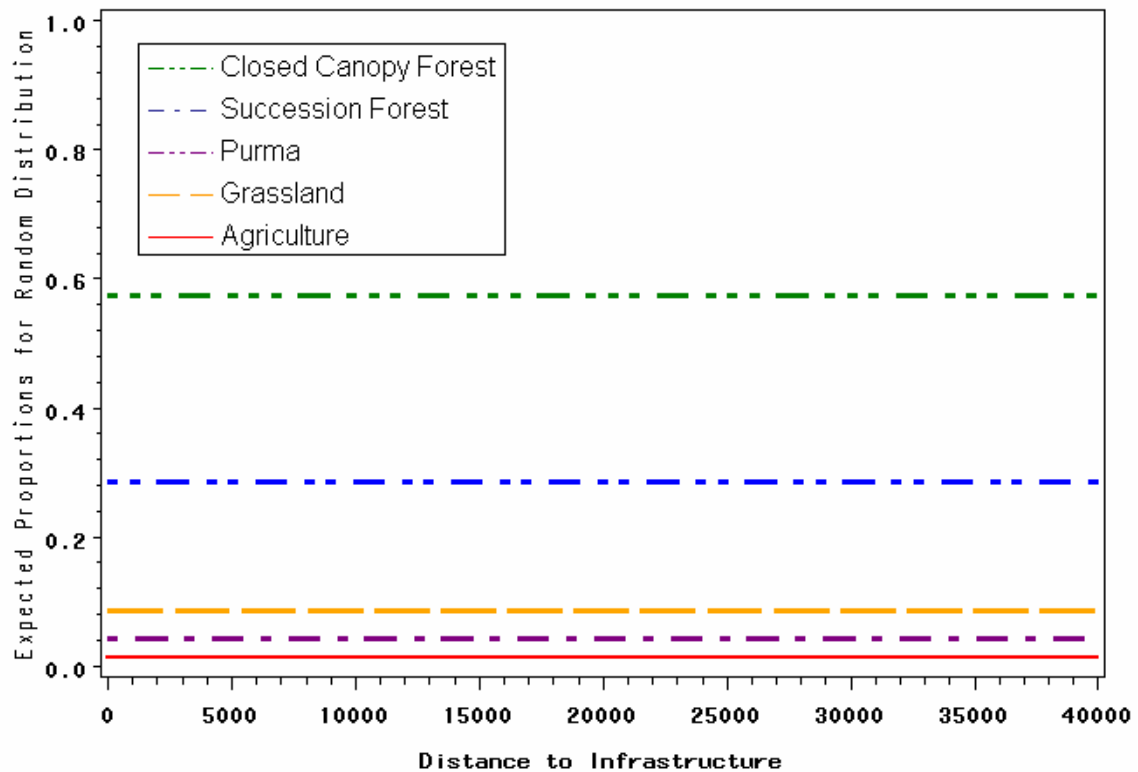


Figure 27. Expected proportions for random distribution of land-cover types as a function of distance to infrastructure features. The proportion of each land-cover is expected to be the same as the overall proportion for the entire study area. Close Canopy Forest at 0.58 (Green), Succession Forest at 0.29 (Blue), Purma at 0.04 (Yellow), Grassland at 0.08 (Orange), Agriculture at 0.01 (Red).

From the cumulative percentage plots a threshold of 95% for agriculture, grassland and purma was determined at 7,075 meters (Fig.29). This distance threshold was applied to the proportion plots, since at further distances a lot of noise is expected due to a small number of pixels. Further, in order to be able to compare the effects of infrastructure features on each of the different land cover

type the deviation from the expected proportion for random distribution was calculated. The percentage of deviation normalizes the distribution.

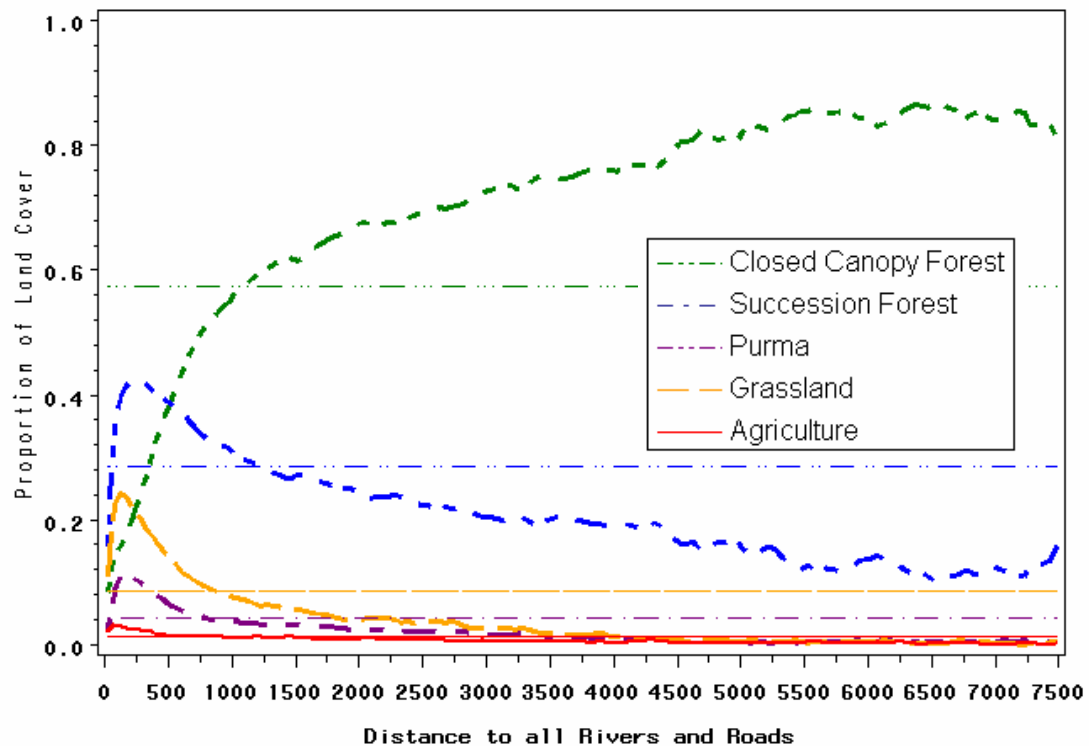


Figure 28. Expected proportions for random distribution (straight lines) and actual proportions of all land-cover types as a function of distance to all rivers and roads. Close-Canopy Forest (Green), Succession Forest (Blue), Purma (Purple), Grassland (Orange), Agriculture (Red).

The graph in Figure 30 shows the percent of deviation for all rivers and roads. It clearly indicates the expected trend, that agriculture and grassland are over proportionally represented at distances of up to about 1,000 meters to rivers and roads. Just the opposite trend is obvious for closed-canopy forest. A

comparison of deviation graphs for different infrastructure combinations shows to what extend each feature can be associated with the distribution of land cover.

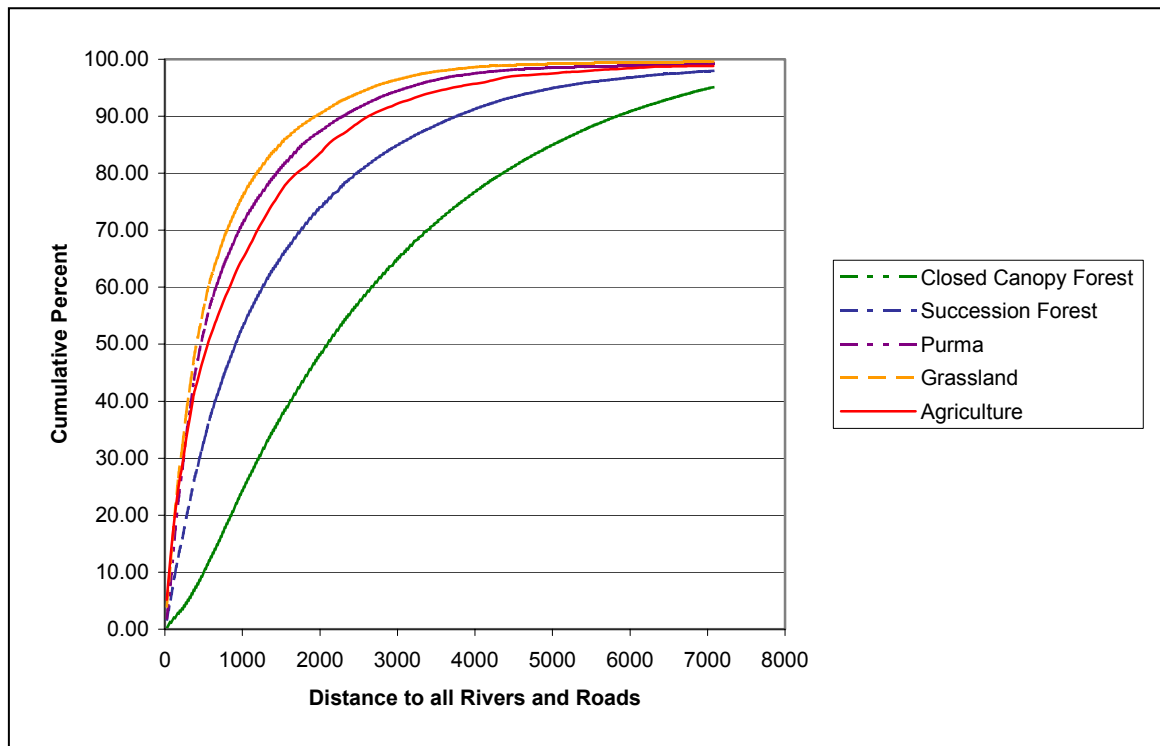


Figure 29. Cumulative percentage for land-cover distributions as a function of distance to all rivers and roads. Threshold is set to 95% of Agriculture, Grassland and Purma, which is at 7,075 meters to rivers and roads.

The various deviation graphs in Figure 31 clearly demonstrate that infrastructure features such as major rivers and major rivers by themselves are not sufficient enough to explain the existing distribution. A curve that shows no inflection points would be expected if the feature were highly correlated to the distribution. On the other hand a curve that changes direction many times and even crosses the line of the expected random distribution at several distances implies that there are other factors that need to be considered to explain the

distribution of land cover. This is the case for each, major roads and major rivers only.

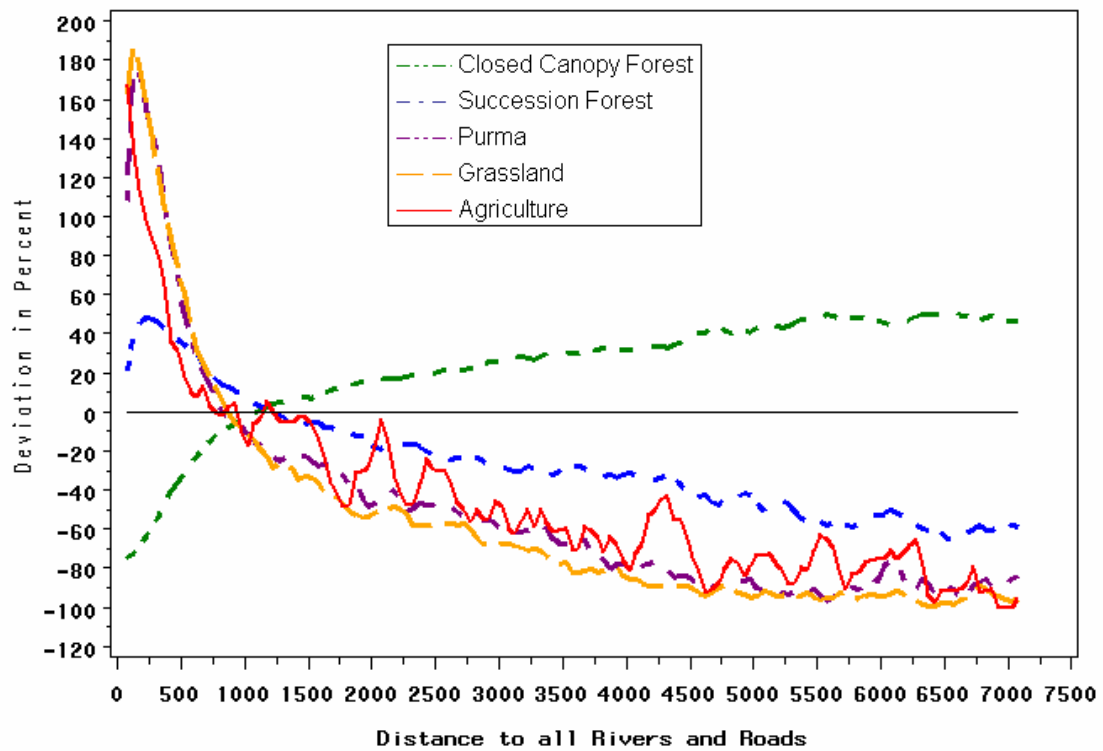


Figure 30. Deviation from expected random distribution of land-cover categories in percent as a function of distance to rivers and roads in meters.

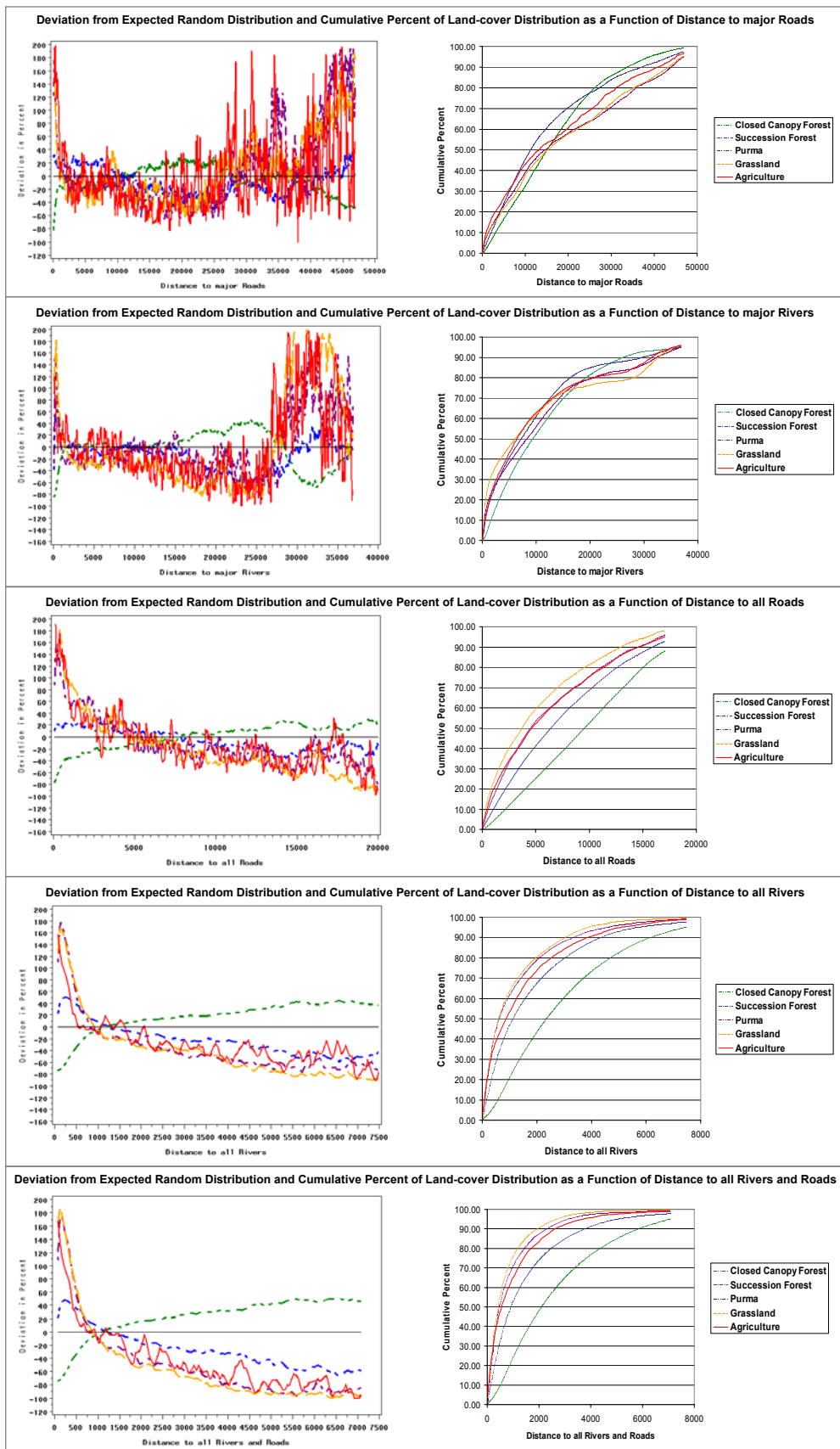


Figure 31. Left: Deviation from expected random distribution of land-cover proportions in percent as a function of distance to different combinations of infrastructure features in meters. Right: Cumulative percent of land-cover distribution as a function of distance to infrastructure in meters.

Distribution of Conflicting and Potential Land Uses as a Function of Distance to Rivers and Roads

As described, land use in the study area is not optimal at present. A high percentage of land is used inappropriately. Conflict areas indicate that the best land use recommendations are not followed even though land that qualifies for agriculture and pasture is left unused. Since distance to rivers and roads seems to play an important role in land-use choices, it is of interest to see what the distribution of underutilized land looks as a proportion of the area that has potential for agro-pastoral activities. On the other hand distribution patterns of land used inappropriately as a proportion of land suitable only for forestry and protection can help to determine whether the distance to rivers and roads is a factor that explains the observation that unsuitable land is used when potentially good land is still underutilized. If distance to rivers and roads is only a minor factor the distribution of both proportion distributions is expected to be relatively constant along a distance gradient from infrastructure features. The KDE proportion plots in Figure 32 indicate that there is a clear trend especially for distances below 1.5 to 2 kilometers. Inappropriate land use as a proportion of potential conflict areas is over represented at close distances and the opposite is true for underutilized areas as a proportion to potential development. Best land use classifications therefore seem not to be a good predictors for people's choices of land use it is rather again distance to rivers and roads that seems to play a more important role.

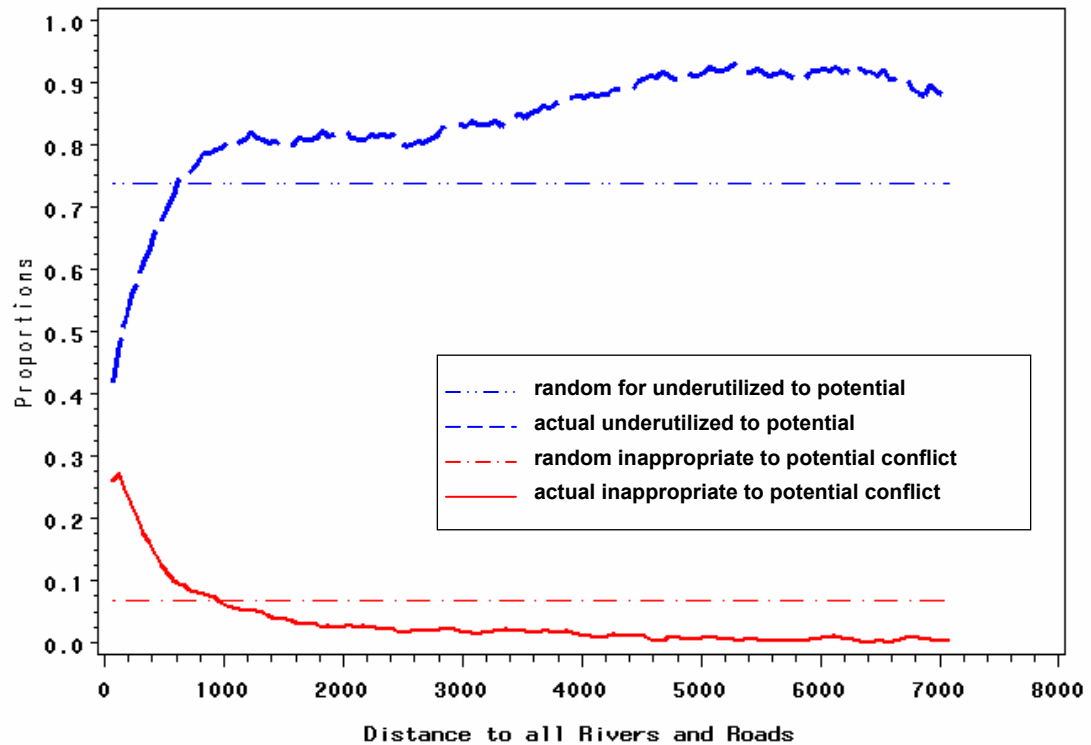


Figure 32. Expected proportions for random distribution and actual proportions of (1) inappropriate use of best land use categories forestry and protection (Red) and (2) potential versus underutilized land (Blue).

Conclusion

The actual use of the land is driven mainly by the ability to access the land and markets as is obvious in the distribution of land cover as a function of distance to a combination of all rivers and roads. Rivers as well as roads by themselves do not provide enough support to explain land cover distribution. When considering all rivers and roads a threshold distance at which disproportional agro-pastoral land cover switches from over represented to under represented is at about 1km. Best land use suggestions seem not to affect the

choice of land use. Differences in abundance of land cover between watersheds are more prevailing than differences between colonist and indigenous groups.

REFERENCES

- Bedoya, E. 1995. The Social and Economic Causes of Deforestation in the Peruvian Amazon Basin: Natives and Colonists. Pages 217-246 *in* M. Painter and W. H. Durham, editors. The Social Causes of Environmental Destruction in Latin America. The University of Michigan Press, Ann Arbor.
- Bedoya, E. 1996. Forty Years of Political Ecology in the Peruvian Upper Forest: The Case of Upper Huallaga. Pages 165 -186 *in* T. N. H. a. R. C. B. L. E. Sponsel, editor. Tropical Deforestation The Human Dimension. Columbia University Press, New York.
- Bilsborrow, R. E., and H. W. O. Okoth Ogendo. 1992. Population-driven Changes in Land Use in Developing Countries. *Ambio* 21:37-44.
- Brondizio, E. S., E. F. Moran, P. Mausel, and Y. Wu. 1994. Land use change in the Amazon Estuary: Patterns of Caboclo settlement and landscape management. *Human Ecology* 22:249-278.
- Chavez, P. S. 1988. An improved Dark-Object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment* 24:459 - 479.
- Chavez, P. S. 1996. Image-based atmospheric corrections - revisited and improved. *Photogrammetric Engineering & Remote Sensing* 62:1025-1036.
- Colby, J. D. 1991. Topographic normalization in rugged terrain. *Photogrammetric Engineering & Remote Sensing* 57:531-537.
- Colby, J. D. 1998. Land cover classification using Landsat TM imagery in the tropical highlands: the influence of anisotropic reflectance. *International Journal of Remote Sensing* 19:1479-1500.
- Congalton, R. G., and K. Green. 1957. Assessing the accuracy of remotely sensed data: principles and practices. Pages 137 *in*. Lewis Publishers, Boca Raton, Florida 33431.
- Cossio-Solano, R. E. 2001. Land Use of Riparian Zones in Two Communities in the Palcazu Basin, Central Andean Amazon, Peru. Master of Science. Florida International University, Maimi, Florida.
- de Koning, G. H. J., P. H. Verburg, A. Veldkamp, and L. O. Fresco. 1999. Multi-scale modelling of land use change dynamics in Ecuador. *Agricultural Systems* 61:77-93.

- Forman, R. T. T. 1995. Land mosaics : the ecology of landscapes and regions. Cambridge University Press, Cambridge, New York.
- Geoghegan, J. 1998. "Socializing the Pixel" and "Pixelizing the Social" in Land-Use and Land-Cover Change. Pages 51-69 *in* D. Liverman, Emilio F. Moran, Ronald R. Rindfuss, Paul C. Stern, editor. People and Pixels: Linking Remote Sensing and Social Science. National Academy Press, Washington, D.C.
- Hall, F. G., D. E. Strebel, J. E. Nickeson, and S. J. Goetz. 1991. Radiometric rectification: toward a common radiometric response among multitemporal, multisensor images. *Remote Sensing of Environment* 35:11-27.
- Herman, B. M., and S. R. Browning. 1965. A numerical solution to the equation of radiative transfer. *Journal of the Atmospheric Sciences* 22:59-66.
- Holdridge, L. R. 1967. Life zone ecology. Tropical Science Center, San Jose, Costa Rica.
- Hughes, R. F. 2000. Ecosystem-Scale Impacts of Deforestation and Land Use in a Humid Tropical Region of Mexico. *Ecological Applications* 10:515-527.
- Jensen, J. R. 1996. Introductory digital image processing: a remote sensing perspective, 2nd edition. Prentice Hall, Upper Saddle River, NJ 07459.
- Johnson, L. B. 1990. Analyzing spatial and temporal phenomena using geographical information systems. *Landscape Ecology* 4:31-43.
- Kammerbauer, J., and C. Ardon. 1999. Land use dynamics and landscape change pattern in a typical watershed in the hillside region of central Honduras. *Agriculture, Ecosystems and Environment* 75:93-100.
- Kneizys, F. X., E. P. Shettle, and L. W. Abreu. 1988. Users guide to LOWTRAN 7.
- Krysanova, V. 1998. Development and test of a spatially distributed hydrological/water quality model for mesoscale watersheds. *Ecological Modelling* 106:261-289.
- Lambin, E. F., X. Baulies, N. Bockstael, G. Fischer, T. Krug, R. Leemans, E. F. Moran, R. R. Rindfuss, Y. Sato, D. Skole, B. L. Turner II, and C. Vogel. 1999. IGBP Report No. 48 and IHDP Report No. 10 Land-Use and Land-Cover Change Implementation Strategy. LUC-CIPO, Bonn.

- MacArthur, R. H., and E. O. Wilson. 1961. The theory of island biogeography. Princeton University Press, Princeton.
- Markham, B. L., and J. L. Barker. 1986. Landsat MSS and TM Post-Calibration Dynamic Ranges, Exoatmospheric Reflectances and At-Satellite Temperatures. EOSAT Technical Notes.
- Mena, C. F. 2001. Deforestation in the Napo Basin: Socioeconomic factors, spatial patterns, and metrics. Master of Science. Florida International University, Miami, Florida.
- Moore, M. M. 1994. Analysis and Mapping of Late-Successional Forests in the Americal Southwest. Pages 108-110 *in* V. A. Sample, editor. Remote Sensing and GIS in Ecosystem Management. Island Press, Washington, D.C.
- Moore, T. 1989. The Social Dimension in the Palcazu: A Pre-Evaluation of the Central Selva Resource Management Projec. Ronco Consulting Corporation, Washington DC.
- Moran, S. M., R. D. Jackson, P. N. Slater, and P. M. Teillet. 1992. Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output. Remote Sensing of Environment 41:169-184.
- NASA. Landsat 7 Science Data Users Handbook. *in*.
- Novo, E. M. 1997. Identification and mapping of the Amazon habitats using a mixing model. International Journal of Remote Sensing 18:663-670.
- Ocaña, J. 1990. Manejo de Bosques Naturales de la Selva Alta de Peru. Un estudio de Caso del Valle del Palcazu. Centro Cientifico Tropical. Ronco Consulting Corporation, Fundacion Peruana Para La Conservacion De La Naturaleza, Lima, Peru.
- O'Neill, R. V., J. R. Krummel, R. H. Gardner, G. Sugihara, B. Jackson, D. L. DeAngelis, B. T. Milne, M. G. Turner, B. Zygmunt, S. W. Christensen, V. H. Dale, and R. L. Graham. 1988. Indices of landscape pattern. Landscape Ecology 1:153-162.
- Pastor, J. 1992. Using Simulation Models and Geographic Information Systems to Integrate Ecosystem and Landscape Ecology. Pages 324-346 *in* R. J. Naiman, editor. Watershed Management Balancing Sustainability of Environmental Change. Springer Verlag, New York.

- Pearson, S. M. 1994. Ecological Perspective: Understanding the Impacts of Forest Fragmentation. Pages 178-191 *in* V. A. Sample, editor. Remote Sensing and GIS in Ecosystem Management. Island Press, Washington, D.C.
- Pfaff, A. S. P. 1999. What Drives Deforestation in the Brazilian Amazon? *Journal of Environmental Economics and Management* 37:26-43.
- Rindfuss, R. R., and P. C. Stern. 1998. Linking Remote Sensing and Social Science: The Need and the Challenges. Pages 1-27 *in* D. Liverman, Emilio F. Moran, Ronald R. Rindfuss, Paul C. Stern, editor. People and Pixels: Linking Remote Sensing and Social Science. National Academy Press, Washington, D.C.
- Saatchi, R. Nelson, Podest, and Holt. 2000. Mapping land cover types in the Amazon Basin using 1 km JERS-1 mosaic. *International Journal of Remote Sensing* 21:1201-1234.
- Santos-Granero, F., and F. Barclay. 1998. Selva Central: history, economy, and land use in Peruvian Amazonia. Smithsonian Institution Press, Washington, DC.
- Schott, J. R., C. Salvaggio, and W. K. Volchok. 1988. Radiometric scene normalization using pseudoinvariant features. *Remote Sensing of Environment* 26:1-16.
- Smith, J., T. Lin, and K. Ranson. 1980. The lambertian assumption and landsat data. *Photogrammetric Engineering & Remote Sensing* 46:1183-1189.
- Su, Z. 2000. Remote sensing of land use and vegetation for mesoscale hydrological studies. *International Journal of Remote Sensing* 21:213-233.
- Swain, P. H., and S. M. Davis. 1978. Remote Sensing: The quantitative approach., McGraw-Hill, New York.
- Tanré, D., C. Deroo, P. Duhaut, M. Herman, and J. J. Mockette. 1990. Description of a computer code to simulate the satellite signal in the solar spectrum: the 5S Code. *International Journal of Remote Sensing* 2:659-668.
- Teillet, P. M., and S. R. P. 1991. Terrain elevation and sensor altitude dependence in a semi-analytical atmosphere code. *Canadian Journal of Remote Sensing* 17:36-44.
- Turner, M. G. 1990. Spatial and temporal analysis of landscape patterns. *Landscape Ecology* 4:21-30.

- Verburg, P. H., G. H. J. de Koning, K. Kok, A. Veldkamp, and J. Bouma. 1999. A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. *Ecological Modelling* 116:45-61.
- Vink, A. P. A. 1975. Land use in advancing agriculture. Springer Verlag, Berlin.
- Wood, C. H., and D. Skole. 1998. Linking Satellite, Census, and Survey Data to Study Deforestation in the Brazilian Amazon. Pages 70-93 *in* D. Liverman, Emilio F. Moran, Ronald R. Rindfuss, Paul C. Stern, editor. *People and Pixels: Linking Remote Sensing and Social Science*. National Academy Press, Washington, D.C.
- Yuan, D., and C. D. Elvidge. 1996. Comparison of relative radiometric normalization techniques. *ISPRS Journal of Photogrammetry and Remote Sensing* 51:117-126.

APPENDIX A

Regression Results

Model: MODEL1

Dependent Variable: B1_767

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.01451	0.01451	2388.92	<.0001
Error	12	0.00007287	0.00000607		
Corrected Total	13	0.01458			

Root MSE	0.00246	R-Square	0.9950
Dependent Mean	0.12929	Adj R-Sq	0.9946
Coeff Var	1.90599		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.01934	0.00234	8.25	<.0001
B1_667	1	0.83333	0.01705	48.88	<.0001

The REG Procedure

Model: MODEL1

Dependent Variable: B2_767

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.02378	0.02378	3213.04	<.0001
Error	12	0.00008880	0.00000740		
Corrected Total	13	0.02387			

Root MSE	0.00272	R-Square	0.9963
Dependent Mean	0.11431	Adj R-Sq	0.9960
Coeff Var	2.37974		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.01149	0.00195	5.88	<.0001
B2_667	1	0.87075	0.01536	56.68	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: B3_767

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.04552	0.04552	2954.69	<.0001
Error	12	0.00018487	0.00001541		
Corrected Total	13	0.04571			

Root MSE	0.00393	R-Square	0.9960
Dependent Mean	0.10449	Adj R-Sq	0.9956
Coeff Var	3.75652		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.00840	0.00206	4.09	0.0015
B3_667	1	0.87970	0.01618	54.36	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: B4_767

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.01595	0.01595	1111.28	<.0001
Error	12	0.00017227	0.00001436		
Corrected Total	13	0.01613			

Root MSE	0.00379	R-Square	0.9893
Dependent Mean	0.21390	Adj R-Sq	0.9884
Coeff Var	1.77135		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.02595	0.00573	4.53	0.0007
B4_667	1	0.85532	0.02566	33.34	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: B5_767

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.04044	0.04044	699.39	<.0001
Error	12	0.00069394	0.00005783		
Corrected Total	13	0.04114			

Root MSE	0.00760	R-Square	0.9831
Dependent Mean	0.15100	Adj R-Sq	0.9817
Coeff Var	5.03613		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-0.01849	0.00672	-2.75	0.0176
B5_667	1	0.99884	0.03777	26.45	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: B7_767

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.05907	0.05907	1133.74	<.0001
Error	12	0.00062526	0.00005211		
Corrected Total	13	0.05970			

Root MSE	0.00722	R-Square	0.9895
Dependent Mean	0.09804	Adj R-Sq	0.9887
Coeff Var	7.36260		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-0.00053441	0.00351	-0.15	0.8814
B7_667	1	0.98282	0.02919	33.67	<.0001

APPENDIX B

Two Sample Paired t-test for the Means of B1_767_cost and B1_667_cost

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B1_767_cost	40	0.047871	0.0305	0.0048
B1_667_cost	40	0.049972	0.0376	0.0059

Hypothesis Test

Null hypothesis: Mean of (B1_767_cost - B1_667_cost) = 0
 Alternative: Mean of (B1_767_cost - B1_667_cost) \neq 0

t Statistic	Df	Prob > t
-1.834	39	0.0743

Two Sample Paired t-test for the Means of B2_767_cost and B2_667_cost

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B2_767_cost	40	0.055274	0.0353	0.0056
B2_667_cost	40	0.056838	0.0403	0.0064

Hypothesis Test

Null hypothesis: Mean of (B2_767_cost - B2_667_cost) = 0
 Alternative: Mean of (B2_767_cost - B2_667_cost) \neq 0

t Statistic	Df	Prob > t
-1.870	39	0.0691

Two Sample Paired t-test for the Means of B3_767_cost and B3_667_cost

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B3_767_cost	40	0.049054	0.0476	0.0075
B3_667_cost	40	0.050107	0.0542	0.0086

Hypothesis Test

Null hypothesis: Mean of (B3_767_cost - B3_667_cost) = 0
 Alternative: Mean of (B3_767_cost - B3_667_cost) \neq 0

t Statistic	Df	Prob > t
-0.984	39	0.3313

Two Sample Paired t-test for the Means of B4_767_cost and B4_667_cost

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B4_767_cost	40	0.293277	0.0697	0.011
B4_667_cost	40	0.306895	0.0697	0.011

Hypothesis Test

Null hypothesis: Mean of (B4_767_cost - B4_667_cost) = 0
 Alternative: Mean of (B4_767_cost - B4_667_cost) \neq 0

t Statistic	Df	Prob > t
-15.079	39	<.0001

Two Sample Paired t-test for the Means of B5_767_cost and B5_667_cost

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B5_767_cost	40	0.118798	0.0581	0.0092
B5_667_cost	40	0.130933	0.057	0.009

Hypothesis Test

Null hypothesis: Mean of (B5_767_cost - B5_667_cost) = 0
 Alternative: Mean of (B5_767_cost - B5_667_cost) \neq 0

t Statistic	Df	Prob > t
-20.964	39	<.0001

Two Sample Paired t-test for the Means of B7_767_cost and B7_667_cost

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B7_767_cost	40	0.0709	0.048	0.0076
B7_667_cost	40	0.073562	0.0479	0.0076

Hypothesis Test

Null hypothesis: Mean of (B7_767_cost - B7_667_cost) = 0
 Alternative: Mean of (B7_767_cost - B7_667_cost) \neq 0

t Statistic	Df	Prob > t
-7.540	39	<.0001

Two Sample Paired t-test for the Means of B1_767_reg and B1_667_reg

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B1_767_reg	40	0.108736	0.0229	0.0036
B1_667_reg	40	0.108901	0.0229	0.0036

Hypothesis Test

Null hypothesis: Mean of (B1_767_reg - B1_667_reg) = 0
 Alternative: Mean of (B1_767_reg - B1_667_reg) \neq 0

t Statistic	Df	Prob > t
-1.059	39	0.2961

Two Sample Paired t-test for the Means of B2_767_reg and B2_667_reg

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B2_767_reg	40	0.090819	0.03	0.0047
B2_667_reg	40	0.090523	0.0298	0.0047

Hypothesis Test

Null hypothesis: Mean of (B2_767_reg - B2_667_reg) = 0
 Alternative: Mean of (B2_767_reg - B2_667_reg) \neq 0

t Statistic	Df	Prob > t
1.380	39	0.1754

Two Sample Paired t-test for the Means of B3_767_reg and B3_667_reg

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B3_767_reg	40	0.068655	0.0404	0.0064
B3_667_reg	40	0.068986	0.0404	0.0064

Hypothesis Test

Null hypothesis: Mean of (B3_767_reg - B3_667_reg) = 0
 Alternative: Mean of (B3_767_reg - B3_667_reg) \neq 0

t Statistic	Df	Prob > t
-1.770	39	0.0845

Two Sample Paired t-test for the Means of B4_767_reg and B4_667_reg

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B4_767_reg	40	0.254199	0.0609	0.0096
B4_667_reg	40	0.251302	0.0523	0.0083

Hypothesis Test

Null hypothesis: Mean of (B4_767_reg - B4_667_reg) = 0
 Alternative: Mean of (B4_767_reg - B4_667_reg) \neq 0

t Statistic	Df	Prob > t
1.884	39	0.0671

Two Sample Paired t-test for the Means of B5_767_reg and B5_667_reg

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B5_767_reg	40	0.139338	0.0504	0.008
B5_667_reg	40	0.138816	0.0482	0.0076

Hypothesis Test

Null hypothesis: Mean of (B5_767_reg - B5_667_reg) = 0
 Alternative: Mean of (B5_767_reg - B5_667_reg) \neq 0

t Statistic	Df	Prob > t
0.895	39	0.3764

Two Sample Paired t-test for the Means of B7_767_reg and B7_667_reg

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B7_767_reg	40	0.063702	0.0408	0.0064
B7_667_reg	40	0.064092	0.0399	0.0063

Hypothesis Test

Null hypothesis: Mean of (B7_767_reg - B7_667_reg) = 0
 Alternative: Mean of (B7_767_reg - B7_667_reg) \neq 0

t Statistic	Df	Prob > t
-1.205	39	0.2355

Two Sample Paired t-test for the Means of B1_767_eos and B1_667_eos

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B1_767_eos	40	0.108736	0.0229	0.0036
B1_667_eos	40	0.107605	0.0277	0.0044

Hypothesis Test

Null hypothesis: Mean of (B1_767_eos - B1_667_eos) = 0
 Alternative: Mean of (B1_767_eos - B1_667_eos) \neq 0

t Statistic	Df	Prob > t
1.457	39	0.1531

Two Sample Paired t-test for the Means of B2_767_eos and B2_667_eos

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B2_767_eos	40	0.090819	0.03	0.0047
B2_667_eos	40	0.090968	0.0343	0.0054

Hypothesis Test

Null hypothesis: Mean of (B2_767_eos - B2_667_eos) = 0
 Alternative: Mean of (B2_767_eos - B2_667_eos) \neq 0

t Statistic	Df	Prob > t
-0.210	39	0.8351

Two Sample Paired t-test for the Means of B3_767_eos and B3_667_eos

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B3_767_eos	40	0.068655	0.0404	0.0064
B3_667_eos	40	0.068831	0.0461	0.0073

Hypothesis Test

Null hypothesis: Mean of (B3_767_eos - B3_667_eos) = 0
 Alternative: Mean of (B3_767_eos - B3_667_eos) \neq 0

t Statistic	Df	Prob > t
-0.193	39	0.8482

Two Sample Paired t-test for the Means of B4_767_eos and B4_667_eos

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B4_767_eos	40	0.254199	0.0609	0.0096
B4_667_eos	40	0.263773	0.0611	0.0097

Hypothesis Test

Null hypothesis: Mean of (B4_767_eos - B4_667_eos) = 0
 Alternative: Mean of (B4_767_eos - B4_667_eos) \neq 0

t Statistic	Df	Prob > t
-12.100	39	<.0001

Two Sample Paired t-test for the Means of B5_767_eos and B5_667_eos

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B5_767_eos	40	0.139338	0.0504	0.008
B5_667_eos	40	0.157593	0.0484	0.0077

Hypothesis Test

Null hypothesis: Mean of (B5_767_eos - B5_667_eos) = 0
 Alternative: Mean of (B5_767_eos - B5_667_eos) \neq 0

t Statistic	Df	Prob > t
-32.370	39	<.0001

Two Sample Paired t-test for the Means of B7_767_eos and B7_667_eos

Sample Statistics

Group	N	Mean	Std. Dev.	Std. Error
B7_767_eos	40	0.063702	0.0408	0.0064
B7_667_eos	40	0.0659	0.0407	0.0064

Hypothesis Test

Null hypothesis: Mean of (B7_767_eos - B7_667_eos) = 0
 Alternative: Mean of (B7_767_eos - B7_667_eos) \neq 0

t Statistic	Df	Prob > t
-7.333	39	<.0001

APPENDIX C

Signature Listing for COST Corrected Images

Number of signatures: 7

Number of layers: 6

Signature: Closed-Canopy Forest

Number of pixels: 450

Statistics				
Layer	Minimum	Maximum	Mean	Sigma
1	0.000	58.000	31.962	9.714
2	20.000	67.000	49.038	7.357
3	11.000	51.000	34.953	6.835
4	95.000	163.000	127.198	10.534
5	27.000	91.000	52.782	10.036
7	27.000	75.000	44.971	7.787

Covariance						
Layer	1	2	3	4	5	7
1	94.362	29.006	24.713	-12.545	-15.941	-2.146
2	29.006	54.130	24.817	1.050	-9.580	0.039
3	24.713	24.817	46.722	-10.314	-13.787	-0.739
4	-12.545	1.050	-10.314	110.965	57.342	27.462
5	-15.941	-9.580	-13.787	57.342	100.723	55.248
7	-2.146	0.039	-0.739	27.462	55.248	60.638

Signature: Succession Forest

Number of pixels: 450

Statistics				
Layer	Minimum	Maximum	Mean	Sigma
1	0.000	58.000	34.587	8.821
2	34.000	79.000	54.500	7.155
3	15.000	55.000	35.344	6.626
4	112.000	212.000	157.467	13.833
5	35.000	100.000	72.393	11.880
7	31.000	78.000	54.538	8.172

Covariance						
Layer	1	2	3	4	5	7
1	77.815	18.100	17.637	0.730	-11.539	-5.450
2	18.100	51.190	19.970	35.437	5.663	5.131
3	17.637	19.970	43.910	0.135	-10.775	0.687
4	0.730	35.437	0.135	191.341	80.268	29.499
5	-11.539	5.663	-10.775	80.268	141.126	71.545
7	-5.450	5.131	0.687	29.499	71.545	66.788

Signature: Purma
Number of pixels: 450

Statistics				
Layer	Minimum	Maximum	Mean	Sigma
1	11.000	63.000	38.742	9.175
2	44.000	99.000	68.793	9.630
3	18.000	62.000	40.356	6.345
4	157.000	248.000	196.582	17.134
5	56.000	144.000	98.451	13.696
7	37.000	99.000	71.024	10.018

Covariance						
Layer	1	2	3	4	5	7
1	84.174	28.566	16.457	0.995	8.275	9.516
2	28.566	92.730	25.831	67.820	46.120	34.335
3	16.457	25.831	40.256	4.748	12.875	13.308
4	0.995	67.820	4.748	293.558	136.472	71.260
5	8.275	46.120	12.875	136.472	187.576	108.902
7	9.516	34.335	13.308	71.260	108.902	100.354

Signature: Grassland
Number of pixels: 450

Statistics				
Layer	Minimum	Maximum	Mean	Sigma
1	34.000	102.000	63.133	10.100
2	74.000	134.000	107.289	10.770
3	44.000	119.000	82.782	13.540
4	112.000	230.000	174.467	19.953
5	92.000	221.000	142.918	20.749
7	67.000	223.000	118.360	20.498

Covariance						
Layer	1	2	3	4	5	7
1	102.004	42.692	80.884	-29.296	93.632	103.673
2	42.692	115.983	78.230	62.607	116.917	95.007
3	80.884	78.230	183.329	-103.083	175.318	206.696
4	-29.296	62.607	-103.083	398.107	19.976	-108.972
5	93.632	116.917	175.318	19.976	430.534	374.014
7	103.673	95.007	206.696	-108.972	374.014	420.182

Signature: Agriculture
Number of pixels: 450

Layer	Statistics			
	Minimum	Maximum	Mean	Sigma
1	40.000	131.000	78.276	15.128
2	54.000	148.000	93.249	15.210
3	62.000	186.000	117.091	21.280
4	49.000	146.000	90.471	14.814
5	88.000	253.000	168.287	29.379
7	97.000	255.000	187.918	33.782

Layer	Covariance					
	1	2	3	4	5	7
1	228.841	173.989	226.754	48.890	238.965	180.145
2	173.989	231.350	257.879	104.406	283.893	195.299
3	226.754	257.879	452.827	104.171	454.967	276.894
4	48.890	104.406	104.171	219.466	174.239	-100.387
5	238.965	283.893	454.967	174.239	863.118	582.019
7	180.145	195.299	276.894	-100.387	582.019	1141.198

Signature: Bare Surface
Number of pixels: 450

Layer	Statistics			
	Minimum	Maximum	Mean	Sigma
1	58.000	255.000	210.649	46.522
2	67.000	255.000	227.051	41.056
3	87.000	255.000	233.567	37.363
4	15.000	171.000	101.529	17.998
5	54.000	246.000	155.640	38.475
7	102.000	255.000	225.558	34.804

Layer	Covariance					
	1	2	3	4	5	7
1	2164.291	1758.806	1516.937	263.870	554.876	904.397
2	1758.806	1685.621	1476.501	284.514	502.482	843.105
3	1516.937	1476.501	1395.997	222.570	432.251	790.028
4	263.870	284.514	222.570	323.933	478.701	284.192
5	554.876	502.482	432.251	478.701	1480.356	989.722
7	904.397	843.105	790.028	284.192	989.722	1211.334

Signature: Water
Number of pixels: 450

Layer	Statistics			
	Minimum	Maximum	Mean	Sigma
1	23.000	124.000	66.460	14.559
2	44.000	173.000	78.138	14.487
3	29.000	183.000	67.391	18.219
4	0.000	100.000	16.147	14.419
5	0.000	34.000	0.691	3.509
7	0.000	40.000	7.104	8.661

Layer	Covariance					
	1	2	3	4	5	7
1	211.959	154.910	153.025	54.797	3.735	43.513
2	154.910	209.861	208.812	80.741	7.577	50.732
3	153.025	208.812	331.931	120.210	7.633	69.249
4	54.797	80.741	120.210	207.921	10.825	62.802
5	3.735	7.577	7.633	10.825	12.316	10.669
7	43.513	50.732	69.249	62.802	10.669	75.007

Signature Listing for Regression Corrected Image

Number of signatures: 7

Number of layers: 6

Signature: Closed-Canopy Forest

Number of pixels: 450

Statistics				
Layer	Minimum	Maximum	Mean	Sigma
1	121.000	148.000	134.393	4.742
2	93.000	122.000	110.178	4.591
3	52.000	80.000	67.989	4.918
4	103.000	167.000	138.038	10.409
5	52.000	106.000	75.973	8.207
7	30.000	77.000	48.704	7.576

Covariance						
Layer	1	2	3	4	5	7
1	22.484	7.162	6.608	-2.296	-0.831	1.239
2	7.162	21.073	8.764	6.450	1.887	2.558
3	6.608	8.764	24.189	-0.334	0.873	3.774
4	-2.296	6.450	-0.334	108.339	40.723	21.356
5	-0.831	1.887	0.873	40.723	67.353	42.977
7	1.239	2.558	3.774	21.356	42.977	57.394

Signature: Succession Forest

Number of pixels: 450

Statistics				
Layer	Minimum	Maximum	Mean	Sigma
1	121.000	148.000	135.858	4.228
2	101.000	129.000	113.827	4.527
3	55.000	83.000	68.633	4.718
4	124.000	220.000	168.562	14.504
5	63.000	119.000	93.320	9.856
7	34.000	81.000	57.982	7.997

Covariance						
Layer	1	2	3	4	5	7
1	17.877	4.659	3.836	5.913	-0.012	-1.408
2	4.659	20.491	8.001	29.300	8.033	4.765
3	3.836	8.001	22.255	11.507	4.256	4.089
4	5.913	29.300	11.507	210.376	61.737	25.498
5	-0.012	8.033	4.256	61.737	97.131	59.039
7	-1.408	4.765	4.089	25.498	59.039	63.955

Signature: Purma
Number of pixels: 450

Statistics				
Layer	Minimum	Maximum	Mean	Sigma
1	126.000	149.000	137.791	4.578
2	108.000	142.000	122.969	6.402
3	58.000	91.000	72.458	5.238
4	163.000	255.000	208.969	14.973
5	75.000	155.000	117.551	11.430
7	40.000	101.000	74.562	9.822

Covariance						
Layer	1	2	3	4	5	7
1	20.958	10.198	6.419	4.406	5.977	5.521
2	10.198	40.988	15.640	35.442	22.708	21.325
3	6.419	15.640	27.438	8.952	13.404	13.383
4	4.406	35.442	8.952	224.195	80.944	52.908
5	5.977	22.708	13.404	80.944	130.644	89.097
7	5.521	21.325	13.383	52.908	89.097	96.469

Signature: Grassland
Number of pixels: 450

Statistics				
Layer	Minimum	Maximum	Mean	Sigma
1	136.000	166.000	149.991	5.131
2	126.000	170.000	148.947	8.031
3	77.000	138.000	107.082	11.104
4	119.000	242.000	187.073	22.037
5	110.000	235.000	158.789	20.307
7	71.000	228.000	121.949	20.799

Covariance						
Layer	1	2	3	4	5	7
1	26.325	17.714	35.186	-5.752	51.842	56.919
2	17.714	64.492	49.131	76.305	96.960	76.783
3	35.186	49.131	123.301	-61.084	148.657	175.071
4	-5.752	76.305	-61.084	485.618	76.428	-74.615
5	51.842	96.960	148.657	76.428	412.376	371.419
7	56.919	76.783	175.071	-74.615	371.419	432.596

Signature: Agriculture
Number of pixels: 450

Layer	Statistics			
	Minimum	Maximum	Mean	Sigma
1	139.000	179.000	157.364	6.936
2	114.000	172.000	139.340	9.763
3	91.000	184.000	134.584	16.893
4	55.000	152.000	99.411	15.356
5	111.000	255.000	182.204	26.077
7	101.000	255.000	191.382	33.815

Layer	Covariance					
	1	2	3	4	5	7
1	48.103	49.653	80.183	20.228	91.972	87.511
2	49.653	95.316	129.467	67.853	157.612	125.199
3	80.183	129.467	285.375	83.643	318.234	220.533
4	20.228	67.853	83.643	235.793	148.564	-114.837
5	91.972	157.612	318.234	148.564	679.994	533.211
7	87.511	125.199	220.533	-114.837	533.211	1143.444

Signature: Bare Surface
Number of pixels: 450

Layer	Statistics			
	Minimum	Maximum	Mean	Sigma
1	148.000	255.000	225.431	24.444
2	122.000	255.000	232.964	29.380
3	110.000	255.000	234.920	32.807
4	27.000	175.000	110.822	17.459
5	75.000	247.000	170.482	33.710
7	104.000	255.000	228.193	33.942

Layer	Covariance					
	1	2	3	4	5	7
1	597.524	644.848	682.353	116.393	197.179	464.694
2	644.848	863.166	927.565	207.577	326.213	616.308
3	682.353	927.565	1076.274	207.008	364.671	711.842
4	116.393	207.577	207.008	304.815	389.284	260.894
5	197.179	326.213	364.671	389.284	1136.379	860.613
7	464.694	616.308	711.842	260.894	860.613	1152.040

Signature: Water
Number of pixels: 450

Layer	Statistics			
	Minimum	Maximum	Mean	Sigma
1	131.000	184.000	152.024	7.861
2	108.000	197.000	129.744	10.119
3	65.000	194.000	94.647	15.101
4	8.000	108.000	22.969	14.071
5	0.000	53.000	5.442	8.677
7	0.000	44.000	9.964	9.525

Layer	Covariance					
	1	2	3	4	5	7
1	61.801	60.792	67.770	10.865	24.116	29.845
2	60.792	102.396	119.355	34.442	35.066	42.240
3	67.770	119.355	228.042	86.495	45.756	62.756
4	10.865	34.442	86.495	197.995	49.816	53.823
5	24.116	35.066	45.756	49.816	75.298	67.927
7	29.845	42.240	62.756	53.823	67.927	90.729

Signature Listing for Uncorrected Images

Number of signatures: 7

Number of layers: 6

Signature: Closed-Canopy Forest

Number of pixels: 450

Layer	Statistics			
	Minimum	Maximum	Mean	Sigma
1	0.000	64.000	35.382	11.150
2	19.000	69.000	50.609	7.894
3	11.000	52.000	35.676	7.361
4	95.000	163.000	128.096	10.351
5	50.000	113.000	74.993	10.612
7	32.000	79.000	50.293	7.621

Layer	Covariance					
	1	2	3	4	5	7
1	124.321	47.651	42.400	-17.429	-48.561	-8.455
2	47.651	62.310	32.975	0.804	-27.156	-2.070
3	42.400	32.975	54.180	-9.383	-28.281	-1.896
4	-17.429	0.804	-9.383	107.151	53.880	25.776
5	-48.561	-27.156	-28.281	53.880	112.617	53.343
7	-8.455	-2.070	-1.896	25.776	53.343	58.083

Signature: Succession Forest

Number of pixels: 450

Layer	Statistics			
	Minimum	Maximum	Mean	Sigma
1	0.000	64.000	37.169	10.224
2	34.000	78.000	55.629	7.266
3	15.000	56.000	35.882	7.137
4	114.000	212.000	158.184	13.687
5	57.000	120.000	92.773	12.066
7	36.000	83.000	59.598	7.976

Layer	Covariance					
	1	2	3	4	5	7
1	104.537	30.972	33.991	-4.503	-45.178	-10.297
2	30.972	52.800	26.117	32.307	-10.220	2.995
3	33.991	26.117	50.933	1.608	-25.216	-0.477
4	-4.503	32.307	1.608	187.331	73.026	28.094
5	-45.178	-10.220	-25.216	73.026	145.583	65.583
7	-10.297	2.995	-0.477	28.094	65.583	63.617

Signature: Purma
Number of pixels: 450

Statistics				
Layer	Minimum	Maximum	Mean	Sigma
1	11.000	64.000	41.416	9.914
2	44.000	98.000	69.540	9.373
3	18.000	62.000	41.011	6.591
4	158.000	248.000	197.253	16.787
5	76.000	159.000	115.551	13.501
7	42.000	103.000	75.576	9.743

Covariance						
Layer	1	2	3	4	5	7
1	98.292	27.285	22.035	-28.965	-25.258	0.384
2	27.285	87.857	26.508	54.030	32.911	29.849
3	22.035	26.508	43.441	-1.074	2.495	11.392
4	-28.965	54.030	-1.074	281.811	139.778	68.188
5	-25.258	32.911	2.495	139.778	182.284	99.310
7	0.384	29.849	11.392	68.188	99.310	94.931

Signature: Grassland
Number of pixels: 450

Statistics				
Layer	Minimum	Maximum	Mean	Sigma
1	33.000	98.000	65.000	10.736
2	73.000	136.000	107.778	11.242
3	44.000	120.000	83.224	13.585
4	112.000	230.000	175.207	19.933
5	107.000	222.000	154.673	18.182
7	71.000	224.000	121.802	20.001

Covariance						
Layer	1	2	3	4	5	7
1	115.269	59.258	93.223	-23.621	78.187	109.301
2	59.258	126.391	84.916	65.523	94.883	96.112
3	93.223	84.916	184.549	-99.715	145.196	201.969
4	-23.621	65.523	-99.715	397.331	16.885	-103.554
5	78.187	94.883	145.196	16.885	330.586	315.287
7	109.301	96.112	201.969	-103.554	315.287	400.034

Signature: Agriculture
Number of pixels: 450

Layer	Statistics			
	Minimum	Maximum	Mean	Sigma
1	38.000	125.000	79.684	13.689
2	54.000	146.000	93.936	14.838
3	62.000	185.000	117.318	21.075
4	51.000	146.000	91.418	14.602
5	103.000	253.000	176.816	26.569
7	101.000	255.000	189.611	32.974

Layer	Covariance					
	1	2	3	4	5	7
1	187.379	150.347	198.938	31.482	173.572	174.309
2	150.347	220.158	247.833	96.392	241.645	190.320
3	198.938	247.833	444.142	97.847	402.297	269.988
4	31.482	96.392	97.847	213.224	153.166	-96.855
5	173.572	241.645	402.297	153.166	705.888	495.926
7	174.309	190.320	269.988	-96.855	495.926	1087.263

Signature: Bare Surface
Number of pixels: 450

Layer	Statistics			
	Minimum	Maximum	Mean	Sigma
1	64.000	255.000	208.944	45.265
2	69.000	255.000	226.993	40.788
3	87.000	255.000	233.542	37.266
4	15.000	171.000	102.511	17.766
5	74.000	248.000	165.640	34.562
7	106.000	255.000	226.344	33.917

Layer	Covariance					
	1	2	3	4	5	7
1	2048.886	1682.452	1461.836	242.407	464.283	869.231
2	1682.452	1663.641	1463.371	278.115	465.071	823.978
3	1461.836	1463.371	1388.779	217.522	394.944	770.922
4	242.407	278.115	217.522	315.636	419.164	274.487
5	464.283	465.071	394.944	419.164	1194.525	848.991
7	869.231	823.978	770.922	274.487	848.991	1150.333

Signature: Water
Number of pixels: 450

Layer	Statistics			
	Minimum	Maximum	Mean	Sigma
1	22.000	127.000	68.113	16.605
2	44.000	174.000	78.836	15.201
3	29.000	184.000	67.882	18.366
4	0.000	100.000	17.233	14.794
5	0.000	48.000	7.351	7.636
7	0.000	45.000	12.278	9.236

Layer	Covariance					
	1	2	3	4	5	7
1	275.731	197.010	177.869	87.029	0.025	57.423
2	197.010	231.082	220.914	97.074	8.688	56.667
3	177.869	220.914	337.316	129.099	25.897	72.592
4	87.029	97.074	129.099	218.852	36.539	69.628
5	0.025	8.688	25.897	36.539	58.309	44.187
7	57.423	56.667	72.592	69.628	44.187	85.310

APPENDIX D

Signature Separability Listing for COST Correction

Distance measure: Transformed Divergence

Using bands: 1 2 3 4 5 7

Taken 4 at a time

Class

- 1 Closed-Canopy Forest
- 2 Succession
- 3 Purma
- 4 Grassland
- 5 Agriculture
- 6 Bare Surface
- 7 Water
- 8 Shadow

Best Minimum Separability

Bands	AVE	MIN	Class Pairs:							
			1: 2	1: 3	1: 4	1: 5	1: 6	1: 7	1: 8	
			2: 3	2: 4	2: 5	2: 6	2: 7	2: 8	3: 4	
			3: 5	3: 6	3: 7	3: 8	4: 5	4: 6	4: 7	
			4: 8	5: 6	5: 7	5: 8	6: 7	6: 8	7: 8	
1 3 4 7	1985	1691	1691	2000	2000	2000	2000	2000	2000	2000
			1902	2000	2000	2000	2000	2000	2000	2000
			2000	2000	2000	2000	1999	2000	2000	2000
			2000	2000	2000	2000	2000	2000	2000	2000

Best Average Separability

Bands	AVE	MIN	Class Pairs:							
			1: 2	1: 3	1: 4	1: 5	1: 6	1: 7	: 8	
			2: 3	2: 4	2: 5	2: 6	2: 7	2: 8	3: 4	
			3: 5	3: 6	3: 7	3: 8	4: 5	4: 6	4: 7	
			4: 8	5: 6	5: 7	5: 8	6: 7	6: 8	7: 8	
1 3 4 7	1985	1691	1691	2000	2000	2000	2000	2000	2000	2000
			1902	2000	2000	2000	2000	2000	2000	2000
			2000	2000	2000	2000	1999	2000	2000	2000
			2000	2000	2000	2000	2000	2000	2000	2000

Signature Separability Listing for Regression Correction

Distance measure: Transformed Divergence

Using bands: 1 2 3 4 5 7

Taken 4 at a time

Class

- 1 Closed-Canopy Forest
- 2 Succession Forest
- 3 Purma
- 4 Grassland
- 5 Agriculture
- 6 Bare Surface
- 7 Water
- 8 Shadow

Best Minimum Separability

Bands	AVE	MIN	Class Pairs:							
			1: 2	1: 3	1: 4	1: 5	1: 6	1: 7	1: 8	
			2: 3	2: 4	2: 5	2: 6	2: 7	2: 8	3: 4	
			3: 5	3: 6	3: 7	3: 8	4: 5	4: 6	4: 7	
			4: 8	5: 6	5: 7	5: 8	6: 7	6: 8	7: 8	
1 3 4 7	1984	1661	1661	2000	2000	2000	2000	2000	2000	
			1903	2000	2000	2000	2000	2000	2000	
			2000	2000	2000	2000	1999	2000	2000	
			2000	2000	2000	2000	2000	2000	2000	

Best Average Separability

Bands	AVE	MIN	Class Pairs:							
			1: 2	1: 3	1: 4	1: 5	1: 6	1: 7	1: 8	
			2: 3	2: 4	2: 5	2: 6	2: 7	2: 8	3: 4	
			3: 5	3: 6	3: 7	3: 8	4: 5	4: 6	4: 7	
			4: 8	5: 6	5: 7	5: 8	6: 7	6: 8	7: 8	
1 3 4 7	1984	1661	1661	2000	2000	2000	2000	2000	2000	
			1903	2000	2000	2000	2000	2000	2000	
			2000	2000	2000	2000	1999	2000	2000	
			2000	2000	2000	2000	2000	2000	2000	

Signature Separability Listing for Uncorrected

Distance measure: Transformed Divergence
 Using bands: 1 2 3 4 5 7
 Taken 4 at a time

Class

- 1 Closed-Canopy Forest
- 2 Succession Forest
- 3 Purma
- 4 Grassland
- 5 Agriculture
- 6 Bare Surface
- 7 Water
- 8 Shadow

Best Minimum Separability

Bands	AVE	MIN	Class	Pairs:						
			1: 2	1: 3	1: 4	1: 5	1: 6	1: 7	1: 8	
			2: 3	2: 4	2: 5	2: 6	2: 7	2: 8	3: 4	
			3: 5	3: 6	3: 7	3: 8	4: 5	4: 6	4: 7	
			4: 8	5: 6	5: 7	5: 8	6: 7	6: 8	7: 8	
1 3 4 7	1984	1661	1661	2000	2000	2000	2000	2000	2000	2000
			1903	2000	2000	2000	2000	2000	2000	2000
			2000	2000	2000	2000	1999	2000	2000	2000
			2000	2000	2000	2000	2000	2000	2000	2000

Best Average Separability

Bands	AVE	MIN	Class	Pairs:						
			1: 2	1: 3	1: 4	1: 5	1: 6	1: 7	1: 8	
			2: 3	2: 4	2: 5	2: 6	2: 7	2: 8	3: 4	
			3: 5	3: 6	3: 7	3: 8	4: 5	4: 6	4: 7	
			4: 8	5: 6	5: 7	5: 8	6: 7	6: 8	7: 8	
1 3 4 7	1984	1661	1661	2000	2000	2000	2000	2000	2000	2000
			1903	2000	2000	2000	2000	2000	2000	2000
			2000	2000	2000	2000	1999	2000	2000	2000
			2000	2000	2000	2000	2000	2000	2000	2000

APPENDIX E

Crossvalidation Results

Abbreviations:

CCF – Closed-Canopy Forest

SF – Succession Forest

P – Purma

G – Grassland

A – Agriculture

BS – Bare Surface

W – Water

Uncorrected Bands 1 3 4 7

Pixel-Level Crossvalidation

The FREQ Procedure

Table of Reference by Classified

Reference Classified

Frequency Row Pct	CCF	SF	P	G	A	BS	W	Total
CCF	418 92.89	32 7.11	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	450
SF	61 13.56	370 82.22	19 4.22	0 0.00	0 0.00	0 0.00	0 0.00	450
P	0 0.00	43 9.56	405 90.00	2 0.44	0 0.00	0 0.00	0 0.00	450
G	2 0.44	2 0.44	49 10.89	391 86.89	6 1.33	0 0.00	0 0.00	450
A	4 0.89	0 0.00	0 0.00	1 0.22	445 98.89	0 0.00	0 0.00	450
BS	0 0.00	0 0.00	0 0.00	3 0.67	28 6.22	419 93.11	0 0.00	450
W	2 0.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	448 99.56	450
Total	487	447	473	397	479	419	448	3150

Statistics for Table of Reference by Classified

Test of Symmetry

Statistic (S)	106.2185
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9059	0.0057	0.8948	0.9170
Weighted Kappa	0.9606	0.0027	0.9552	0.9660

Sample Size = 3150

Uncorrected Bands 1 3 4 7
Polygon-Level Crossvalidation

The FREQ Procedure

Table of Reference by Classified

Reference Classified

Frequency Row Pct	CCF	SF	P	G	A	BS	W	Total
CCF	1236 91.56	114 8.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1350
SF	189 14.00	1095 81.11	66 4.89	0 0.00	0 0.00	0 0.00	0 0.00	1350
P	0 0.00	141 10.44	1203 89.11	6 0.44	0 0.00	0 0.00	0 0.00	1350
G	6 0.44	6 0.44	150 11.11	1170 86.67	18 1.33	0 0.00	0 0.00	1350
A	12 0.89	0 0.00	0 0.00	3 0.22	1335 98.89	0 0.00	0 0.00	1350
BS	0 0.00	0 0.00	0 0.00	9 0.67	87 6.44	1254 92.89	0 0.00	1350
W	6 0.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1449	1356	1419	1188	1440	1254	1344	9450

Statistics for Table of Reference by Classified

Test of Symmetry

Statistic (S)	315.3756
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8996	0.0034	0.8930	0.9062
Weighted Kappa	0.9582	0.0016	0.9551	0.9614

Sample Size = 9450

Uncorrected Bands 1 3 4 7
Region-Level Crossvalidation

The FREQ Procedure

Table of Reference by Classified

Reference Classified

Frequency Row Pct	CCF	SF	P	G	A	BS	W	Total
CCF	1224 90.67	126 9.33	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1350
SF	183 13.56	1098 81.33	69 5.11	0 0.00	0 0.00	0 0.00	0 0.00	1350
P	0 0.00	177 13.11	1167 86.44	6 0.44	0 0.00	0 0.00	0 0.00	1350
G	12 0.89	6 0.44	141 10.44	1170 86.67	21 1.56	0 0.00	0 0.00	1350
A	9 0.67	0 0.00	0 0.00	3 0.22	1338 99.11	0 0.00	0 0.00	1350
BS	0 0.00	0 0.00	0 0.00	9 0.67	87 6.44	1254 92.89	0 0.00	1350
W	6 0.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1434	1407	1377	1188	1446	1254	1344	9450

Statistics for Table of Reference by Classified

Test of Symmetry

Statistic (S) 324.4088
DF 21
Pr > S <.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8944	0.0034	0.8877	0.9012
Weighted Kappa	0.9562	0.0016	0.9529	0.9594

Sample Size = 9450

Uncorrected Bands 1 3 4 7
Image-Level Crossvalidation

The FREQ Procedure

Table of Reference by Classified

Reference Classified

Frequency Row Pct	CCF	SF	P	G	A	BS	W	Total
CCF	1230 91.11	120 8.89	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1350
SF	171 12.67	1053 78.00	126 9.33	0 0.00	0 0.00	0 0.00	0 0.00	1350
P	0 0.00	252 18.67	1092 80.89	6 0.44	0 0.00	0 0.00	0 0.00	1350
G	9 0.67	6 0.44	123 9.11	1191 88.22	21 1.56	0 0.00	0 0.00	1350
A	3 0.22	0 0.00	0 0.00	6 0.44	1341 99.33	0 0.00	0 0.00	1350
BS	0 0.00	0 0.00	0 0.00	9 0.67	102 7.56	1239 91.78	0 0.00	1350
W	6 0.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1419	1431	1341	1212	1464	1239	1344	9450

Statistics for Table of Reference by Classified

Test of Symmetry

Statistic (S) 300.3878
DF 21
Pr > S <.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8815	0.0036	0.8744	0.8886
Weighted Kappa	0.9524	0.0016	0.9492	0.9556

Sample Size = 9450

Regression Corrected Bands 1 3 4 7
Pixel-Level Crossvalidation

The FREQ Procedure

Table of Reference by Classified

Reference Classified

Frequency Row Pct	CCF	SF	P	G	A	BS	W	Total
CCF	422 93.78	28 6.22	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	450
SF	61 13.56	356 79.11	33 7.33	0 0.00	0 0.00	0 0.00	0 0.00	450
P	0 0.00	29 6.44	419 93.11	2 0.44	0 0.00	0 0.00	0 0.00	450
G	5 1.11	3 0.67	51 11.33	385 85.56	6 1.33	0 0.00	0 0.00	450
A	3 0.67	0 0.00	0 0.00	1 0.22	446 99.11	0 0.00	0 0.00	450
BS	0 0.00	0 0.00	0 0.00	3 0.67	28 6.22	419 93.11	0 0.00	450
W	2 0.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	448 99.56	450
Total	493	416	503	391	480	419	448	3150

Statistics for Table of Reference by Classified

Test of Symmetry

Statistic (S)	105.3673
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9056	0.0057	0.8944	0.9167
Weighted Kappa	0.9599	0.0028	0.9544	0.9654

Sample Size = 3150

Regression Corrected Bands 1 3 4 7
Polygon-Level Crossvalidation

The FREQ Procedure

Table of Reference by Classified

Reference Classified

Frequency Row Pct	CCF	SF	P	G	A	BS	W	Total
CCF	1263 93.56	87 6.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1350
SF	195 14.44	1053 78.00	102 7.56	0 0.00	0 0.00	0 0.00	0 0.00	1350
P	0 0.00	93 6.89	1251 92.67	6 0.44	0 0.00	0 0.00	0 0.00	1350
G	15 1.11	9 0.67	162 12.00	1146 84.89	18 1.33	0 0.00	0 0.00	1350
A	9 0.67	0 0.00	0 0.00	3 0.22	1338 99.11	0 0.00	0 0.00	1350
BS	0 0.00	0 0.00	0 0.00	9 0.67	87 6.44	1254 92.89	0 0.00	1350
W	6 0.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1488	1242	1515	1164	1443	1254	1344	9450

Statistics for Table of Reference by Classified

Test of Symmetry

Statistic (S) 332.3485
DF 21
Pr > S <.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9011	0.0033	0.8946	0.9077
Weighted Kappa	0.9582	0.0016	0.9550	0.9614

Sample Size = 9450

Regression Corrected Bands 1 3 4 7
Region-Level Crossvalidation

The FREQ Procedure

Table of Reference by Classified

Reference Classified

Frequency Row Pct	CCF	SF	P	G	A	BS	W	Total
CCF	1257 93.11	93 6.89	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1350
SF	189 14.00	1053 78.00	108 8.00	0 0.00	0 0.00	0 0.00	0 0.00	1350
P	0 0.00	93 6.89	1251 92.67	6 0.44	0 0.00	0 0.00	0 0.00	1350
G	15 1.11	12 0.89	165 12.22	1140 84.44	18 1.33	0 0.00	0 0.00	1350
A	9 0.67	0 0.00	0 0.00	3 0.22	1338 99.11	0 0.00	0 0.00	1350
BS	0 0.00	0 0.00	0 0.00	9 0.67	90 6.67	1251 92.67	0 0.00	1350
W	6 0.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1476	1251	1524	1158	1446	1251	1344	9450

Statistics for Table of Reference by Classified

Test of Symmetry

Statistic (S) 333.3566
DF 21
Pr > S <.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8993	0.0034	0.8927	0.9059
Weighted Kappa	0.9574	0.0016	0.9541	0.9606

Sample Size = 9450

Regression Corrected Bands 1 3 4 7
Image-Level Crossvalidation

The FREQ Procedure

Table of Reference by Classified

Reference Classified

Frequency Row Pct	CCF	SF	P	G	A	BS	W	Total
CCF	1203 89.11	147 10.89	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1350
SF	195 14.44	1035 76.67	120 8.89	0 0.00	0 0.00	0 0.00	0 0.00	1350
P	0 0.00	90 6.67	1254 92.89	6 0.44	0 0.00	0 0.00	0 0.00	1350
G	15 1.11	18 1.33	195 14.44	1104 81.78	18 1.33	0 0.00	0 0.00	1350
A	6 0.44	0 0.00	0 0.00	6 0.44	1338 99.11	0 0.00	0 0.00	1350
BS	0 0.00	0 0.00	0 0.00	9 0.67	87 6.44	1254 92.89	0 0.00	1350
W	6 0.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1425	1290	1569	1125	1443	1254	1344	9450

Statistics for Table of Reference by Classified

Test of Symmetry

Statistic (S) 335.7390
DF 21
Pr > S <.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8867	0.0036	0.8797	0.8936
Weighted Kappa	0.9527	0.0017	0.9494	0.9560

Sample Size = 9450

COST Corrected Bands 1 3 4 7
Pixel-Level Crossvalidation

The FREQ Procedure

Table of Reference by Classified

Reference Classified

Frequency Row Pct	CCF	SF	P	G	A	BS	W	Total
CCF	412 91.56	38 8.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	450
SF	60 13.33	370 82.22	20 4.44	0 0.00	0 0.00	0 0.00	0 0.00	450
P	0 0.00	43 9.56	405 90.00	2 0.44	0 0.00	0 0.00	0 0.00	450
G	2 0.44	2 0.44	48 10.67	392 87.11	6 1.33	0 0.00	0 0.00	450
A	3 0.67	0 0.00	0 0.00	1 0.22	446 99.11	0 0.00	0 0.00	450
BS	0 0.00	0 0.00	0 0.00	3 0.67	31 6.89	416 92.44	0 0.00	450
W	2 0.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	448 99.56	450
Total	479	453	473	398	483	416	448	3150

Statistics for Table of Reference by Classified

Test of Symmetry

Statistic (S)	102.2270
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9033	0.0057	0.8921	0.9146
Weighted Kappa	0.9600	0.0027	0.9547	0.9653

Sample Size = 3150

COST Corrected Bands 1 3 4 7
Polygon-Level Crossvalidation

The FREQ Procedure

Table of Reference by Classified

Reference Classified

Frequency Row Pct	CCF	SF	P	G	A	BS	W	Total
CCF	1236 91.56	114 8.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1350
SF	186 13.78	1101 81.56	63 4.67	0 0.00	0 0.00	0 0.00	0 0.00	1350
P	0 0.00	153 11.33	1191 88.22	6 0.44	0 0.00	0 0.00	0 0.00	1350
G	6 0.44	6 0.44	150 11.11	1170 86.67	18 1.33	0 0.00	0 0.00	1350
A	9 0.67	0 0.00	0 0.00	3 0.22	1338 99.11	0 0.00	0 0.00	1350
BS	0 0.00	0 0.00	0 0.00	9 0.67	93 6.89	1248 92.44	0 0.00	1350
W	6 0.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1443	1374	1404	1188	1449	1248	1344	9450

Statistics for Table of Reference by Classified

Test of Symmetry

Statistic (S)	327.4174
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8985	0.0034	0.8919	0.9051
Weighted Kappa	0.9582	0.0016	0.9551	0.9614

Sample Size = 9450

COST Corrected Bands 1 3 4 7
Region-Level Crossvalidation

The FREQ Procedure

Table of Reference by Classified

Reference Classified

Frequency Row Pct	CCF	SF	P	G	A	BS	W	Total
CCF	1209 89.56	141 10.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1350
SF	186 13.78	1083 80.22	81 6.00	0 0.00	0 0.00	0 0.00	0 0.00	1350
P	0 0.00	180 13.33	1164 86.22	6 0.44	0 0.00	0 0.00	0 0.00	1350
G	12 0.89	6 0.44	141 10.44	1173 86.89	18 1.33	0 0.00	0 0.00	1350
A	9 0.67	0 0.00	0 0.00	3 0.22	1338 99.11	0 0.00	0 0.00	1350
BS	0 0.00	0 0.00	0 0.00	9 0.67	87 6.44	1254 92.89	0 0.00	1350
W	6 0.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1422	1410	1386	1191	1443	1254	1344	9450

Statistics for Table of Reference by Classified

Test of Symmetry

Statistic (S) 307.4383
DF 21
Pr > S <.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8907	0.0035	0.8839	0.8976
Weighted Kappa	0.9548	0.0017	0.9515	0.9580

Sample Size = 9450

COST Corrected Bands 1 3 4 7
Image-Level Crossvalidation

The FREQ Procedure

Table of Reference by Classified

Reference Classified

Frequency Row Pct	CCF	SF	P	G	A	BS	W	Total
CCF	1212 89.78	138 10.22	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1350
SF	186 13.78	1023 75.78	141 10.44	0 0.00	0 0.00	0 0.00	0 0.00	1350
P	0 0.00	255 18.89	1086 80.44	9 0.67	0 0.00	0 0.00	0 0.00	1350
G	9 0.67	9 0.67	144 10.67	1164 86.22	24 1.78	0 0.00	0 0.00	1350
A	9 0.67	0 0.00	0 0.00	6 0.44	1335 98.89	0 0.00	0 0.00	1350
BS	0 0.00	0 0.00	0 0.00	9 0.67	111 8.22	1230 91.11	0 0.00	1350
W	6 0.44	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1422	1425	1371	1188	1470	1230	1344	9450

Statistics for Table of Reference by Classified

Test of Symmetry

Statistic (S)	322.8469
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8696	0.0038	0.8622	0.8770
Weighted Kappa	0.9470	0.0017	0.9436	0.9504

Sample Size = 9450

Pixel-Level
Uncorrected v. Regression Corrected

The FREQ Procedure

Table 1 of Reference by Classified
Controlling for Regression

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	446 99.11	0 0.00	3 0.67	1 0.22	0 0.00	0 0.00	0 0.00	450
BS	28 6.22	419 93.11	0 0.00	3 0.67	0 0.00	0 0.00	0 0.00	450
CCF	0 0.00	0 0.00	422 93.78	0 0.00	0 0.00	28 6.22	0 0.00	450
G	6 1.33	0 0.00	5 1.11	385 85.56	51 11.33	3 0.67	0 0.00	450
P	0 0.00	0 0.00	0 0.00	2 0.44	419 93.11	29 6.44	0 0.00	450
SF	0 0.00	0 0.00	61 13.56	0 0.00	33 7.33	356 79.11	0 0.00	450
W	0 0.00	0 0.00	2 0.44	0 0.00	0 0.00	0 0.00	448 99.56	450
Total	480	419	493	391	503	416	448	3150

Statistics for Table 1 of Reference by Classified
Controlling for Regression

Test of Symmetry

Statistic (S)	105.3673
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9056	0.0057	0.8944	0.9167
Weighted Kappa	0.9359	0.0045	0.9272	0.9447

Sample Size = 3150

Pixel-Level
Uncorrected v. Regression Corrected

The FREQ Procedure

Table 2 of Reference by Classified
Controlling for Uncorrected

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	445 98.89	0 0.00	4 0.89	1 0.22	0 0.00	0 0.00	0 0.00	450
BS	28 6.22	419 93.11	0 0.00	3 0.67	0 0.00	0 0.00	0 0.00	450
CCF	0 0.00	0 0.00	418 92.89	0 0.00	0 0.00	32 7.11	0 0.00	450
G	6 1.33	0 0.00	2 0.44	391 86.89	49 10.89	2 0.44	0 0.00	450
P	0 0.00	0 0.00	0 0.00	2 0.44	405 90.00	43 9.56	0 0.00	450
SF	0 0.00	0 0.00	61 13.56	0 0.00	19 4.22	370 82.22	0 0.00	450
W	0 0.00	0 0.00	2 0.44	0 0.00	0 0.00	0 0.00	448 99.56	450
Total	479	419	487	397	473	447	448	3150

Statistics for Table 2 of Reference by Classified
Controlling for Uncorrected

Test of Symmetry

Statistic (S)	106.2185
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9059	0.0057	0.8948	0.9170
Weighted Kappa	0.9351	0.0045	0.9262	0.9440

Sample Size = 3150

Pixel-Level
Uncorrected v. Regression Corrected

The FREQ Procedure

Summary Statistics for Reference by Classified

Overall Kappa Coefficients

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9057	0.0040	0.8979	0.9136
Weighted Kappa	0.9355	0.0032	0.9293	0.9418

Tests for Equal Kappa Coefficients

Statistic	Chi-Square	DF	Pr > ChiSq
Simple Kappa	0.0021	1	0.9631
Weighted Kappa	0.0185	1	0.8918

Total Sample Size = 6300

Pixel-Level
Uncorrected v. COST Corrected

The FREQ Procedure

Table 1 of Reference by Classified
Controlling for COST

Reference		Classified						
Frequency								
Row Pct	A	BS	CCF	G	P	SF	W	Total
A	446 99.11	0 0.00	3 0.67	1 0.22	0 0.00	0 0.00	0 0.00	450
BS	31 6.89	416 92.44	0 0.00	3 0.67	0 0.00	0 0.00	0 0.00	450
CCF	0 0.00	0 0.00	412 91.56	0 0.00	0 0.00	38 8.44	0 0.00	450
G	6 1.33	0 0.00	2 0.44	392 87.11	48 10.67	2 0.44	0 0.00	450
P	0 0.00	0 0.00	0 0.00	2 0.44	405 90.00	43 9.56	0 0.00	450
SF	0 0.00	0 0.00	60 13.33	0 0.00	20 4.44	370 82.22	0 0.00	450
W	0 0.00	0 0.00	2 0.44	0 0.00	0 0.00	0 0.00	448 99.56	450
Total	483	416	479	398	473	453	448	3150

Statistics for Table 1 of Reference by Classified
Controlling for COST

Test of Symmetry

Statistic (S)	102.2270
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9033	0.0057	0.8921	0.9146
Weighted Kappa	0.9329	0.0046	0.9238	0.9420

Sample Size = 3150

Pixel-Level
Uncorrected v. COST Corrected

The FREQ Procedure

Table 2 of Reference by Classified
Controlling for Uncorrected

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	445 98.89	0 0.00	4 0.89	1 0.22	0 0.00	0 0.00	0 0.00	450
BS	28 6.22	419 93.11	0 0.00	3 0.67	0 0.00	0 0.00	0 0.00	450
CCF	0 0.00	0 0.00	418 92.89	0 0.00	0 0.00	32 7.11	0 0.00	450
G	6 1.33	0 0.00	2 0.44	391 86.89	49 10.89	2 0.44	0 0.00	450
P	0 0.00	0 0.00	0 0.00	2 0.44	405 90.00	43 9.56	0 0.00	450
SF	0 0.00	0 0.00	61 13.56	0 0.00	19 4.22	370 82.22	0 0.00	450
W	0 0.00	0 0.00	2 0.44	0 0.00	0 0.00	0 0.00	448 99.56	450
Total	479	419	487	397	473	447	448	3150

Statistics for Table 2 of Reference by Classified
Controlling for Uncorrected

Test of Symmetry

Statistic (S)	106.2185
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9059	0.0057	0.8948	0.9170
Weighted Kappa	0.9351	0.0045	0.9262	0.9440

Sample Size = 3150

Pixel-Level
Uncorrected v. COST Corrected

The FREQ Procedure

Summary Statistics for Reference by Classified

Overall Kappa Coefficients

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9046	0.0040	0.8968	0.9125
Weighted Kappa	0.9340	0.0032	0.9277	0.9404

Tests for Equal Kappa Coefficients

Statistic	Chi-Square	DF	Pr > ChiSq
Simple Kappa	0.1037	1	0.7475
Weighted Kappa	0.1112	1	0.7388

Total Sample Size = 6300

Pixel-Level
COST v. Regression Corrected

The FREQ Procedure

Table 1 of Reference by Classified
Controlling for COST

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	446 99.11	0 0.00	3 0.67	1 0.22	0 0.00	0 0.00	0 0.00	450
BS	31 6.89	416 92.44	0 0.00	3 0.67	0 0.00	0 0.00	0 0.00	450
CCF	0 0.00	0 0.00	412 91.56	0 0.00	0 0.00	38 8.44	0 0.00	450
G	6 1.33	0 0.00	2 0.44	392 87.11	48 10.67	2 0.44	0 0.00	450
P	0 0.00	0 0.00	0 0.00	2 0.44	405 90.00	43 9.56	0 0.00	450
SF	0 0.00	0 0.00	60 13.33	0 0.00	20 4.44	370 82.22	0 0.00	450
W	0 0.00	0 0.00	2 0.44	0 0.00	0 0.00	0 0.00	448 99.56	450
Total	483	416	479	398	473	453	448	3150

Statistics for Table 1 of Reference by Classified
Controlling for COST

Test of Symmetry

Statistic (S)	102.2270
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9033	0.0057	0.8921	0.9146
Weighted Kappa	0.9329	0.0046	0.9238	0.9420

Sample Size = 3150

Pixel-Level
COST v. Regression Corrected

The FREQ Procedure

Table 2 of Reference by Classified
Controlling for Regression

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	446 99.11	0 0.00	3 0.67	1 0.22	0 0.00	0 0.00	0 0.00	450
BS	28 6.22	419 93.11	0 0.00	3 0.67	0 0.00	0 0.00	0 0.00	450
CCF	0 0.00	0 0.00	422 93.78	0 0.00	0 0.00	28 6.22	0 0.00	450
G	6 1.33	0 0.00	5 1.11	385 85.56	51 11.33	3 0.67	0 0.00	450
P	0 0.00	0 0.00	0 0.00	2 0.44	419 93.11	29 6.44	0 0.00	450
SF	0 0.00	0 0.00	61 13.56	0 0.00	33 7.33	356 79.11	0 0.00	450
W	0 0.00	0 0.00	2 0.44	0 0.00	0 0.00	0 0.00	448 99.56	450
Total	480	419	493	391	503	416	448	3150

Statistics for Table 2 of Reference by Classified
Controlling for Regression

Test of Symmetry

Statistic (S)	105.3673
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9056	0.0057	0.8944	0.9167
Weighted Kappa	0.9359	0.0045	0.9272	0.9447

Sample Size = 3150

Pixel-Level
COST v. Regression Corrected

The FREQ Procedure

Summary Statistics for Reference by Classified

Overall Kappa Coefficients

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9045	0.0040	0.8966	0.9124
Weighted Kappa	0.9345	0.0032	0.9282	0.9408

Tests for Equal Kappa Coefficients

Statistic	Chi-Square	DF	Pr > ChiSq
Simple Kappa	0.0761	1	0.7827
Weighted Kappa	0.2212	1	0.6382

Total Sample Size = 6300

Polygon-Level
Uncorrected v. Regression Corrected

The FREQ Procedure

Table 1 of Reference by Classified
Controlling for Regression

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1338 99.11	0 0.00	9 0.67	3 0.22	0 0.00	0 0.00	0 0.00	1350
BS	87 6.44	1254 92.89	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1263 93.56	0 0.00	0 0.00	87 6.44	0 0.00	1350
G	18 1.33	0 0.00	15 1.11	1146 84.89	162 12.00	9 0.67	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1251 92.67	93 6.89	0 0.00	1350
SF	0 0.00	0 0.00	195 14.44	0 0.00	102 7.56	1053 78.00	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1443	1254	1488	1164	1515	1242	1344	9450

Statistics for Table 1 of Reference by Classified
Controlling for Regression

Test of Symmetry

Statistic (S)	332.3485
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9011	0.0033	0.8946	0.9077
Weighted Kappa	0.9329	0.0027	0.9277	0.9381

Sample Size = 9450

Polygon-Level
Uncorrected v. Regression Corrected

The FREQ Procedure

Table 2 of Reference by Classified
Controlling for Uncorrected

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1335 98.89	0 0.00	12 0.89	3 0.22	0 0.00	0 0.00	0 0.00	1350
BS	87 6.44	1254 92.89	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1236 91.56	0 0.00	0 0.00	114 8.44	0 0.00	1350
G	18 1.33	0 0.00	6 0.44	1170 86.67	150 11.11	6 0.44	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1203 89.11	141 10.44	0 0.00	1350
SF	0 0.00	0 0.00	189 14.00	0 0.00	66 4.89	1095 81.11	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1440	1254	1449	1188	1419	1356	1344	9450

Statistics for Table 2 of Reference by Classified
Controlling for Uncorrected

Test of Symmetry

Statistic (S)	315.3756
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8996	0.0034	0.8930	0.9062
Weighted Kappa	0.9305	0.0027	0.9252	0.9359

Sample Size = 9450

Polygon-Level
Uncorrected v. Regression Corrected

The FREQ Procedure

Summary Statistics for Reference by Classified

Overall Kappa Coefficients

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9004	0.0024	0.8957	0.9050
Weighted Kappa	0.9317	0.0019	0.9280	0.9355

Tests for Equal Kappa Coefficients

Statistic	Chi-Square	DF	Pr > ChiSq
Simple Kappa	0.0976	1	0.7547
Weighted Kappa	0.3874	1	0.5337

Total Sample Size = 18900

Polygon-Level
Uncorrected v. COST Corrected

The FREQ Procedure

Table 1 of Reference by Classified
Controlling for COST

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1338 99.11	0 0.00	9 0.67	3 0.22	0 0.00	0 0.00	0 0.00	1350
BS	93 6.89	1248 92.44	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1236 91.56	0 0.00	0 0.00	114 8.44	0 0.00	1350
G	18 1.33	0 0.00	6 0.44	1170 86.67	150 11.11	6 0.44	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1191 88.22	153 11.33	0 0.00	1350
SF	0 0.00	0 0.00	186 13.78	0 0.00	63 4.67	1101 81.56	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1449	1248	1443	1188	1404	1374	1344	9450

Statistics for Table 1 of Reference by Classified
Controlling for COST

Test of Symmetry

Statistic (S)	327.4174
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8985	0.0034	0.8919	0.9051
Weighted Kappa	0.9306	0.0027	0.9253	0.9359

Sample Size = 9450

Polygon-Level
Uncorrected v. COST Corrected

The FREQ Procedure

Table 2 of Reference by Classified
Controlling for Uncorrected

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1335 98.89	0 0.00	12 0.89	3 0.22	0 0.00	0 0.00	0 0.00	1350
BS	87 6.44	1254 92.89	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1236 91.56	0 0.00	0 0.00	114 8.44	0 0.00	1350
G	18 1.33	0 0.00	6 0.44	1170 86.67	150 11.11	6 0.44	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1203 89.11	141 10.44	0 0.00	1350
SF	0 0.00	0 0.00	189 14.00	0 0.00	66 4.89	1095 81.11	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1440	1254	1449	1188	1419	1356	1344	9450

Statistics for Table 2 of Reference by Classified
Controlling for Uncorrected

Test of Symmetry

Statistic (S)	315.3756
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8996	0.0034	0.8930	0.9062
Weighted Kappa	0.9305	0.0027	0.9252	0.9359

Sample Size = 9450

Polygon-Level
Uncorrected v. COST Corrected

The FREQ Procedure

Summary Statistics for Reference by Classified

Overall Kappa Coefficients

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8991	0.0024	0.8944	0.9038
Weighted Kappa	0.9306	0.0019	0.9268	0.9343

Tests for Equal Kappa Coefficients

Statistic	Chi-Square	DF	Pr > ChiSq
Simple Kappa	0.0543	1	0.8158
Weighted Kappa	0.0002	1	0.9897

Total Sample Size = 18900

Polygon-Level
COST v. Regression Corrected

The FREQ Procedure

Table 1 of Reference by Classified
Controlling for COST

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1338 99.11	0 0.00	9 0.67	3 0.22	0 0.00	0 0.00	0 0.00	1350
BS	93 6.89	1248 92.44	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1236 91.56	0 0.00	0 0.00	114 8.44	0 0.00	1350
G	18 1.33	0 0.00	6 0.44	1170 86.67	150 11.11	6 0.44	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1191 88.22	153 11.33	0 0.00	1350
SF	0 0.00	0 0.00	186 13.78	0 0.00	63 4.67	1101 81.56	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1449	1248	1443	1188	1404	1374	1344	9450

Statistics for Table 1 of Reference by Classified
Controlling for COST

Test of Symmetry

Statistic (S) 327.4174
DF 21
Pr > S <.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8985	0.0034	0.8919	0.9051
Weighted Kappa	0.9306	0.0027	0.9253	0.9359

Sample Size = 9450

Polygon-Level
COST v. Regression Corrected

The FREQ Procedure

Table 2 of Reference by Classified
Controlling for Regression

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1338 99.11	0 0.00	9 0.67	3 0.22	0 0.00	0 0.00	0 0.00	1350
BS	87 6.44	1254 92.89	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1263 93.56	0 0.00	0 0.00	87 6.44	0 0.00	1350
G	18 1.33	0 0.00	15 1.11	1146 84.89	162 12.00	9 0.67	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1251 92.67	93 6.89	0 0.00	1350
SF	0 0.00	0 0.00	195 14.44	0 0.00	102 7.56	1053 78.00	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1443	1254	1488	1164	1515	1242	1344	9450

Statistics for Table 2 of Reference by Classified
Controlling for Regression

Test of Symmetry

Statistic (S)	332.3485
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.9011	0.0033	0.8946	0.9077
Weighted Kappa	0.9329	0.0027	0.9277	0.9381

Sample Size = 9450

Polygon-Level
COST v. Regression Corrected

The FREQ Procedure

Summary Statistics for Reference by Classified

Overall Kappa Coefficients

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8998	0.0024	0.8952	0.9045
Weighted Kappa	0.9318	0.0019	0.9281	0.9355

Tests for Equal Kappa Coefficients

Statistic	Chi-Square	DF	Pr > ChiSq
Simple Kappa	0.2975	1	0.5854
Weighted Kappa	0.3730	1	0.5414

Total Sample Size = 18900

Region-Level
Uncorrected v. Regression Corrected

The FREQ Procedure

Table 1 of Reference by Classified
Controlling for Regression

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1338 99.11	0 0.00	9 0.67	3 0.22	0 0.00	0 0.00	0 0.00	1350
BS	90 6.67	1251 92.67	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1257 93.11	0 0.00	0 0.00	93 6.89	0 0.00	1350
G	18 1.33	0 0.00	15 1.11	1140 84.44	165 12.22	12 0.89	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1251 92.67	93 6.89	0 0.00	1350
SF	0 0.00	0 0.00	189 14.00	0 0.00	108 8.00	1053 78.00	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1446	1251	1476	1158	1524	1251	1344	9450

Statistics for Table 1 of Reference by Classified
Controlling for Regression

Test of Symmetry

Statistic (S)	333.3566
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8993	0.0034	0.8927	0.9059
Weighted Kappa	0.9321	0.0027	0.9269	0.9373

Sample Size = 9450

Region-Level
Uncorrected v. Regression Corrected

The FREQ Procedure

Table 2 of Reference by Classified
Controlling for Uncorrected

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1338 99.11	0 0.00	9 0.67	3 0.22	0 0.00	0 0.00	0 0.00	1350
BS	87 6.44	1254 92.89	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1224 90.67	0 0.00	0 0.00	126 9.33	0 0.00	1350
G	21 1.56	0 0.00	12 0.89	1170 86.67	141 10.44	6 0.44	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1167 86.44	177 13.11	0 0.00	1350
SF	0 0.00	0 0.00	183 13.56	0 0.00	69 5.11	1098 81.33	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1446	1254	1434	1188	1377	1407	1344	9450

Statistics for Table 2 of Reference by Classified
Controlling for Uncorrected

Test of Symmetry

Statistic (S)	324.4088
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8944	0.0034	0.8877	0.9012
Weighted Kappa	0.9280	0.0027	0.9226	0.9334

Sample Size = 9450

Region-Level
Uncorrected v. Regression Corrected

The FREQ Procedure

Summary Statistics for Reference by Classified

Overall Kappa Coefficients

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8969	0.0024	0.8922	0.9016
Weighted Kappa	0.9301	0.0019	0.9264	0.9339

Tests for Equal Kappa Coefficients

Statistic	Chi-Square	DF	Pr > ChiSq
Simple Kappa	0.9994	1	0.3175
Weighted Kappa	1.1415	1	0.2853

Total Sample Size = 18900

Region-Level
Uncorrected v. COST Corrected

The FREQ Procedure

Table 1 of Reference by Classified
Controlling for COST

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1338 99.11	0 0.00	9 0.67	3 0.22	0 0.00	0 0.00	0 0.00	1350
BS	87 6.44	1254 92.89	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1209 89.56	0 0.00	0 0.00	141 10.44	0 0.00	1350
G	18 1.33	0 0.00	12 0.89	1173 86.89	141 10.44	6 0.44	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1164 86.22	180 13.33	0 0.00	1350
SF	0 0.00	0 0.00	186 13.78	0 0.00	81 6.00	1083 80.22	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1443	1254	1422	1191	1386	1410	1344	9450

Statistics for Table 1 of Reference by Classified
Controlling for COST

Test of Symmetry

Statistic (S)	307.4383
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8907	0.0035	0.8839	0.8976
Weighted Kappa	0.9252	0.0028	0.9197	0.9307

Sample Size = 9450

Region-Level
Uncorrected v. COST Corrected

The FREQ Procedure

Table 2 of Reference by Classified
Controlling for Uncorrected

Reference		Classified						
Frequency								
Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1338 99.11	0 0.00	9 0.67	3 0.22	0 0.00	0 0.00	0 0.00	1350
BS	87 6.44	1254 92.89	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1224 90.67	0 0.00	0 0.00	126 9.33	0 0.00	1350
G	21 1.56	0 0.00	12 0.89	1170 86.67	141 10.44	6 0.44	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1167 86.44	177 13.11	0 0.00	1350
SF	0 0.00	0 0.00	183 13.56	0 0.00	69 5.11	1098 81.33	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1446	1254	1434	1188	1377	1407	1344	9450

Statistics for Table 2 of Reference by Classified
Controlling for Uncorrected

Test of Symmetry

Statistic (S)	324.4088
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8944	0.0034	0.8877	0.9012
Weighted Kappa	0.9280	0.0027	0.9226	0.9334

Sample Size = 9450

Region-Level
Uncorrected v. COST Corrected

The FREQ Procedure

Summary Statistics for Reference by Classified

Overall Kappa Coefficients

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8926	0.0025	0.8878	0.8974
Weighted Kappa	0.9266	0.0020	0.9228	0.9305

Tests for Equal Kappa Coefficients

Statistic	Chi-Square	DF	Pr > ChiSq
Simple Kappa	0.5698	1	0.4503
Weighted Kappa	0.4993	1	0.4798

Total Sample Size = 18900

Region-Level
COST v. Regression Corrected

The FREQ Procedure

Table 1 of Reference by Classified
Controlling for COST

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1338 99.11	0 0.00	9 0.67	3 0.22	0 0.00	0 0.00	0 0.00	1350
BS	87 6.44	1254 92.89	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1209 89.56	0 0.00	0 0.00	141 10.44	0 0.00	1350
G	18 1.33	0 0.00	12 0.89	1173 86.89	141 10.44	6 0.44	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1164 86.22	180 13.33	0 0.00	1350
SF	0 0.00	0 0.00	186 13.78	0 0.00	81 6.00	1083 80.22	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1443	1254	1422	1191	1386	1410	1344	9450

Statistics for Table 1 of Reference by Classified
Controlling for COST

Test of Symmetry

Statistic (S)	307.4383
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8907	0.0035	0.8839	0.8976
Weighted Kappa	0.9252	0.0028	0.9197	0.9307

Sample Size = 9450

Region-Level
COST v. Regression Corrected

The FREQ Procedure

Table 2 of Reference by Classified
Controlling for Regression

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1338 99.11	0 0.00	9 0.67	3 0.22	0 0.00	0 0.00	0 0.00	1350
BS	90 6.67	1251 92.67	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1257 93.11	0 0.00	0 0.00	93 6.89	0 0.00	1350
G	18 1.33	0 0.00	15 1.11	1140 84.44	165 12.22	12 0.89	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1251 92.67	93 6.89	0 0.00	1350
SF	0 0.00	0 0.00	189 14.00	0 0.00	108 8.00	1053 78.00	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1446	1251	1476	1158	1524	1251	1344	9450

Statistics for Table 2 of Reference by Classified
Controlling for Regression

Test of Symmetry

Statistic (S)	333.3566
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8993	0.0034	0.8927	0.9059
Weighted Kappa	0.9321	0.0027	0.9269	0.9373

Sample Size = 9450

Region-Level
COST v. Regression Corrected

The FREQ Procedure

Summary Statistics for Reference by Classified

Overall Kappa Coefficients

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8952	0.0024	0.8904	0.8999
Weighted Kappa	0.9288	0.0019	0.9251	0.9326

Tests for Equal Kappa Coefficients

Statistic	Chi-Square	DF	Pr > ChiSq
Simple Kappa	3.0785	1	0.0793
Weighted Kappa	3.1524	1	0.0758

Total Sample Size = 18900

Image-Level
Uncorrected v. Regression Corrected

The FREQ Procedure

Table 1 of Reference by Classified
Controlling for Regression

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1338 99.11	0 0.00	6 0.44	6 0.44	0 0.00	0 0.00	0 0.00	1350
BS	87 6.44	1254 92.89	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1203 89.11	0 0.00	0 0.00	147 10.89	0 0.00	1350
G	18 1.33	0 0.00	15 1.11	1104 81.78	195 14.44	18 1.33	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1254 92.89	90 6.67	0 0.00	1350
SF	0 0.00	0 0.00	195 14.44	0 0.00	120 8.89	1035 76.67	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1443	1254	1425	1125	1569	1290	1344	9450

Statistics for Table 1 of Reference by Classified
Controlling for Regression

Test of Symmetry

Statistic (S)	335.7390
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8867	0.0036	0.8797	0.8936
Weighted Kappa	0.9215	0.0029	0.9158	0.9271

Sample Size = 9450

Image-Level
Uncorrected v. Regression Corrected

The FREQ Procedure

Table 2 of Reference by Classified
Controlling for Uncorrected

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1341 99.33	0 0.00	3 0.22	6 0.44	0 0.00	0 0.00	0 0.00	1350
BS	102 7.56	1239 91.78	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1230 91.11	0 0.00	0 0.00	120 8.89	0 0.00	1350
G	21 1.56	0 0.00	9 0.67	1191 88.22	123 9.11	6 0.44	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1092 80.89	252 18.67	0 0.00	1350
SF	0 0.00	0 0.00	171 12.67	0 0.00	126 9.33	1053 78.00	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1464	1239	1419	1212	1341	1431	1344	9450

Statistics for Table 2 of Reference by Classified
Controlling for Uncorrected

Test of Symmetry

Statistic (S)	300.3878
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8815	0.0036	0.8744	0.8886
Weighted Kappa	0.9249	0.0027	0.9196	0.9302

Sample Size = 9450

Image-Level
Uncorrected v. Regression Corrected

The FREQ Procedure

Summary Statistics for Reference by Classified

Overall Kappa Coefficients

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8841	0.0025	0.8792	0.8891
Weighted Kappa	0.9233	0.0020	0.9194	0.9272

Tests for Equal Kappa Coefficients

Statistic	Chi-Square	DF	Pr > ChiSq
Simple Kappa	1.0438	1	0.3069
Weighted Kappa	0.7403	1	0.3896

Total Sample Size = 18900

Image-Level
Uncorrected v. COST Corrected

The FREQ Procedure

Table 1 of Reference by Classified
Controlling for COST

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1335 98.89	0 0.00	9 0.67	6 0.44	0 0.00	0 0.00	0 0.00	1350
BS	111 8.22	1230 91.11	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1212 89.78	0 0.00	0 0.00	138 10.22	0 0.00	1350
G	24 1.78	0 0.00	9 0.67	1164 86.22	144 10.67	9 0.67	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	9 0.67	1086 80.44	255 18.89	0 0.00	1350
SF	0 0.00	0 0.00	186 13.78	0 0.00	141 10.44	1023 75.78	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1470	1230	1422	1188	1371	1425	1344	9450

Statistics for Table 1 of Reference by Classified
Controlling for COST

Test of Symmetry

Statistic (S)	322.8469
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8696	0.0038	0.8622	0.8770
Weighted Kappa	0.9168	0.0029	0.9111	0.9224

Sample Size = 9450

Image-Level
Uncorrected v. COST Corrected

The FREQ Procedure

Table 2 of Reference by Classified
Controlling for Uncorrected

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1341 99.33	0 0.00	3 0.22	6 0.44	0 0.00	0 0.00	0 0.00	1350
BS	102 7.56	1239 91.78	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1230 91.11	0 0.00	0 0.00	120 8.89	0 0.00	1350
G	21 1.56	0 0.00	9 0.67	1191 88.22	123 9.11	6 0.44	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1092 80.89	252 18.67	0 0.00	1350
SF	0 0.00	0 0.00	171 12.67	0 0.00	126 9.33	1053 78.00	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1464	1239	1419	1212	1341	1431	1344	9450

Statistics for Table 2 of Reference by Classified
Controlling for Uncorrected

Test of Symmetry

Statistic (S)	300.3878
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8815	0.0036	0.8744	0.8886
Weighted Kappa	0.9249	0.0027	0.9196	0.9302

Sample Size = 9450

Image-Level
Uncorrected v. COST Corrected

The FREQ Procedure

Summary Statistics for Reference by Classified

Overall Kappa Coefficients

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8758	0.0026	0.8707	0.8809
Weighted Kappa	0.9210	0.0020	0.9172	0.9249

Tests for Equal Kappa Coefficients

Statistic	Chi-Square	DF	Pr > ChiSq
Simple Kappa	5.1190	1	0.0237
Weighted Kappa	4.2203	1	0.0399

Total Sample Size = 18900

Image-Level
COST v. Regression Corrected

The FREQ Procedure

Table 1 of Reference by Classified
Controlling for COST

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1335 98.89	0 0.00	9 0.67	6 0.44	0 0.00	0 0.00	0 0.00	1350
BS	111 8.22	1230 91.11	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1212 89.78	0 0.00	0 0.00	138 10.22	0 0.00	1350
G	24 1.78	0 0.00	9 0.67	1164 86.22	144 10.67	9 0.67	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	9 0.67	1086 80.44	255 18.89	0 0.00	1350
SF	0 0.00	0 0.00	186 13.78	0 0.00	141 10.44	1023 75.78	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1470	1230	1422	1188	1371	1425	1344	9450

Statistics for Table 1 of Reference by Classified
Controlling for COST

Test of Symmetry

Statistic (S)	322.8469
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8696	0.0038	0.8622	0.8770
Weighted Kappa	0.9168	0.0029	0.9111	0.9224

Sample Size = 9450

Image-Level
COST v. Regression Corrected

The FREQ Procedure

Table 2 of Reference by Classified
Controlling for Regression

Reference Classified

Frequency Row Pct	A	BS	CCF	G	P	SF	W	Total
A	1338 99.11	0 0.00	6 0.44	6 0.44	0 0.00	0 0.00	0 0.00	1350
BS	87 6.44	1254 92.89	0 0.00	9 0.67	0 0.00	0 0.00	0 0.00	1350
CCF	0 0.00	0 0.00	1203 89.11	0 0.00	0 0.00	147 10.89	0 0.00	1350
G	18 1.33	0 0.00	15 1.11	1104 81.78	195 14.44	18 1.33	0 0.00	1350
P	0 0.00	0 0.00	0 0.00	6 0.44	1254 92.89	90 6.67	0 0.00	1350
SF	0 0.00	0 0.00	195 14.44	0 0.00	120 8.89	1035 76.67	0 0.00	1350
W	0 0.00	0 0.00	6 0.44	0 0.00	0 0.00	0 0.00	1344 99.56	1350
Total	1443	1254	1425	1125	1569	1290	1344	9450

Statistics for Table 2 of Reference by Classified
Controlling for Regression

Test of Symmetry

Statistic (S)	335.7390
DF	21
Pr > S	<.0001

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8867	0.0036	0.8797	0.8936
Weighted Kappa	0.9215	0.0029	0.9158	0.9271

Sample Size = 9450

Image-Level
COST v. Regression Corrected

The FREQ Procedure

Summary Statistics for Reference by Classified

Overall Kappa Coefficients

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.8787	0.0026	0.8736	0.8838
Weighted Kappa	0.9191	0.0020	0.9151	0.9231

Tests for Equal Kappa Coefficients

Statistic	Chi-Square	DF	Pr > ChiSq
Simple Kappa	10.7863	1	0.0010
Weighted Kappa	1.3469	1	0.2458

Total Sample Size = 18900