UNIVERSITA' DEGLI STUDI
DI ROMA TOR VERGATA

Dipartimento di Ingegneria Civile e Ingegneria Informatica

GeoInformation Doctorate

Monitoring Forests:
Parameters Estimation and Vegetation Classification with Multisource Remote Sensing Data

A thesis submitted in partial fulfillment for the PhD degree
(Dottore di Ricerca)

Candidate**: MSc. Gaia Vaglio Laurin**

Supervisors: **Prof. Leila Guerriero, Ing. Fabio Del Frate**

January 2014

Abstract

The work presented in this thesis covers two main areas of forest research with remote sensing data: the classification of forested landscapes, conducted in a tropical and an Alpine montane region, and the estimation of parameters of forestry interest, namely above ground biomass and the Shanon-Wiener arboreal diversity index.

The thesis first introduces the need of monitoring forested landscapes, their changes and their resources, illustrating objectives, motivations and areas of innovation in the present research. The material and methods adopted in the research, with specifications on the study areas, and a short thesis outline, are also presented in the Introduction chapter. A short overview of techniques and sensors used in classification and estimation of the two forest parameters of interest is presented in Chapter 2, followed by the identification of some of the most recent challenges in remote sensing applied to forest studies, which have been object of the present thesis. In Chapter 3 the first case study is introduced, as published in Remote Sensing of Environment, addressing the integration of airborne lidar and vegetation types derived from aerial photography for mapping aboveground biomass. Chapter 4 presents the research paper as published by the International Journal of Remote Sensing, dealing with discrimination of vegetation types in alpine sites with ALOS PALSAR, RADARSAT-2 and lidar derived information. Chapter 5 illustrates the third case study, which is about optical and SAR sensor synergies for forest and land cover mapping in a tropical site in West Africa, according to the paper published in the International Journal of Applied Earth Observation and Geoinformation. In Chapter 6, the case study addresses the aboveground biomass estimation in an Africa tropical forest with lidar and hyperspectral data, a paper at its second review in the ISPRS Journal of Photogrammetry and Remote Sensing. The last case study is presented in Chapter 7, and deals with biodiversity mapping in a tropical West African forest with hyperspectral data, and is also a paper at its second review in PlosONE. Finally, the research summary and conclusions are presented in Chapter 8.

Acknowledgments

These research years have been among the most exciting and interesting of my life, but this study effort was only possible thanks to the many persons who supported me. First, all my colleagues of the EO lab, with whom I shared ideas, hypothesis, laughs, hopes, troubles, and more and more for a long time: Lino, Giorgio, Chiara, Irene, Antonio, Matteo, Daniele, Reza, Ruggero, Andrei, Simone, Zina, Cristina. Three persons really inspired me more than others, and tried to transmit me the 'sense' of scientific research: Riccardo Valentini (La Tuscia University) who guided me with his vision and pushed and supported me toward invaluable scientific experiences; Qi Chen (University of Hawaii) who hosted and assisted me in all the ways before, during and after my months at the University of Hawaii; David Coomes (University of Cambridge) who always provided opportunity for research, exchange and collaboration. All of them gave me their trust, patience and time, and I really hope I will have a chance to keep on sharing and collaborating with them. My supervisors at the University of Rome Tor Vergata, Leila Guerriero and Fabio Del Frate, offered continuous assistance, encouragement and support during these years. Finally I have to thank my family: all my love to my mother who followed me on the other side of the globe and to my daughter with her sweetness and patience.

Table of Contents

Chapter 1

Introduction

Forest is defined as land spanning more than 0.5 hectares with trees higher than 5 meters and a canopy cover of more than 10 percent, or trees able to reach these thresholds in situ. It does not include land that is predominantly under agricultural or urban land use (FAO 2010). Forest ecosystems are characterized by the dominant vegetation type, stand structure, climate, soil type, and topography; local climate determine the biome level division in tropical, boreal and temperate forests.
In the last three decades the international community has debated on climate change and global warming, and since the 1994 the United Nation Framework Convention on Climate Change entered into force, with the ultimate aim of preventing dangerous human interference with the climate system. Carbon dioxide has constantly increased in the last decades (Fig. 1) and the CO_2 emissions are the first responsible for greenhouse effects, modifying the radiative balance of the earth, which results in increased heat absorbed and trapped into the atmosphere and thus in global warming (NOAA 2007) .
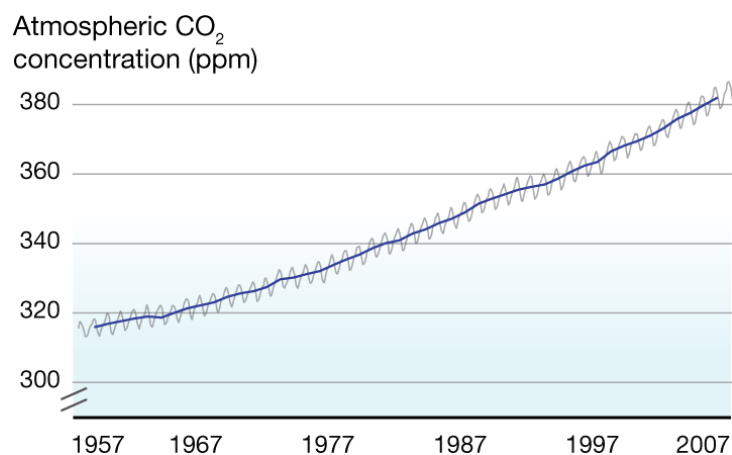
Figure 1 – The Keeling curve: atmospheric carbon dioxide concentration in parts per million in the last 50 years (NOAA 2007).

Oceans are the major sinks of carbon on earth, but soil and vegetation are the first responsible – through photosynthesis – of CO_2 removal from atmosphere (IPCC/GRID-Arendal 2001), with about half of forest biomass made by carbon. Deforestation is considered the responsible of about 10-20% of global annual greenhouse gases emissions.

Considering the constraints of reducing emissions from industrialized countries and the increasing emission from emerging economies (i.e. Brazil, India), to avoid deforestation and forest degradation is possibly the best option to quickly and efficiently reduce carbon emissions and mitigate on-going climate change. This is one of the main reasons behind the increase in forest studies in recent years.

Monitoring of forest resources is therefore essential and it can be realized by means of retrieval and classification of remote sensing data, which allow generalizing to large areas the local in situ observations. Estimation of forest biophysical and ecological parameters, among which are found woody biomass and forest biodiversity, is important for forest inventory, management and for scientific purposes (Parresol, 1999). Classification of forests, discrimination of different forest types, and mapping their extent is also essential to management and conservation activities and to assess degradation. Both retrieval and classification activities, based on remote sensing data, are needed for the full understanding of ecosystem functioning in a changing climate scenario, and its management and conservation.

1.1 Thesis objectives, motivations, innovation

The main goal of this research is to innovatively use remote sensing data to produce information on important forest characteristics, such as forest parameters and classification into distinguishable vegetation classes.

The main motivation behind this research is the desire to contribute to forest conservation by means of improving methods and tools for its monitoring, bringing ecology and engineering issues closer. Forests are complex ecosystems, having different and often site-specific characteristics. The selection of multiple case studies, dealing with both classification and retrieval in different biomes, has been a good exercise to

face the multitude of issues and monitoring needs that forest resource demand. Most forest studies are located in temperate and boreal regions, and another motivation of this research is to help to enrich scientific studies over tropical rainforests, especially in Africa. In fact, few references are found about conducting forest studies in Africa on the basis of remote sensing data. Even less researches are interested to the West African area, where the tropical study site of this research is located and which is considered a specific ecological region affected by intense degradation of natural resources due to anthropogenic causes.

The innovation in the research is conducted adopting the following approaches in data analysis:

1.   *Investigating data fusion and integration, thus the combination of different active and passive sensors and ancillary datasets to improve the accuracy of retrieved parameters and of classifications, accordingly to forest characteristics, data availability and specific aims.*

In the different case studies multispectral, hyperspectral, SAR, LiDAR and ancillary data were used and integrated, showing that the simultaneous use of different remote sensing data types or derivates is significantly beneficial. For instance, the integration of optical and SAR data has recently gained interest as it can help solving complex classification problems and monitoring needs. Optical and microwave remote sensing data are complementary to each other: optical data basically measure the physical properties of observed objects, whereas the SAR data provide more information about geometric properties of the objects (Lillesand and Kiefer 1994). SAR data are now widely available, and have an all-weather capability which is a very important feature for tropical cloudy regions. SAR data are also available at different frequencies, which have different abilities to penetrate vegetation and potentially can bring in complementary information on vegetation characteristics.

In the last decade, airborne LiDAR demonstrated to be a valuable tool for forest monitoring, and nowadays costs associated to data acquisitions have been greatly reduced, allowing for surveying entire forested areas or to use LiDAR as a sampling

tool to replace ground truth. In the same way, hyperspectral data use is expected to increase due to the forthcoming launch of new satellite missions. Studies based on these innovative data types can help to better understand their potential and usefulness in the context of forest monitoring.

The use of two different SAR types for vegetation discrimination, having different frequencies and polarizations, with the comparison of the SAR-based results against those obtained by an optical sensor, is illustrated in Chapter 3, in which the integration of satellite data with ancillary data derived from LiDAR is also presented. An example of integration of optical and SAR data is then presented in Chapter 4 for the classification of a complex tropical forested area. LiDAR use to retrieve above ground biomass is presented in Chapter 5, in which integrated data on forest types (derived from aerial imagery) proved to be useful to improve the AGB estimates. The fusion of LiDAR and hyperspectral data is attempted in Chapter 6, as a way to improve AGB estimation in a tropical forest, while an innovative use of hyperspectral data is carried out for the estimation of the arboreal diversity index, the Shannon-Wiener index, which is presented in Chapter 7.

2.   *Adopting different statistical modeling tools and classification algorithms, selected to better respond to specific datasets and issues, as a way to improve results.*
In this thesis work, Mixed Effects, Partial Least Squares Regression, and Random Forests were the adopted techniques for estimating AGB and forest biodiversity, while Neural Networks were tested in forest classification and vegetation type discrimination and compared to Maximum Likelihood more traditional approach. These innovative techniques were chosen to face specific problems found in the analysis of the different datasets, which are better detailed in the case studies in following paragraphs.
Mixed Effects was firstly introduced in 1918 by Ronald Fisher (Fisher 1918) to study the correlations of trait values between relatives. Since then, mixed modeling has become a major area of statistical research, with applications in many disciplines where multiple correlated measurements are made on each unit of interest. They are prominently used in research involving human and animal subjects in fields ranging from genetics to marketing, and have also been used in industrial statistics. In the remote sensing area, Mixed-effects models have been recently used to estimate canopy

height from satellite LiDAR (GLAS) data (Chen 2010) and tree diameter from airborne discrete-return LiDAR data (Salas et al., 2010).

Partial Least Squares Regression was introduced by the Swedish statistician Herman Wold, who then developed it with his son, Svante Wold (Wold et al. 2001). Although the original applications were in the social sciences, PLS regression is today most widely used in chemometrics and related areas. It is also used in bioinformatics, sensometrics, neuroscience and anthropology. PLS regression has been previously employed in spectral and chemical analysis of tropical forests (Asner and Martin 2008), for AGB estimation (Lei et al. 2012; Goodenough et al. 2005), and as a method for dealing with large hyperspectral dataset (Peerbay et al. 2013).

Random forests are an ensemble learning method for classification (and regression) . The algorithm for inducing a random forest was developed by Leo Breiman and Adele Cutler in 2001 (Breiman 2001). More recently, major advances in this area have come from Microsoft Research (Criminisi et al. 2001) which incorporate and extend the earlier work from Breiman. Random Forests is an algorithm which became popular in remote sensing studies in recent years (Cutler et al. 2007; Pal 2005).

Artificial neural networks are computational models inspired by animal central nervous systems (in particular the brain) that are capable of machine learning and pattern recognition. Like other machine learning methods, neural networks have been used to solve a wide variety of tasks that are hard to solve using ordinary rule-based programming, including computer vision and speech recognition. Warren McCulloch and Walter Pitts in 1943 firstly created a computational model for neural networks based on mathematics and algorithms (McCulloch and Pitts, 1943). Subsequently, much advancement in the technique was realized by different researchers and Neural Networks have been widely applied to solve complex remote sensing problems (Del Frate and Solimini, 2004; Atkinson and Tatnall, 1997).

For the retrieval of forest parameters, the selection of the modeling approach is usually data-driven, and an initial exploratory analysis of the ground truth and remote sensing data can reveal relevant information to detect data issues, such as multicollinearity or non-linearity which can violate the assumption of traditional statistical regression

techniques. The limited availability of ground truth data, typical in forestry research, represents another issue to take into consideration for model selection. In the present research three different approaches have been adopted to model the biophysical and ecological parameters of interest, with both the purpose of experimenting innovative techniques and finding the best solution to fit the data. Mixed Effects Models (MEM) have been employed to take advantage of the availability of ancillary data on vegetation types (Chapter 3); Partial Least Squares Regression (PLSR) and Random Forests (RF) have been used to deal with a complex set of remote sensing input data and for their ability to provide information on the most relevant remote sensing inputs that contributed to the models (Chapters 6 and 7). Multilinear regression technique has also been adopted as a benchmark for results comparison in two of the three regression case studies.

A MEM is a statistical model containing both fixed effects and random effects. These models are particularly useful where measurements are made on clusters of related statistical units, such as field data collected in plots belonging to different vegetation classes. Linear MEM describe the relationship between a continuous response variable and some covariates that have been measured or observed along with the response, where at least one of the covariates in the model is a categorical variable. This categorical variable may represent the study location, or more generally, the observational unit such as the vegetation type in case study in Chapter 3. Parameters associated with the particular categories of a covariate are sometimes called the "effects". If the set of possible values (or levels in the case of categorical variables) of the covariate is fixed and reproducible we model the covariate using fixed-effects parameters. If the levels that we observed represent a random sample from the set of all possible levels we incorporate random effects. Thus we can distinguish between "fixed-effects parameters", which are indeed parameters in the statistical model, and "random effects", which, strictly speaking, are not parameters: random effects are unobserved random variables. Random effects contribute only to the covariance structure of the data, but their presence often introduces correlations between cases as well. In the case study in Chapter 5, about estimation of above ground biomass based on LiDAR data, the random effects are caused by the presence of different vegetation types, identified in

the ancillary information by means of a vegetation map. Though the fixed effect is the primary interest in most studies or experiments, it is possible to adjust for the covariance structure of the data, in our example the vegetation types, and in this way increase the accuracy of the model by means of incorporation of categorical values. Thus, modeling variance structure is probably the most powerful feature of MEM, which allows correlation among observations. Additionally, MEM are useful when the number of observation is limited and does not allow for setting up a specific model for each cluster of statistical units. Technical literature on MEM include: Pinhero and Bates (2000), Brown and Prescott (2006) and Faraway (2004).

PLSR is a recent technique that generalizes and combines features from principal component analysis and multiple regression. It is particularly useful when there is a very large set of independent variables and multicollinearity issues. Several approaches have been developed to cope with this problem; one approach is to eliminate some predictors using step-wise methods. Another one, called principal component regression, is to perform a principal component analysis of the predictors matrix and then use those principal components as regressors on the dependent variable. The orthogonality of the principal components eliminates the multicollinearity problem, but, the problem of choosing an optimum subset of predictors remains. A possible strategy is to keep only a few of the first components, but they are chosen to explain the independent variables rather than the dependent one, and so there is no guarantee that the principal components, which explain X , are relevant for Y. PLSR instead finds components from X that are also relevant for Y. In the case study illustrated in Chapter 6, the Y variable is above ground biomass, while the X variables are the huge number of bands values and statistical metrics derived by hyperspectral and LiDAR datasets. Specifically, PLSR regression searches for a set of components (called latent vectors) that perform a simultaneous decomposition of X and Y with the constraint that these components explain as much as possible of the covariance between X and Y. This step generalizes PCA. It is followed by a regression step where the decomposition of X is used to predict Y (Abdi 2003). An additional feature of PLSR is the calculation of the Variable of Importance in the Projection (VIP) to evaluate importance of individual

predictors for estimation (Peerbhay et al. 2013). Technical details on PLSR are found in Wold et al. (2001).

Random Forests (RF) (Breiman 2001) is an ensemble learning method for regression and classification, which creates multiple decision trees and provides in output a regression model that is the mode of the regression output by individual trees. The main principle behind ensemble methods is that a group of weak learners can come together to form a strong learner. Each classifier or regressor, individually, is a weak learner, while all taken together are a strong learner. RF starts with a standard machine learning technique, called a decision tree which, in ensemble terms, corresponds to a weak learner. In a decision tree, an input is entered at the top and as it traverses down the tree the data gets bucketed into smaller and smaller sets. RF combines trees and, in ensemble terms, the trees are weak learners and the random forest is a strong learner. Thus, RF method combines bagging (Breiman 1996), which is a bootstrap aggregating method performing a resampling that creates n dataset composed by observation randomly selected from the original dataset, and the random selection of features in order to build a collection of decision trees with controlled variation. Bagging improves the stability and accuracy of machine learning algorithms, reducing variance and avoiding overfitting. Out-of-bag samples, so those not included in the previous bagging procedure used to build the trees, are used to calculate the error rate, eliminating the need for a test set or cross-validation; because a large number of trees are grown, generalization error is limited. Furthermore, RF can estimates the importance of a regression variable by looking at how much prediction error increases when out-of-bag data for that variable is permuted while all others are left unchanged (Liaw and Wiener 2002). RF runs efficiently on large data bases and can handle thousands of input variables without variable deletion or pruning needs. Parameter tuning in RF is relatively easy but in regression, values out of the range of those occurring in the training sample cannot be predicted. In the case study presented in Chapter 7, RF received in input hundreds of collinear features derived from the contiguous hyperspectral bands. Furthermore, as the available ground truth was obtained from a limited number of field sites, the possibility to avoid separating a cross-validation subset from limited field data constitutes an advantage. Additional technical details are

found in the official RF website
(http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm), as well as in
Cutler et al. (2007), Evans and Cushman (2009) and Prasad et al. (2006).

With respect to the classification problem, the main goal of image classification in a
forested landscape is often a thematic map describing forest characteristics (Wulder
1998). Image classifications label pixels into classes, or categories, based on distinctive
patterns of digital numbers or reflectance values, attempting to find patterns in the
spectral or structural response in relation to land cover groups known to be present. The
classification procedures are normally grouped either as supervised or unsupervised. For
more accurate information, a supervised classification is recommended as it involves
field observation data and a priori knowledge of land characteristics of the study area
(Lillesand and Kiefer 1994).

Among the supervised methods for image classification, the most popular one is the
maximum likelihood (ML) approach. This approach is a statistical decision rule method
that examines the probability of a pixel to belong to each class, with assignment of the
pixel to the class with the highest probability. It has the underlying assumption of a
normal distribution of the data within each class and may be biased with unequally
sized training classes (Jensen 1996). This is a limiting assumption that cannot be always
be met in a real world dataset and which justifies the use of more flexible and efficient
approaches.

Neural Network (NN) is a machine learning classification approach, which needs no
assumption on data distribution and is based on interconnected networks of simple
processing elements (Rumelhart et al. 1986). NN has been widely used in remote
sensing due mainly to the ability to perform accurate classifications, particularly when
the feature space is complex and the source data have different statistical distributions
(Atkinson and Tatnall 1997; Del Frate and Solimini 2004). Feed forward back-
propagation neural network is probably among the most popular neural network
approaches and has proven to enhance the classification of forest regenerating stages
(Liu et al. 2005). In this research, Neural Networks has been used to perform the two

complex classification tasks presented in Chapter 3 and 4, and ML has been used as a benchmark to evaluate the performance of NN machine learning approach.

1.2 Materials and methods

This thesis is divided into two main topics:

(*1*) land cover classification and vegetation type discrimination in complex forest landscapes: case studies are in a tropical and an alpine sites, using SAR and LiDAR data, and SAR and optical data respectively, with traditional maximum likelihood approach used as a benchmark against Neural Network machine learning

(*2*) estimation of biophysical and ecological forest parameters: case studies are in a temperate and in a tropical site, using LiDAR, and LiDAR and hyperspectral respectively, to retrieve AGB and the Shannon-Wiener biodiversity index. Ancillary vegetation type information was also use for the temperate site.

Detailed description of the remote sensing data corrections, images preparation, filtering, ground truth data, and of the classification or statistical modeling techniques are presented in chapters 3 to 7, which are related to specific case studies.

Here a summary table of the remote sensing data directly used for this thesis is presented, with the case studies in which were employed, and the algorithms used for tests.

Table 1: Materials of the study

Case study	Study area	Algorithm	Data type	Spatial Resolution
Vegetation type discrimination (classification)	Alps, Bozen	Maximum Likelihood and Neural Networks	ALOS PALSAR	20m
			RADARSAT-2	20m
			LiDAR-derived Canopy Height Model	20m
			SPOT 5	20m
Land cover (classification)	Gola, Sierra Leone	Maximum Likelihood and Neural Networks	Landsat TM	30m
			ALOS AVNIR	10m
			ALOS PALSAR	10m
AGB modeling (regression)	Sierra Nevada	Mixed Effects Models	Discrete return LiDAR	Points per meter: 2-4
AGB modeling (regression)	Gola, Sierra Leone	Partial Least Squares	Discrete return LiDAR	Points per meter: 10

		Regression	Hyperspectral AISA Eagle	1m
Shannon-Wiener index modeling (regression)	Gola, Sierra Leone	Random Forests Regression	Hyperspectral AISA Eagle	1m

1.2.1 The Sierra Nevada, U.S.A.

The Sierra Nevada study site is located in the United States Forest Service Sagehen Creek Experimental Forest in California, on the eastern slope of the Sierra Nevada. The area covers approximately 3925 ha, in an elevation range of 1862 m to 2670 m with average slope of 18%. The Sagehen Experimental Forest has a Mediterranean type climate with cold, wet winters and warm, dry summers. Annual precipitation is about 847 mm; snowfall accounts for more than 80 percent of the annual precipitation. Five major vegetation cover types can be found in the experimental forest: grass, shrub, mixed conifer, true fir, and conifer plantation. Mixed conifer stand is a mixture of several co-dominant species including ponderosa pine (*Pinus ponderosa*), Jeffrey pine (*Pinus Jeffreyi*), sugar pine *(Pinus lambertiana)*, white fir *(Abies concolor)*, red fir *(Abies magnifica)*, and incense cedar *(Calocedrus decurrens)*. Mixed conifer stands are found in higher elevations, mainly on the slopes south of Sagehen Creek. The true fir forest cover type occurs on northeast- and northwest-facing, high-elevation slopes south of Sagehen Creek in moist soil areas. Red fir is the dominant tree species, growing on deep, moist soils. White fir is the major associated species in lower elevations, while mountain hemlock *(Tsuga mertensiana)* and mountain mahogany *(Cercocarpus betuloides)* are associates at higher elevations. Other associated species are western white pine *(Pinus monticola)*, lodgepole pine (*Pinus contorta*), Jeffrey pine, and western juniper (*Juniperus occidentalis*). Non-forested areas include fens, wet and dry montane meadows and shrub fields.

1.2.2 The Alps, Bozen, Italy

The selected site is located in the Autonomous Province of Bolzano (Bozen), South Tyrol, in Northern Italy, in an area of approximately 30×30 km with elevations from 224 to 3343 m. Four main vegetation types were identified according to structural

criteria and ancillary data: needle-leaved forest, broadleaved forest, shrubs and dwarf pines, and grasslands. Broadleaved forest is typical of sub-mountain and basal planes, with mesophilous species (i.e. *Fagus sylvatica*) and thermophilous or thermomesophilous examples (i.e. *Quercus, Carpinus, Castanea* genera). Needle-leaved forest is found in the upper part of the mountain plane, with trees from different genera such as *Picea, Larix*, and *Pinus,* and in the cooler areas of the submountain plane mixed with broadleaved ones. Grasslands are frequently composed of tens of genera and are found in both the alpine plane (intermixed with short woody vegetation) and the lower planes, with flat terrain and deeper soil, where they are grazed by livestock. Few very resistant species can also be found in the the nival plane, which is dominated by lichens and mosses. Shrubs and dwarf pines are characteristic of an alpine plane – due to high winds, the vegetation has horizontal growth and its height is reduced, usually below 1 m. *Pinus mugo* formations are prevalent in the study site.

1.2.3 The Gola Rainforest National Park, Sierra Leone

This area is located along the border of Sierra Leone and Liberia and has been the subject of three of the presented case studies. In the first, land cover classification, the study area is larger, covering not only the Gola Rainforest National Park (GRNP) but also most of the Liberian Gola National Forest. In the second and third case studies (estimation of AGB and of Shannon-Wiener biodiversity index) the research has interested only the GRNP. Forests of the whole region are classified mainly as lowland moist evergreen with some drier parts occurrence, and are dominated by *Fabaceae*, *Euphorbiaceae* and *Sterculiaceae* families (Cole 1993). Outside the protected areas, in the north-western range, there is a fragmented landscape comprising small patches of disturbed forest, farmbushes, plantations, active agriculture areas, bare lands and settlements; in the southeastern range outside of GRNP and in Liberia the landscape is sparsely populated and dominated by forests. Overall, the area is characterized by a moist tropical climate with annual rainfall around 2500–3000 mm, a wet season lasting from May to October, and an altitude in the range 70–410 m without abrupt elevation changes. The dry period occurs between December and March, and corresponds to the semi-deciduous phenological stage of vegetation in the moist forest.

1.3 Thesis outline

A general overview of remote sensing in forested landscapes is presented in Chapter 2, with reference to land cover mapping and retrieval of parameters in the framework of forest research. In the same chapter, a review of the most recent challenges and problems in these two forest remote sensing main topics, classification and retrieval, follows.

The discrimination of vegetation types in an Alpine area is the case study presented in Chapter 3. Different SAR data (ALOS PALSAR and RADARSAT-2) have been used to perform a complex discrimination task taking advantage of data fusion with a Canopy Height Model derived by a LiDAR survey. The study compares Maximum Likelihood and Neural Networks algorithms abilities to perform the vegetation discrimination task, highlighting the advantages of machine learning methods and SAR data.

Chapter 4 presents the classification of a complex tropical forested landscape using both optical (Landsat TM and ALOS AVNIR) and SAR data (ALOS PALSAR), illustrating the benefits of data fusion, of Neural Networks for classification, of SAR data in tropical context, of textural features, and of different spatial resolution to capture useful forest information. A land cover map of this previously unstudied area is provided.

Chapter 5 focuses on the retrieval of above ground biomass in Sierra Nevada forest using airborne LiDAR and ancillary information on vegetation types, derived from aerial photography. The advantages of using Mixed Effects Models, which allow integration of vegetation types, against conventional multilinear regression is illustrated. The reported tests allow speculating on the importance of species-specific allometric equations and of highly precise geolocation of field data.

In Chapter 6 the AGB of a tropical forest is estimated by means of an innovative fusion of LiDAR and hyperspectral airborne data. This is one of the few studies dealing with these joined datasets and one of the few available estimations of forest biomass in Africa. It takes advantage of Partial Least Squared Regression to deal with the complex and multiple inputs generated by LiDAR metrics and hyperspectral contiguous reflectance bands.

The arboreal biodiversity of a tropical forest, a proxy for overall ecosystem biodiversity, is retrieved using hyperspectral data in a case study presented in Chapter 7. The study

illustrates that standard deviation of hyperspectral bands is most important in estimating the Shannon-Wiener index collected in the field, a measure of trees diversity and abundance. Random forests has been used to deal with the high number of collinear inputs and provided a mean to understand the most informative spectral regions. The use of derivative of reflectance and vegetation indices has been also tested.

Chapter 8 summarizes the thesis contributions, the achievements obtained in the applied case studies, the research questions answered, and draws conclusions and recommendations.

1.4 References

Abdi H., 2003. Partial Least Squares (PLS) Regression. In: Lewis-Beck M., Bryman, A., Futing T. (Eds.) Encyclopedia of Social Sciences Research Methods. Thousand Oaks (CA): Sage

Asner G.P. and Martin R.E., 2008. Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels. Remote Sensing of Environment, 112(10), 3958-3970.

Atkinson, P.M. and Tatnall, A.R.L., 1997. Introduction Neural networks in remote sensing. International Journal of Remote Sensing, 18(4): 699-709.

Breiman L. 1996. Bagging predictors. Machine Learning 24 (2): 123‑140.

Breiman L. 2001. Random Forests. Machine Learning 45 (1): 5‑32.

Brown H., and Prescott R., 2006. Applied mixed models in medicine. Wiley Ltd.

Chen Q. (2010). Retrieving canopy height of forests and woodlands over mountainous areas in the Pacific coast region using satellite laser altimetry. Remote Sensing of Environment, 114, 1610–1627.

Cole N.H.A., 1993. Floristic Associations in the Gola Rain forests: A Proposed biosphere reserve. Journal of Pure and Applied Science, 2:35-50

Criminisi, A., Shotton, J., Konukoglu E., 2011. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. Foundations and Trends in Computer Vision 7: 81–227.

Cutler D. R., Edwards T. C. Beard K. H., Cutler A., and Hess K. T., 2007. Random forests for classification in ecology. Ecology 88:2783-2792.

Del Frate F. and Solimini D., 2004. On neural network algorithms for retrieving forest biomass from SAR data. Geoscience and Remote Sensing, IEEE Transactions on, 42(1), 24-34.

Evans J. S. and S.A. Cushman, 2009. Gradient modeling of conifer species using random forests. Landscape Ecology 24(5): 673-683.

FAO, 2010. Global forest resources assessment. Terms and definitions. http://www.fao.org/docrep/014/am665e/am665e00.pdf (Accessed on: August 26, 2013).

Faraway J.J., 2004. Linear models with R. CRC Press.

Fisher, RA (1918). "The correlation between relatives on the supposition of Mendelian inheritance". Transactions of the Royal Society of Edinburgh 52 (2): 399–433.

Goodenough D.G., Li J.Y., Dyk A., 2006. Combining hyperspectral remote sensing and physical modeling for applications in land ecosystems. IEEE International Geoscience And Remote Sensing Symposium (IGARSS 2006), Denver, Colorado, Vol. 1-8.

IPCC/GRID-Arendal, 2001. Special Report on Land Use, Land-Use Change And Forestry. http://www.grida.no/publications/other/ipcc_sr/ (Accessed on: August 26, 2013).

Jensen J.R., 1996. Introductory Digital Image Processing : A remote Sensing Perspective. Prentice Hall.

Lei C., Ju C., Cai T., Jing X., Wei X., Di X., 2012. Estimating canopy closure density and above-ground tree biomass using partial least square methods in Chinese boreal forests.Journal of Forestry Research, Volume 23, Issue 2, pp 191-196.

Lillesand T.M. and Kiefe, R.W., 1994. Remote Sensing and Image Interpretation. John Wiley & Sons, Inc.

Liu J., Shao G., Zhu H. and Liu S., 2005. A neural network approach for enhancing information extraction from multispectral image data. Canadian Journal of Remote Sensing, 31(6): 432-438.

McCulloch, W. and Pitts W., 1943. A Logical Calculus of Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics 5 (4): 115–133.

NOAA Earth System Research Laboratory, 2007. Monthly mean atmospheric carbon dioxide at Mauna Loa Observatory, Hawaii. http://www.esrl.noaa.gov/gmd/ccgg/ - trends/co2_data_mlo.html (Accessed on: August 26, 2013)

Pal M., 2005. Random forest classifier for remote sensing classification. International Journal of Remote Sensing, 26(1), 217-222.

Parresol, R., 1999. Assessing Tree and Stand Biomass: A Review with Examples and Critical Comparisons. Forest Science, 45: 573-593.

Peerbhay K. Y., Mutanga O., Ismail R., 2013. Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu–Natal, South Africa. ISPRS Journal of Photogrammetry and Remote Sensing, Volume 79, Pages 19–28.

Pinheiro, J. C., and Bates, D. M., 2000. Linear mixed-effects models: basic concepts and examples (pp. 3-56). Springer New York.

Prasad A.M., Iverson L.R., and Liaw A., 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. Ecosystems 9:181-199.

Rumelhart D.E., Hinton G.E. and Williams R.J., 1986. Learning internal Representations by Error Propagation. In: D.E. Rumelhart, J.L. McClelland and P.R. Group (Editors), Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press, Cambridge, pp. 318-362.

Salas C., Ene L., Gregoire T. G., Næsset E., Gobakken T., 2010. Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models. Remote Sensing of Environment, 114, 1277–1285.

Wold S., Sjöström M., Eriksson L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58 (2): 109–130.

Wulder M.A., 1998. Optical remote sensing techniques for the assessment of forest inventory and biophysical parameters. Progress in Physical Geography, 22(4): 449-476.

Chapter 2

Remote sensing of forested landscapes

Recent researches in earth observation of forested landscapes are so advanced that is now possible to set up operational services to map land cover, forests, and their associated characteristics and changes. Examples are the Corine Land Cover program of the European Union, the Forest Resource Assessment of the United Nations Food and Agriculture Organization, the European Forest Fire Information System and the Forest layers of the Copernicus Land Monitoring Service realized by the Joint Research Center of the EU. Most of these operational mapping services are based on optical remote sensing data from sensors acquiring at different spatial resolutions. Forest landscape monitoring exercises have been carried out with any sensor, and with the broad range of currently available satellites and tools, so that the successful approaches are multiple. Due to the vastness of the topic, in the first paragraph of this chapter only a short review of the main approaches and sensors for land cover mapping is presented.

Also developed, but less consolidated, is the ability to retrieve forest parameters and specifically those of interest in this study: the main approaches of remote sensing estimation of AGB and biodiversity are briefly reviewed in two dedicated paragraphs. Biomass estimation has received increased attention in the last decade due to its relevancy to climate change and mitigation program such as REDD+. Differently from land cover mapping, in this case the most useful sensors are the active ones, such as radar and LiDAR, able to provide structural information on vegetation which is related to its biomass content. However, a clear methodology for AGB estimation has not been developed yet, and research efforts are still undergoing to try to overcome existing limitations and test the effectiveness of different tools in specific forested environments. Biodiversity estimation from remote sensing has been attempted for years with contrasting results. However, considering the different levels at which biodiversity can be estimated (ecosystem, species and genetic levels), and its indirect link with the electromagnetic signal, its accurate estimation is even more challenging and it represents an area of active and innovative research.

The remaining paragraphs in this chapter focus on the analysis of the recent challenges and problems in classification of vegetation and estimation of forest parameters by means of remote sensing. In fact, with the usefulness of remote sensing for natural remote sensing monitoring already proven, most of the recent scientific efforts are devoted to the identification of effective techniques to improve the results, or to overcome specific problems in forestry study. These challenges correspond to the specific research questions that the thesis addresses, and are related to both classification and retrieval tasks.

Specifically, a first set of questions is related to the usefulness of ancillary data in improving the results of classifications and AGB estimations, considering the availability of accessory datasets over many areas and from local administrations (paragraphs 2.3.1 and 2.3.2). Then, the effectiveness of integration of different data type, again for both classification and retrieval, which is favored by the increased availability of many sensors with different characteristics, such as optical, SAR, hyperspectral and LiDAR, is treated in paragraphs 2.3.3 and 2.3.4. Finally, the recurrent problem of geolocation in forest studies is considered in paragraph 2.3.5, evaluating how relevant is to AGB estimation the availability of accurate field data positional information.

2.1 Land cover mapping

The most popular sensors for land cover classification are optical ones, e.g. Landsat data, possibly for their historical availability, the high temporal resolution, the existing processing facilities in terms of software and algorithms, and the range of spatial resolutions allowing for different monitoring purposes, such as 15m (ASTER), 30m (Landsat), 250m and 500m (MODIS), and 1km (AVHRR). After image preprocessing, the classification task can be performed, with its complexity greatly depending on the selected classification algorithm, the number of classes to be identified, their similarity, and the heterogeneity and fragmentation of the territory.

In case of forested landscapes, Vegetation Indices (VIs) are used as an additional input to increase the discrimination of different forest classes or forest disturbances, based on phenological changes among different tree species (Wolter et al. 1995). The indices are

also useful to assess burn severity due to forest fires (Epting et al. 2005). Despite some advantages, the sensitivity of vegetation indices is rather limited to some wavelength windows, such as red or near infrared bands, and the saturation problem also reduces their applicability (Huete et al. 1997).

In addition to spectral information, also the spatial information provided by texture features is useful in image classification. Texture information involves information from neighboring pixels which is important to characterize the identified objects or regions of interest in an image (Haralick et al. 1973). In the present research, texture has been successfully used for the discrimination of vegetation types characterized by spectral similarities, and for the classification of a complex forested area. In Chapter 3 and 4 literature relevant to texture use is also presented and discussed.

The use of SAR data in land cover classification represents a great advantage when persistent cloud cover prevents optical data usage, like in tropical regions (Salas et al. 2002). Classification of vegetation types in tropical forest landscape has especially benefitted from lower frequency data: first studies took advantage of L-band JERS-1 data (Miranda et al. 1998) and then many researches and applications were developed based on ALOS PALSAR L-band data (Rosenqvist et al. 2008), including the two cases here presented.

A greater challenge in remote sensing is however the combination of optical and SAR sensors for improving the classification accuracy (Lu 2006), as demonstrated by Kuplich et al. (2005) combining JERS-1, SIR-C and X SAR and optical bands from Landsat TM for discriminating regenerating forest stages in the tropical Amazon. Attempts to combine optical and SAR data are also experimented in the present study (Chapter 4) to classify a tropical forested landscape.

Recently, LiDAR and hyperspectral sensors have been also used in vegetation classification efforts, often leading to the identification of single species (Korpela et al. 2010; Zhang et al. 2011). LiDAR provides highly detailed 3D structural information which can help in distinguishing among vegetation different physiognomy. LiDAR data are presently provided only by airborne surveys: the high acquisition cost allows the use of this data only over limited areas. LiDAR-derived data, such as Canopy Height Models, can be a cost effective way for improving classification and are often available

as by-products by local administrations; such an example is provided in the case study in Chapter 3 for an Alpine forested region.

Hyperspectral sensor, collecting hundreds of contiguous narrow bands in the VIS to SWIR regions of the spectra, are able to provide information on the biochemical components of leaves, and have been used alone and in conjunction with LiDAR data to characterize vegetation species and biodiversity (Asner and Martin 2008; Baldek and Asner 2013; Carlson et al. 2007; Leutner et al. 2012). The availability of both hyperspectral airborne and satellite data, and the forthcoming new space missions (EnMAP, PRISMA) make this data very promising for detailed ecological mapping.

2.2 Estimation of forest parameters

Among the different biophysical and ecological parameters which characterize forested landscapes, two are especially relevant with respect to main international environmental agreement: above ground biomass and biodiversity.

Above Ground Biomass (AGB) or biomass, terms here used interchangeably, is defined as all living biomass above the soil including stem, stump, branches, bark, foliage and seeds (IPCC 2003). Since over 50% of the forest dry biomass is carbon, AGB and its temporal changes is an important indicator for carbon sequestration estimation and climate change studies. Furthermore, its quantification is critical for the system of economic incentives to developing countries in preparation under the UN REDD+ framework. This, as well as national requirements for internal natural resources management, implies the need to develop precise procedures to quantify forest biomass. Those procedures are not yet established at international level; in spite of increasing studies and funding devoted to this topic in recent years, there is still a critical need of researching on AGB mapping methods and options. Biomass is usually expressed as mass per unit area, e.g. Mega gram (Mg)/ha (Brown 1997).

Biodiversity is defined in terms of genes, species and ecosystems, corresponding to three fundamental and hierarchically-related levels of biological organization. At least 40 per cent of the world's economy and 80 per cent of the needs of the poors are derived from biological resources. Forests are biologically diverse systems, representing some of the richest biological areas on Earth, as they offer a variety of habitats for plants,

animals and micro-organisms. However, forest biodiversity is increasingly threatened as a result of deforestation, fragmentation, climate change and other stressors, and it is a main target of the UN Convention of Biological Diversity. In forests, arboreal diversity can be considered a proxy measure of overall biodiversity (Gentry 1988). A common index to measure biodiversity is the

$$\text{Shannon-Wiener Index:} \qquad H' = -\sum_{i=1}^{N} pi \ln pi$$

where pi is the percentage of individuals belonging to the species ith in a field plot including N species.

2.2.1 Biomass estimation

The main approaches to biomass estimation in forested areas are direct field measurement and remote sensing. Field measurement is considered to be accurate but proves to be very costly and time consuming (De Gier 2003); often destructive sampling for developing specific allometric relationships between biomass and field measured parameters is required. Remote sensing approach has become an efficient technique, especially taking into account the recent availability of data with increased temporal, spectral and spatial resolutions. Modeling of forest biophysical parameters, such as biomass, from remote sensing data means that the forest properties are retrieved using spectral and/or structural information from the data, and the predicted results are then validated using a set of field measurements.

GIS-based (or combine and assign) approaches (Lu 2006) are also common, and take advantage of pre-existing datasets to reduce costs, usually attaching AGB-range values derived from literature or national inventories to land cover classes. These methods are as good as the input data used to set them, and are often limited by the availability of fine scale land cover maps and accurate AGB records per each class.

Many remote sensing studies explored the potential of optical satellite, LIDAR data, or SAR data for estimating forest biomass.

Multispectral optical data have been widely used for estimating AGB, often joined with data-derived features such as vegetation indices or textures (Foody and Cox 1994;

Houghton et al., 2001; Lu et al. 2004). Anyhow, the scattered energy perceived by optical sensors originates mainly from the upper canopy layer, causing a saturation problem at high vegetation density and hampering the use of this data type (Lu 2005; Steininger 2000). In tropical regions the dense cloud cover represents an additional issue against optical data use for AGB estimation.

On the other side, microwave remote sensing data are insensitive to the cloud-free and daylight conditions needed for optical image acquisition, and SAR data have been also widely employed for estimation of forest biophysical properties (Austin et al. 2003; Englhart et al. 2011; Fransson and Israelsson 1999; Kuplich et al. 2005; Rauste 2005). Empirical models of microwave data showed sensitivity to the density, shape, length, dielectric properties, and orientation of the scatterers (Kingsley and Quegan 1999). The X-band (2.4 – 3.75 cm) and C-band (3.75 – 7.5 cm) SAR data are useful to generate digital terrain models if interferometric pairs are available, and for discriminating the top canopy of vegetated land. In facts, due to their limited penetration ability as optical data, limitations for biomass retrieval at these frequencies exist: the backscattered response originates mainly from the upper canopy region and saturation at low vegetation densities occurs, for instance making difficult to distinguish forest regeneration after disturbance (Saatchi et al. 1997). In contrast, L-band SAR data (15 – 30cm) showed good ability for modeling the forest parameters under dense vegetation (Luckman et al. 1997; Rauste 2005) and using dual polarized SAR data discrimination between different forest succession stages was demonstrated feasible (Ranson et al. 1997; Rignot et al. 1997). Still, saturation limits exist at this frequency around 100 Mg/ha (far beyond the common 300 Mg/ha tropical forest values), and are dependent on the geometry of data acquisition, the polarization compositions, and the complexity of vegetation structure (Imhoff 1995; Ferrazzoli and Guerriero 1995). The use of polarimetry SAR interferometry (PolInSAR) data is an alternative but technically demanding approach to estimate AGB more accurately (Hajnsek et al. 2009). The BIOMASS P-band polarimetric and inteferometric SAR forthcoming ESA Earth Explorer is thus expected to provide unprecedented data for AGB estimation (Scipal et al. 2010).

Laser altimetry or Light Detection And Ranging (LiDAR) is an active remote sensing technology that determines ranges (distances) by taking the product of the speed of light

and the time required for an emitted laser to travel to a target object. The intense pulses density allows for laser returns both from the canopy and the ground, thus collecting data on the vertical distribution of vegetation in forests (Lim et al. 2003). Biomass is highly correlated to vegetation height metrics derived from LiDAR even at very high AGB densities (Lefsky et al. 2002). Much research has been done to estimate AGB using LiDAR data from ground, airborne and satellite platforms (e.g. Asner et al. 2009; Dubayah et al. 2010; Lefsky et al. 1999; Nelson et al. 2009; Ni-Meister et al. 2010); in all cases the accuracy of LiDAR-based AGB estimations confirmed this to be the most precise tool for detecting biomass in different ecosystems. However, the failure of the LiDAR instrument (GLAS) on board the IceSAT NASA satellite in 2012 and the high costs associated with airborne acquisitions limit the use of this instrument for monitoring purposes.

Hyperspectral sensors, recording the reflectance of a large number of fine resolution spectral bands in the VIS-NIR or VIS-SWIR range, are another frontier technology in remote sensing. Hyperspectral data can capture information on the biochemical composition of the upper canopy layer and has been used for forest type or species classification, and estimation of biophysical and biochemical properties and health status (Asner and Martin 2008; Koch 2010; Goodenough et al. 2006). Moreover, the ecosystem information recorded by hyperspectral data may relate to plant functional types –such as whether a species is light demanding -which could in turn affect wood density and thus biomass content (Baker et al. 2004; Chave et al. 2009).

Thus, integrating lidar and hyperspectral data for biomass estimation is a promising research area (Koch, 2010) which has seen only limited research, especially considering the forthcoming opportunities from future hyperspectral missions (EnMap, PRISMA, MERIS).

For biomass mapping purposes, the combination of different sensors is considered the most promising and innovative approach (Lu 2006; Lu et al. 2012; Tian et al. 2012). In the present research biomass estimation is attempted with an innovative combination of LiDAR and hyperspectral data and LiDAR and aerial photography-derived data (Chapter 6 and 3).

2.2.2 Biodiversity estimation

There are two general approaches to the remote sensing of biodiversity, as defined by Turner et al. (2003). One is the direct remote sensing of individual organisms, species assemblages, or ecological communities from airborne or satellite sensors with high spatial and spectral resolutions. For instance, hyperspectral sensors slice the electromagnetic spectrum into many discrete spectral bands, enabling the detection of spectral signatures that are characteristic of certain plant species or communities. The other approach is the indirect remote sensing of biodiversity through reliance on environmental parameters as proxies. For example, many species are restricted to discrete habitats that can be clearly identified remotely. By combining information about the known habitat requirements of species with maps of land cover derived from satellite imagery, precise estimates of potential species ranges and patterns of species richness are possible.

Palmer et al. (2002) proposed the spectral variation hypothesis to explain why electromagnetic measures are related to biological diversity: the spectral variation of reflectance values is correlated with spatial variation in the environment by means of landscape structure and complexity. Habitat heterogeneity is further linked to niche complexity which is known to enhance species diversity (Simonson et al. 2012). Considering tree diversity in forests, the remote sensing measures are in relationship with the chemical and structural properties of the vascular species.

Different studies revealed that the variability in vegetation biodiversity can only be partially captured with medium spatial resolutions and multispectral data (Carlson et al. 2007; Foody and Cutler 2003; Gould 2000; Rocchini et al. 2004, 2007). On the other hand, indirect mapping methods cannot be used where land cover information is lacking or where the cover is homogeneous, such as in tropical forests (Nagendra 2001), while fine scale biodiversity maps are needed by land managers and scientists, because they can provide an understanding of species distributions on a scale commensurate with conservation, management and policy development activities (Carlson et al. 2007). The increased availability of very high spectral resolution sensors has provided the opportunity to conduct detailed ecological studies on terrestrial ecosystems characteristics (Kumar et al. 2001; Thenkabail et al. 2004), but few previous researches

related hyperspectral data to biodiversity (Carlson et al 2007; Kalacksa et al. 2007). The present research offers an example of modeling the Shannon-Wiener biodiversity index in a tropical forest using airborne hyperspectral data (Chapter 7), developed with the aim of contributing to the identification of advanced tools to detect and monitor the distribution of life diversity on Earth.

2.3 Recent challenges in forest studies

This section deals with specific problems currently faced in forestry research, which correspond to the research questions that this study attempts to answer.

2.3.1 Ancillary data usefulness in AGB LiDAR-based estimations

Vegetation height metrics derived from LiDAR have been found to be highly correlated to biomass even at very high AGB densities (Gonzalez et al. 2010; Means et al. 1999). However, this relationship could vary across different vegetation types, as observed in different studies, because the biomass at the individual tree level is determined not only by canopy structure but also by floristic-related factors such as trunk taper and wood density (Chave et al. 2006; Niklas 1995). Drake et al. (2003) found that the relationships between LiDAR metrics and AGB differ between two tropical forest sites even after the models had adjusted for deciduousness of canopy trees, and attributed the differences to the underlying allometric relationships between stem diameter and AGB. Næsset and Gobakken (2008) estimated AGB in young and mature coniferous forests located in different areas of Norway, finding significant improvement of models when variables including tree species composition were included. Ni-Meister et al. (2010) found that the relationships between biomass and canopy structure are distinctly different for deciduous and conifer trees in temperate forests in New England, U.S.

Thus the inclusion of floristic information on forests can improve the models and should be attempted, as it can provide those accurate estimations needed in order to inform national policies and international treaties regarding forest management and carbon sequestration (Malmsheimer et al. 2011). This information is always available in non-

tropical countries, for which fine scale land cover databases are maintained (i.e. CORINE in EU), but is sometimes available also in developing countries thanks to advancements in research and international cooperation efforts.

One approach for incorporating ancillary forest information into biomass estimation is to stratify the forest plots according to vegetation types, each one having a separate statistical model (e.g., MacLean and Krabill 1986; Nelson et al. 1988). However, such an approach has practical and theoretical limitations. First, in most previous studies, only a limited number (typically 20–60 in total) of field plots were available for biomass modeling due to issues such as accessibility and cost. The stratification of a study area will lead to even fewer number of field plots per vegetation type, making it difficult to fit reliable statistical models for each vegetation type. Another problem of such an approach is that it assumes that the field data contains an exhaustive list of all vegetation types which exist in a given area. This is hardly true for natural forests because the vegetation types collected through field measurements are typically only a sample of the vegetation types which exist over that area.

Recent advances in mixed-effects modeling can circumvent the aforementioned problems. In a conventional statistical model, the regression coefficients (such as intercept and slopes) are treated as constants. However, in mixed-effects models, these coefficients could be modeled as random Gaussian variables with their specific values varying among vegetation types. This approach makes it feasible to estimate biomass even when the sample size per vegetation type is small.

Mixed effects modeling tool was applied in Chapter 5, in which the integration of ancillary floristic information into LiDAR based estimation was tested. The study area was located in a mixed and conifer forest in Sierra Nevada (US) where a LiDAR survey was conducted and where the availability of aerial photographs allowed the USDA Forest Service to develop a detailed vegetation type map. The study demonstrated the feasibility of this integration, which resulted in increased accuracy in estimates, and suggest a wider use of ancillary information into AGB studies.

2.3.2. Ancillary data (LiDAR-derived) usefulness in discriminating vegetation types

The usefulness of vegetation type identification is evident in forest studies, because single species or vegetation association need different management options, and because forest parameters can be better estimated at this level (as discussed in paragraph 2.3.1), thus providing more accurate estimates important for carbon accounting, timber industry, resource conservation and management. But the task of classification of vegetation types by remote sensing data is not always straightforward, especially when the spectral signatures of the objects to be identified are similar. The monitoring of natural areas is routinely conducted in most industrialized countries and, despite great advances in recent years in the availability of new remote sensing instruments, monitoring is mainly conducted with optical sensors. Thus, it is important to test advanced sensors, such as those exploiting very high-resolution synthetic aperture radar (SAR), hyperspectral, or LiDAR, for their integration in operational activities, and to evaluate their contribution to improve the quantity and quality of environmental information. Furthermore, some of these advanced remote sensing datasets can be easily obtained; often LiDAR data most commonly used for forestry and mapping applications are not produced by dedicated flights, with technicians exploiting the raster Canopy Height Model, which provides a measure of the height of the upper canopy for each pixel of vegetation in the surveyed area (Kraus and Pfeifer 1998), available at either a low cost or even for free from surveys carried out for purposes other than vegetation applications (Corona et al. 2012).

The increased availability of these LiDAR-derived products, released by local administrations due to widespread lidar use in topographic mapping, can be very valuable in the support of vegetation monitoring efforts. These data sets could be integrated in routine monitoring activities if their use is proven to be beneficial in increasing the ability to discriminate and map natural vegetation.

This issue is faced in Chapter 3, where the discrimination of natural vegetation based on SAR sensors in performed in an Alpine mountain area, threatened by climate change. Few other attempts exist on the integration of LiDAR into classification of vegetation;

the results are encouraging but those researches were located in other environments and based on optical sensors. For instance, Dowling and Accad (2003) joined the height information generated by LiDAR with digital video to map vegetation types and height classes in a riparian zone in Australia; Bork and Su (2007) compared the classifications obtained by LiDAR, multispectral, and the two combined data types in Canadian rangelands, finding more accurate the one including LiDAR data. Similarly, Geerling et al. (2007) found that the fused Compact Airborne Spectrographic Imager (CASI) spectral and LiDAR information produced better results than single dataset use in a natural floodplain classification in The Netherlands; while Onojeghuo and Blackburn (2011) combined hyperspectral imagery and textural information with LiDAR CHM for the effective mapping of reed bed habitats in the UK. The research presented in Chapter 3 is probably the first attempt based on SAR and LiDAR-derived data to discriminate among vegetation types, and shows how important is the height information provided by LiDAR in increasing the classification accuracy. The research thus offers an example on how improving operational monitoring of natural resources with limited effort, further discussing the additional advantages brought in by SAR, such as all-weather capabilities and retrieval of information on AGB.

2.3.3 Data fusion: evaluating the benefits of optical and RADAR sensors integration for tropical land cover classification

Optical sensors have been the primary data sources for land cover classification since the launch of the Landsat satellite series in early 1970s. In recent years, SAR sensors have emerged as important tools for vegetation studies, being well suited to detect volumetric scattering (Rahman and Sumantyo 2010; Santos et al. 2004; Simard et al. 2000) and offering a supplementary data source when atmospheric conditions hamper optical data use (Lehmann et al. 2012; Lu et al. 2007; Mitchard et al. 2011). Therefore, the combination of the two data types is considered beneficial (Lefsky and Cohen 2003). In fact optical and SAR provide information on different and complementary forest characteristics, such as canopy foliar composition and health, and vegetation

volume and canopy water content. But the benefits of data integration can vary according to landscape and specific sensors characteristics.

As seen in previous chapters, the distinction of different vegetation types is considered a difficult task. In tropical landscape the difficulties are generally represented by the smooth transition between forest successional stages, the presence of different forest types with spectral similarity, such as evergreen and the semi-deciduous forest classes, with latter in leaf-drop period only for few weeks (Lucas et al. 2002; Vieira et al. 2003). The extent of the benefits offered by optical and SAR data integration is explored in Chapter 5 in a forested tropical area located in West Africa, where no previous classification studies exist at a fine spatial resolution scale. Vegetation-oriented classification efforts are especially needed in the region, in which a trans-boundary peace park is planned by the Liberia and Sierra Leone governments.

The results showed that SAR combination with any optical data – TM or AVNIR-2 – always produced the best classification accuracies. Furthermore, the results obtained using only SAR data as classification input went beyond the usual SAR-based forest/non-forest mapping and the obtained accuracy was similar or higher than other classifications based on SAR in African environments (Haack and Bechdol 2000; Herold et al. 2004), enabling the detection of six different classes. It is frequent to experience optical data unavailability in tropical regions, such in this case study in which both optical images (Landsat and ASTER) were impacted by atmospheric conditions. The integrated optical and SAR data offered more information on forests details, with respect to classification based on single data types, but this information was available only where optical data were cloud free, i.e. in imagery sub-portions. Instead, SAR offered full coverage data, and SAR alone can still provide important landscape information. Thus, the use of SAR is suggested as a strategy to cope with persistent cloud cover affecting tropical regions, possibly merged with optical data to increase the information content, even considering different optical sensor for integration with SAR. If available, finer resolution sensors such as AVNIR-2 can help to detail specific subareas of interest, while larger areas or zones covered by clouds can be filled with SAR or optical lower resolution data.

2.3.4 Data fusion: evaluating the integration of LiDAR and hyperspectral sensors for AGB estimation

LiDAR is considered the most effective tool for biomass mapping. This instrument is not affected by the saturation problems characterizing SAR signal in dense forests and, differently from optical sensors, penetrates the forest down to the ground, collecting 3D information on vegetation which is related to AGB by means of allometric relationships. However, LiDAR application in tropical forests has been limited, particularly in Africa. Zolkos et al. (2013) in a comprehensive review identified eight studies carried out with this system in tropical forests, with none in continental Africa; considering the results obtained from tropical LiDAR application, a margin of improvement of accuracies of AGB retrieval seems to exist.

Hyperspectral sensors, recording the reflectance of a large number of fine resolution spectral bands in the VIS-NIR or VIS-SWIR range, are another frontier technology in remote sensing, capturing information on the biochemical composition of the upper canopy layer. The ecosystem information recorded by hyperspectral data may relate to plant functional types, or group of species, which develop in a site in response to environmental constraints, such as water and light availability. The arboreal species can differ substantially in their wood density, which in turns impact the biomass content (Baker et al. 2004; Chave et al. 2009).

Thus, the integration of these two data type may constitute an advantage, as each sensor brings in information which is complementary to the other one. Considering the opportunities from forthcoming hyperspectral missions, such as the Environmental Mapping and Analysis Program (EnMap), the PRecursore IperSpettrale of the application mission (PRISMA) and the Medium Resolution Imaging Spectrometer (MERIS), and the increased availability of LiDAR data, the joined use of the two data types can open new opportunities in ecological monitoring. Despite the interest of this approach, overall the number of published studies on integrating LiDAR and hyperspectral data for biomass estimation is very small, and the few researches that attempted it in boreal, temperate and tropical forests reported only modest or no

improvement in model fit compared to the results from using LiDAR only (Anderson et al. 2008; Clark et al. 2012; Latifi et al. 2012; Swatantran et al. 2011).

In the present thesis, Chapter 6 offers an example of increased accuracy obtained by the addition of hyperspectral features to LiDAR for AGB estimation in a tropical dense forest. Previous researches are scarcely comparable to the present one, either due to different sensors and spatial resolutions used (Anderson et al., 2008; Swatantran et al., 2011), or to the different forest biomes considered. Anyhow, Swatantran et al. (2011) suggested that the predictive power of hyperspectral data could be higher when LiDAR relationships with biomass are weaker, as observed by Anderson et al. (2008) and Roth (2009). This hypothesis is in part confirmed by the case study presented in Chapter 6, in which the LiDAR-AGB relationship in a complex tropical biome is not as high as in less complex temperate forests, and an increase in accuracy was brought by inclusion of hyperspectral data. Other attempts of LiDAR and hyperspectral data fusion at very high resolution are also difficult to compare with the present case study for the different techniques used in data acquisition, processing and analysis. Latifi et al. (2012) used a full waveform lidar and HyMap hyperspectral sensors, reporting minimal improvement in AGB estimates from fused datasets, using Principal Component Regression. The Partial Least Squares regression used in this study is preferable to PCR, as detailed in Chapter 6, which might account for the difference. The only AGB estimation for a tropical area at very high resolution, carried out using a FLI MAP LiDAR (Fugro Aerial & Mobile Mapping Inc.) and the hyperspectral HYDICE sensor reported no improvement by data fusion (Clark et al. 2012). In that case the authors used hyperspectral VIs and spectral mixture fractions, while in this study original hyperspectral bands were used. Possibly, important information for biomass estimation was excluded in deriving the VIs or spectral mixture fractions, with respect to the full information content included in the original bands.

Even considering the positive results obtained in Chapter 6, the status of current research in hyperspectral and LiDAR data fusion is not advanced enough to derive clear conclusions on the effectiveness of the use of joined sensors. In this view, the case study illustrates an example of innovative research in remote sensing of natural resources.

2.3.5 Evaluating the impact of field data geolocation in LiDAR-based AGB estimates

The accuracy of the calibration data used for deriving forest parameters estimates from remote sensing data is influenced not only by the specific earth observation data and the statistical approaches used, but also by the accuracy of the field data geolocations. Low geo-location accuracy can be caused by either the use of low-cost GPS or the existence of dense forests. The use of differential GPS (DGPS) correction can improve the geolocation recorded by low-costs GPS, from the 15 meter nominal GPS accuracy to about 10 cm in case of the best implementations. DGPS uses a network of fixed, ground-based reference stations to broadcast the difference between the positions indicated by the satellite systems and the known fixed positions. These stations broadcast the difference between the measured satellite pseudoranges (ranges including errors) and actual (internally computed) pseudoranges, and receiver stations may correct their pseudoranges by the same amount. The digital correction signal is typically broadcast locally over ground-based transmitters of shorter range. Unfortunately, these ground-based reference stations are not always operating in tropical areas. Furthermore, the bouncing of GPS signal into a dense canopy can result in a delay of the GPS time information which results in an imprecise geolocation recorded by the instrument. This phenomenon is called multipath effect; presently different brands offer multipath correction systems integrated into their GPS, but these effects cannot be completely eliminated.

The effects of geolocation accuracy on the accuracy of forest parameter estimations have been scarcely investigated. Dominy and Duncan (2001) reported the difficulty of quality satellite reception beneath a dense forest canopy, with the degree of spatial error seriously affecting fine-scale vegetation mapping. Miura and Jones (2010) used a Garmin eTrex GPS (average±5.5 m horizontal error) to locate the centers of 25-m radius circular plots for field measurements and related to airborne LiDAR data. They had to manually shift the plots to achieve a better registration between lidar data and field measurements.

In this research, the geolocation problem was investigated and discussed with respect to the estimation of AGB (Chapter 3 and Chapter 6) and of the Shannon-Wiener diversity index (Chapter 7). The availability of both differential GPS coordinates and recreational GPS coordinates for the field plots in Sierra Nevada study site (Chapter 3) made it possible to directly assess the impacts of plot coordinate accuracy on biomass estimation. The use of recreational instead of differential GPS in that study site resulted in a significant decrease of the accuracy of biomass estimates. This degradation in performance due to GPS accuracy will likely vary depending on the site-specific conditions (e.g., canopy structure, spatial heterogeneity, and topography), and emphasize the value of differential GPS to locate field plots for vegetation measurements. Instead, in the Sierra Leone tropical study site (Chapter 6 and 7) field data geolocations were recorded only using a recreational low-cost GPS, and this was considered among the causes of errors associated with the estimates, with both researches suggesting that high quality ground truth data is needed when planning forest parameters estimates in tropical forests.

2.4 References

Anderson J.E., Plourde L.C., Martin M.E., Braswell B.H., Smith M.L., Dubayah R.O., Hofton M.A., Blair J.B., 2008. Integrating waveform LiDAR with hyperspectral imagery for inventory of a northern temperate forest. Remote Sensing of Environment112: 1856–1870.

Asner G.P., Hughes R.F., Varga T.A., Knapp D.E., and Kennedy-Bowdoin T., 2009. Environmental and biotic controls over aboveground biomass throughout a tropical rainforest. Ecosystems, 12, 261–278.

Asner G.P. and Martin R.E., 2008. Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels. Remote Sensing of Environment, 112(10), 3958-3970.

Austin J., Mackey B. and Van Niel K., 2003. Estimating forest biomass using satellite radar: an exploratory study in a temperate Australian Eucalyptus forest. Forest Ecology & Management, 176(1-3): 575 - 583.

Baldeck C.A., and Asner G.P., 2013. Estimating Vegetation Beta Diversity from Airborne Imaging Spectroscopy and Unsupervised Clustering. Remote Sensing, 5(5), 2057-2071.

Baker T.R., Phillips O.L., Malhi Y., Almeida S., Arroyo L., Di Fiore A., Erwin T., Higuchi N., Killeen T.J., Laurance S.G., Laurance W.F., Lewis S.L., Lloyd J., Monteagudo A., Neill D.A., Patino S., Pitman N.C.A., Silva J.N.M., Vasquez Martinez R., 2004. Variation in wood density determines spatial patterns in Amazonian forest biomass. Global Change Biology 10:545–562.

Bork E.W. and Su J.G., 2007. Integrating LIDAR Data and Multispectral Imagery for Enhanced Classification of Rangeland Vegetation: A Meta Analysis. Remote Sensing of Environment 111:11–24.

Brown S., 1997. Estimating Biomass and Biomass Change of Tropical Forests: a Primer. FAO Forestry Paper 134. FAO Food and Agriculture Organization of the United Nations, Rome.

Carlson K.M., Asner G.P., Hughes R.F., Ostertag R., and Martin R.E., 2007. Hyperspectral remote sensing of canopy biodiversity in Hawaiian lowland rainforests. Ecosystems, 10(4), 536-549.

Chave J., Muller-Landau H.C., Baker T.R., Easdale T.A., Ter Steege H., and Webb C.O. , 2006. Regional and phylogenetic variation of wood density across 2456 neotropical tree species. Ecological Applications, 16, 2356–2367.

Chave J., Coomes D., Jansen S., Lewis S.L., Swenson N.G., Zanne A.E., 2009. Towards a worldwide wood economics spectrum. Ecology Letters, Volume 12, Issue 4, pages 351–366.

Clark M.L., Roberts D.A. Ewel J.J., Clark D.B., 2011. Estimation of tropical rain forest aboveground biomass with small-footprint LiDAR and hyperspectral sensors. Remote Sensing of Environment, 115, pp. 2931–2942.

Corona P., Cartisano R., Salvati R., Chirici G., Floris A., Di Martino P., Marchetti M., Scrinzi G., Clementel F., Travaglini D., and Torresan C., 2012. "Airborne Laser Scanning to Support Forest Management." European Journal of Remote Sensing 45: 27–37.

De Gier A., 2003. A new approach to woody biomass assessment in woodlands and shrublands. In: P. Roy (Editor), Geoinformatics for Tropical Ecosystems, India, pp. 161-198.

Dowling, R. and A. Accad. 2003. Vegetation Classification of the Riparian Zone along the Brisbane River, Queensland, Australia, Using Light Detection and Ranging (LiDAR) Data and Forward Looking Digital Video. Canadian Journal of Remote Sensing 29: 556–563.

Drake J.B., Knox R.G., Dubayah R.O., Clark D.B., Condit R., Blair J.B., and Hofton M., 2003. Above-ground biomass estimation in closed canopy Neotropical forests using

lidar remote sensing: Factors affecting the generality of relationships. Global Ecology & Biogeography, 12, 147–159.

Dominy N. J. and Duncan B., 2001. GPS and GIS methods in an African rain forest: applications to tropical ecology and conservation. Conservation Ecology 5(2): 6.

Dubayah R.O., Sheldon S.L., Clark D.B., Hofton M.A., Blair J.B., Hurtt G.C., and Chazdon R.L., 2010. Estimation of tropical forest height and biomass dynamics using LiDAR remote sensing at La Selva, Costa Rica. Journal of Geophysical Research, 115, G00E09.

Englhart S., Keuck V., and Siegert F., 2011. Aboveground biomass retrieval in tropical forests—The potential of combined X-and L-band SAR data use. Remote sensing of environment, 115(5), 1260-1271.

Epting J., Verbyla D.L. and Sorbel B., 2005. Evaluation of Remotely Sensed Indices for Assessing Burn Severity in Interior Alaska using Landsar TM and ETM+. Remote Sensing of Environment, 96(3-4): 328-339.

Ferrazzoli P. and Guerriero L., 1995. Radar sensitivity to tree geometry and woody volume: a model analysis. *Geoscience and Remote Sensing, IEEE Transactions on*, *33*(2), 360-371.

Fisher R.A., 1918. "The correlation between relatives on the supposition of Mendelian inheritance". Transactions of the Royal Society of Edinburgh 52 (2): 399–433.

Foody G.M. and Cox D.P., 1994. Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions. International Journal of Remote Sensing, 15(3): 619-631.

Fransson J.E.S. and Israelsson H., 1999. Estimation of stem volume in boreal forests using ERS-1 C- and JERS-1 L-band SAR data. International Journal of Remote Sensing, 20(1): 123-137.

Geerling G.W., Labrador-Garcia M., Clevers J., Ragas A., and Smits A.J.M., 2007. Classification of Floodplain Vegetation by Data-Fusion of Spectral (CASI) and LiDAR data. International Journal of Remote Sensing 28: 4263–4284.

Gentry A.H., 1992. Tropical forest biodiversity: distributional patterns and their conservational significance. Oikos, 19-28.

Gonzalez P., Asner G.P., Battles J.J., Lefsky M.A., Waring K.M., and Palace M., 2010. Forest carbon densities and uncertainties from LiDAR, QuickBird and field measurements in California. Remote Sensing of Environment, 114(7), 1561–1575.

Goodenough D.G., Li J.Y., Dyk A., 2006. Combining hyperspectral remote sensing and physical modeling for applications in land ecosystems. IEEE International Geoscience And Remote Sensing Symposium (IGARSS 2006), Denver, Colorado, Vol. 1-8.

Gould W., 2000. Remote Sensing of Vegetation, Plant Species Richness, and Regional Biodiversity Hotspots. Ecological Applications Vol. 10, No. 6, pp. 1861-1870.

Haack B. and Bechdol M., 2000. Integrating multisensor data and radar texture measures for land cover mapping. Computers and Geosciences 26 (4), 411–421.

Hajnsek I., Kugler F., Lee S.K. and Papathanassiou K.P., 2009. Tropical forest parameter estimation by means of PolInSAR: The INDREX-II campaign. IEEE Transactions on Geoscience and Remote Sensing, 47(2): 481-493.

Haralick R.M., Shanmugam K. and Dinstein I., 1973. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics, smc 3(6): 610-621.

Herold N.D., Haack B.N., Solomon E., 2004. An evaluation of radar texture for land use/cover extraction in varied landscapes. International Journal of Applied Earth Observation and Geoinformation 5 (2), 113–128.

Houghton R.A., Lawrence K.T., Hackler J.L. and Brown S., 2001. The spatial distribution of forest biomass in the Brazilian Amazon: A comparison of estimates. Global Change Biology, 7(7): 731-746.

Huete A.R., Liu H., Batchily K. and Leeuwen W., 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. Remote Sensing of Environment 59(3): 440-451.

Imhoff M.L., 1995. Radar backscatter and biomass saturation: ramifications for global biomass inventory. IEEE Transactions on Geoscience and Remote Sensing, 33(2): 511-518.

IPCC 2003. Good Practice Guidance for LULUCF – Glossary. http://www.ipcc-nggip.iges.or.jp/public/gpglulucf/gpglulucf_files/Glossary_Acronyms_BasicInfo/Glossary.pdf (Accessed on: August 29, 2013)

Kalacska M., Sanchez-Azofeifa G.A., Rivard B., Caelli T., Peter White H., Calvo-Alvarado J.C., 2007. Ecological fingerprinting of ecosystem succession: Estimating secondary tropical dry forest structure and diversity using imaging spectroscopy, Remote Sensing of Environment, Volume 108, Issue 1, Pages 82-96.

Kingsley S., and Quegan S. 1999. Understanding Radar Systems. SciTech Publishing.

Koch B., 2010. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. ISPRS Journal of

Photogrammetry and Remote Sensing Volume 65, Issue 6, November 2010, Pages 581–590. ISPRS Centenary Celebration Issue

Korpela I., Ørka H. O., Maltamo M., Tokola T., and Hyyppä, J., 2010. Tree species classification using airborne LiDAR–effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. Silva Fennica, 44(2), 319-339.

Kraus, K., and Pfeifer N.1998. "Determination of Terrain Models in Wooded Areas with Airborne Laser Scanner Data." ISPRS Journal of Photogrammetry and Remote Sensing 53: 193–203.

Kumar L., Schmidt K., Dury S. & Skidmore A., 2001. Imaging spectrometry and vegetation science. In Imaging Spectrometry: Basic Principles and Prospective Applications (eds.). F. D. Van der Meer & S. M. De Jong, pp.111-155. Dordrecht: Kluwer Academic Publishers.

Kuplich T.M., Curran P.J. and Atkinson P.M., 2005. Relating SAR image texture to the biomass of regenerating tropical forests. International Journal of Remote Sensing, 26(21): 4829-4854.

Latifi, H., Faßnacht F., Koch B., 2012. Forest structure modeling with combined airborne hyperspectral and LiDAR data. Remote Sensing of Environment, Vol. 121, pp. 10-25

Lefsky M.A., Harding D., Cohen W.B., Parker G., and Shugart H.H., 1999. Surface LiDAR remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. Remote Sensing of Environment, 67, 83–98.

Lefsky M.A., Cohen W.B., Harding D.J., Parker G.G., Acker S.A., and Gower S.T., 2002. Lidar remote sensing of above-ground biomass in three biomes. Global ecology and biogeography, 11(5), 393-399.

Lefsky M.A. and Cohen W.B., 2003. Selection of remotely sensed data. In: Wulder, M.A., Franklin, S.E. (Eds.), Remote Sensing of Forest Environments: Concepts and Case Studies. Kluwer Academic Publishers, Boston, pp. 13–46.

Lehmann E.A., Caccetta P.A., Zhou Z., McNeill S.J., Wu X., Mitchell A.L., 2012. Joint Processing of Landsat and ALOS-PALSAR Data for Forest Mapping and Monitoring. IEEE Transactions on Geoscience and Remote Sensing 50,55–67.

Leutner B. F., Reineking B., Müller J., Bachmann M., Beierkuhnlein C., Dech S., and Wegmann M., 2012. Modelling forest α-diversity and floristic composition—On the added value of LiDAR plus hyperspectral remote sensing. Remote Sensing, 4(9), 2818-2845.

Liaw A. and M. Wiener M., 2002. Classification and Regression by Random Forest. R News 2(3), 18-22.

Lim K., Treitz P., Wulder M., St-Onge B., and Flood M., 2003. LiDAR remote sensing of forest structure. Progress in physical geography, 27(1), 88-106.

Lu D., 2005. Aboveground biomass estimation using Landsat TM data in Brazilian Amazon. International Journal of Remote Sensing, 26(12): 2509-2525.

Lu D., 2006. The potential and challenge of remote sensing- based biomass estimation. International Journal of Remote Sensing, 27(7): 1297-1328.

Lu D., Mausel P., Brondizio E. and Moran E., 2004. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. Forest Ecology and Management, 198(1-3): 149-167.

Lu D., Batistella M., Moran E., 2007. Land-cover classification in the Brazilian Amazon with the integration of Landsat ETM+ and RADARSAT data. International Journal of Remote Sensing 28 (24), 5447–5459.

Lu D., Chen Q., Wang G., Moran E., Batistella M., Zhang M., Vaglio Laurin G. and Saah, D., 2012. Aboveground forest biomass estimation with Landsat and LiDAR data and uncertainty analysis of the estimates. International Journal of Forestry Research, 2012.

Lucas R.M., Honzák M., Amaral I.D., Curran P.J., Foody G.M., 2002. Forest regeneration on abandoned clearances in central Amazonia. International Journal of Remote Sensing 23 (5), 965–988.

Luckman, A., Baker, J., Kuplich, T.M., Corina da Costa, F.Y. and Alejandro, C.F., 1997. A study of the relationship between radar backscatter and regenerating tropical forest biomass for spaceborne SAR instruments. Remote Sensing of Environment, 60(1): 1- 13.

MacLean G.A. and Krabill W.B., 1986. Gross-merchantable timber volume estimation using an airborne LiDAR system. Canadian Journal of Remote Sensing, 12, 7–18.

Malmsheimer R.W., Bowyer J.L., Fried J.S., Gee E., Izlar R.L., Miner R.A., Munn I.A., Oneil E.,and Stewart W.C., 2011. Managing forests because carbonmatters: Integrating energy, products, and land management policy. Journal of Forestry, 109, S7–S48.

Means J.E., Acker S.A., Harding D.J., Blair J.B., Lefsky M.A., Cohen W.B., Harmon M.E., and McKee W.A., 1999. Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the Western Cascades of Oregon. Remote Sensing of Environment, 67, 298–308.

Miranda F.P., Fonseca L.E.N. and Carr J.R., 1998. Semivariogram textural classification of JERS-1 (Fuyo-1) SAR data obtained over a flooded area of the Amazon rainforest. International Journal of Remote Sensing, 19(3): 549-556.

Mitchard E.T.A., Saatchi S.S., White L.J.T., Abernethy K.A., Jeffery K.J., Lewis S.L., Collins M., Lefsky M.A., Leal M.E., Woodhouse I.H., Meir P., 2011. Mapping tropical forest biomass with radar and spaceborne LiDAR: overcomingproblems of high biomass and persistent cloud. Biogeosciences 8 (4), 8781–8815.

Miura N. and Jones S.D., 2010. Characterizing forest ecological structure using pulse types and heights of airborne laser scanning. Remote Sensing of Environment, 114, 1069–1076.

Næsset E., and Gobakken T., 2008. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. Remote Sensing of Environment, 112, 3079–3090.

Nagendra H., 2001. Using remote sensing to assess biodiversity. International Journal of Remote Sensing 22(12): 2377-2400

Nelson R., Krabill W., and Tonelli J., 1988. Estimating forest biomass and volume using airborne laser data. Remote Sensing of Environment, 24, 247–267.

Nelson R., Boudreau J., Gregoire T.G., Margolis H., Næsset E., Gobakken T., and Stahl G., 2009. Estimating Quebec provincial forest resources using ICESat/GLAS. Canadian Journal of Forest Research, 39, 862–881.

Niklas K.J., 1995. Size-dependent allometry of tree height, diameter and trunk-taper. Annals of Botany, 75, 217–227.

Ni-Meister W., Lee S.Y., Strahler A.H., Woodcock C.E., Schaaf C., Yao T.A., Ranson, K.J., Sun G. Q., and Blair J.B., 2010. Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from LiDAR remote sensing. Journal of Geophysical Research-Biogeosciences, 115

Onojeghuo A.O. and Blackburn G.A., 2011. Optimising the Use of Hyperspectral and LiDAR Data for Mapping Reedbed Habitats. Remote Sensing of Environment 115: 2025–2034.

Palmer M.W., Earls P.G., Hoagland B.W., White P.S., Wohlgemuth T., 2002. Quantitative tools for perfecting species lists. Environmetrics 13, 121‑137.

Rahman M.M. and Sumantyo J.T.S., 2010. Mapping tropical forest cover and deforestation using synthetic aperture radar (SAR) images. Applied Geomatics 2 (4),113–121.

Ranson K.J., Sun G., Lang R.H., Chauhan N.S., Cacciola R.J., and Kilic O., 1997. Mapping of boreal forest biomass from spaceborne synthetic aperture radar. Journal of Geophysical Research, 102(D24), 29599-29.

Rauste Y., 2005. Multi-temporal JERS SAR data in boreal forest biomass mapping. Remote Sensing of Environment, 97(2): 263-275.

Rignot E., Salas W.A. and Skole D.L., 1997. Mapping deforestation and secondary growth in Rondonia, Brazil, using imaging radar and thematic mapper data. Remote Sensing of Environment, 59(2): 167- 179.

Rocchini, D., Chiarucci, A., Loiselle, S.A., 2004. Testing the spectral variation hypothesis by using satellite multispectral images. Acta Oecologica 26, 117‑120.

Rocchini, D., Ricotta, C., Chiarucci, A., 2007. Using remote sensing to assess plant species richness: the role of multispectral systems. Applied Vegetation Science, 10: 325-332.

Rosenqvist, A., Shimada, M. and Milne, A.K., 2008. The ALOS Kyoto & carbon initiative, International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3614-3617.

Roth K.L.,2009. A combined lidar and hyperspectral remote sensing analysis for mapping forest biomass. Unpublished master thesis. University of California, Santa Barbara, Dept of Geography.

Saatchi S.S., Soares J.V. and Alves D.S., 1997. Mapping deforestation and land use in amazon rainforest by using SIR-C imagery. Remote Sensing of Environment, 59(2): 191-202.

Salas W.A., Ducey M.J., Rignot E. and Skole D., 2002. Assessment of JERS-1 SAR for monitoring secondary vegetation in Amazonia: I. Spatial and temporal variability in backscatter across a chronosequence of secondary vegetation stands in Rondonia. International Journal of Remote Sensing, 23(7): 1357-1379.

Santos J.R., Neeff T., Dutra L.V., Araujo L.S., Gama F.F., Elmiro M.A.T., 2004. Tropical forest biomass mapping from dual frequency SAR interferometry (X and P-Bands). Twentieth International Society for Photogrametry and Remote Sensing (ISPRS) Congress: GeoImagery Bridging Continents, Istanbul, v. XXXV, pp.1133–1136.

Scipal K., Arcioni M., Chave J., Dall J., Fois F., LeToan T., Lin C.-C., Papathanassiou K., Quegan S., Rocca F., Saatchi S., Shugart H., Ulander L., Williams M., 2010. The BIOMASS mission—An ESA Earth Explorer candidate to measure the BIOMASS of the earth's forests. In Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International (pp. 52-55). IEEE.

Simard M., Saatchi S., De Grandi G.F., 2000. The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest. IEEE Transactions on Geoscience and Remote Sensing 38 (5), 2310–2321.

Simonson W. D., Allen H. D., and Coomes D. A., 2012. Use of an airborne lidar system to model plant species composition and diversity of Mediterranean oak forests. Conservation Biology, 26(5), 840-850.

Steininger M., 2000. Satellite estimation of tropical secondary forest above ground biomass: data from Brazil and Bolivia. International Journal of Remote Sensing, 21(6 & 7): 1139-1157.

Swatantran A., Dubayah R., Roberts D., Hofton M., & Blair J. B., 2011. Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion. Remote Sensing of Environment, 115, 2917–2930.

Thenkabail P.S., Enclona E.A., Ashton M.S., Legg C. and De Dieu M.J., 2004a. Hyperion, IKONOS, ALI and ETM plus sensors in the study of African rainforests. Remote Sensing of Environment, 90, pp. 23-43.

Tian X., Su Z., Chen E., Li Z., van der Tol C., Guo J., and He Q., 2012. Estimation of forest above-ground biomass using multi-parameter remote sensing data over a cold and arid area. International Journal of Applied Earth Observation and Geoinformation, 14(1), 160-168.

Turner W., Spector S., Gardiner N., Fladeland M., Sterling E., and Steininger M., 2003. Remote sensing for biodiversity science and conservation. Trends in ecology & evolution, 18(6), 306-314.

Vieira I.C.G., Almeida A.S., Davidson E.A., Stone T.A., Carvalho C.J.R., Guerrero J.B., 2003. Classifying successional forests using Landsat spectral properties and ecological characteristics in eastern Amazônia. Remote Sensing of Environment 87 (4), 470–481.

Wolter P.T., Mladenoff D.J., Host G.E. and Crow T.R., 1995. Improved Forest Classification in the Northern Lake States using Multi-Temporal Landsat Imagery. Photogrammetric Engineering & Remote Sensing, 61(9): 1129-1143.

Zhang Z., Liu X., Peterson J., and Wright W., 2011. Cool temperate rainforest and adjacent forests classification using airborne LiDAR data. Area, 43(4), 438-448.

Zolkos S.G., Goetz S.J., Dubayah R., 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. Remote Sensing of Environment, 128, pp. 289-298.

Chapter 3

Integration of airborne lidar and vegetation types derived from aerial photography for mapping aboveground live biomass

Integration of airborne lidar and vegetation types derived from aerial photography for mapping aboveground live biomass

Qi Chen [a,*], Gaia Vaglio Laurin [b,c], John J. Battles [d], David Saah [e,f]

[a] Department of Geography, University of Hawai'i at Manoa, 422 Saunders Hall, 2424 Maile Way, Honolulu, HI 96822, USA
[b] Department of Computer, System and Production Engineering, University of Tor Vergata, Rome 00133, Italy
[c] CMCC - Centro Mediterraneo per i Cambiamenti Climatici, via Augusto Imperatore (Euro-Mediterranean Center for Climate Change), Lecce 73100, Italy
[d] Department of Environmental Science, Policy, and Management, 137 Mulford Hall, University of California at Berkeley, Berkeley, CA 94720, USA
[e] Spatial Informatics Group, LLC, 3248 Northampton Ct., Pleasanton, CA 94588, USA
[f] College of Arts and Sciences, Environmental Science, University of San Francisco, San Francisco, CA 94117, USA

ARTICLE INFO

ABSTRACT

The relationship between lidar-derived metrics and biomass could vary across different vegetation types. However, in many studies, there are usually a limited number of field plots associated with each vegetation type, making it difficult to fit reliable statistical models for each vegetation type. To address this problem, this study used mixed-effects modeling to integrate airborne lidar data and vegetation types derived from aerial photographs for biomass mapping over a forest site in the Sierra Nevada mountain range in California, USA. It was found that the incorporation of vegetation types via mixed-effects models can improve biomass estimation from sparse samples. Compared to the use of lidar data alone in multiplicative models, the mixed-effects models could increase the R^2 from 0.77 to 0.83 with RMSE (root mean square error) reduced by 10% (from 80.8 to 72.2 Mg/ha) when the lidar metrics derived from all returns were used. It was also found that the SAF (Society of American Forest) cover types are as powerful as the NVC (National Vegetation Classification) alliance-level vegetation types in the mixed-effects modeling of biomass, implying that the future mapping of vegetation classes could focus on dominant species. This research can be extended to investigate the synergistic use of high spatial resolution satellite imagery, digital image classification, and airborne lidar data for more automatic mapping of vegetation types, biomass, and carbon.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Vegetation biomass, the weight of plant materials that exist over an area, is a critical measure of ecosystem structure and productivity that informs a range of applications such as fire emission calculations (e.g., De Santis et al., 2010), wildlife habitat analysis (e.g., Morris et al., 2009), hydrological modeling (e.g., Ursino, 2007), and greenhouse gas accounting (e.g., De Jong et al., 2010). In particular, accurate estimates of biomass are needed in order to inform national policies and international treaties regarding forest management and carbon sequestration (Malmsheimer et al., 2011).

Lidar is a state-of-the-art remote sensing technology with a proven ability to map aboveground biomass (AGB). The accuracy and sensitivity of the metrics derived from optical and radar imagery (such as NDVI and backscatter coefficient) decline with increasing AGB (Waring et al., 1995). In contrast, vegetation height metrics derived from lidar have been found to be highly correlated to biomass even when the biomass density is very high (Gonzalez et al. 2010, Means et al., 1999). In the past, much research has been done to estimate AGB using airborne discrete-return lidar (e.g., Asner et al., 2009; Banskota et al., 2011; Lim et al., 2003), airborne profiling lidar (e.g., Nelson et al., 2009, 1988; Stahl et al., 2011), airborne waveform lidar (e.g., Dubayah et al. 2010; Lefsky et al., 1999; Ni-Meister et al., 2010), satellite lidar (e.g., Boudreau et al., 2008; Guo et al., 2010; Nelson et al., 2009), and ground-based lidar (e.g., Loudermilk et al., 2009; Ni-Meister et al., 2010). In these applications, statistical models were used to quantify the relationship between biomass measurements and vegetation structure metrics derived from lidar for a number of forest plots or stands. Their performance varies depending on the vegetation conditions, the density of field observations, and the approach used for statistical modeling.

Most of these existing studies have focused on the use of lidar-derived canopy structure metrics, such as height and canopy cover, for biomass estimation. However, studies of plant allometry suggested that biomass at the individual tree level is determined not only by canopy structure but also by factors such as trunk taper and wood density (Chave et al., 2006; Niklas, 1995), which are closely related to the floristic characteristics of the plants. As a result, biomass should be related to vegetation types. For example, Drake et al. (2003) examined the relationships between lidar metrics from an

airborne waveform lidar LVIS (Laser Vegetation Imaging Sensor) and AGB for two study sites in Central America, one in a tropical moist forest in Panama and the other in a tropical wet forest in Costa Rica. They found that the relationships between lidar metrics and AGB differ between these two sites even after the models had adjusted for the fraction of crown area that was deciduous (FCAD) of canopy trees. They attributed the differences to the underlying allometric relationships between stem diameter and AGB in tropical forests. Næsset and Gobakken (2008) estimated the aboveground and belowground biomass for 1395 sample plots in young and mature coniferous forests located in ten different areas within the boreal forest zone of Norway. With one canopy height metric and one canopy density metric derived from airborne discrete-return lidar, they were able to estimate aboveground and belowground biomass with R^2 of 0.82 and 0.77, respectively. When variables including tree species composition were included, the R^2 increased to 0.88 and 0.85. In a recent study, Ni-Meister et al. (2010) found that the relationships between biomass and canopy structure are distinctly different for deciduous and conifer trees in temperate forests in New England, U.S. Their analysis was based on the canopy structure information measured in the field as well as those derived from LVIS and Echidna® validation instrument (EVI), a ground-based lidar system.

The dependence of biomass-canopy structure relationship on vegetation types is well-known (e.g., Nelson et al., 1988; Ni-Meister et al., 2010). One approach for incorporating vegetation type information into biomass estimation is to stratify the forest plots according to vegetation types, for each of which a separate statistical model is developed (e.g., MacLean and Krabill 1986; Nelson et al., 1988). However, such an approach has practical and theoretical limitations. First, in most previous studies, only a limited number (typically 20–60 in total) of field plots were available for biomass modeling due to issues such as accessibility and cost. The stratification of a study area will lead to even fewer number of field plots per vegetation type, making it difficult to fit reliable statistical models for each vegetation type. Another problem of such an approach is that it assumes that the field data contains an exhaustive list of all vegetation types which exist in a given area. This is hardly true for natural forests because the vegetation types collected through field measurements are typically only a sample of the all vegetation types which exist over that area.

Recent advances in mixed-effects modeling can circumvent the aforementioned problems. In a conventional statistical model, the regression coefficients (such as intercept and slopes) are treated as constants. However, in mixed-effects models, these coefficients could be modeled as random Gaussian variables with their specific values varying among vegetation types. This approach makes it feasible to estimate biomass even when the sample size per vegetation type is small. Mixed-effects models have recently been used to estimate canopy height from satellite lidar (GLAS) data (Chen, 2010) and tree diameter from airborne discrete-return lidar data (Salas et al., 2010). Chen (2010) used mixed-effects model to test the generalizability of height estimation from GLAS data within and across three study sites in the Pacific coast region (one conifer site and one woodland site in California and another conifer site in Washington). He found significant random effects between the conifer and woodland sites but not between the two conifer sites. Salas et al. (2010) compared four statistical models including ordinary least squares (OLS), generalized least squares with a non-null correlation structure (GLS), linear mixed-effects model (LME), and geographically weighted regression (GWR) for estimating diameter of individual trees using discrete-return lidar data. They found that LME was significantly better than the other three models. Despite the promising results obtained in these two lidar remote sensing studies, no studies, to our best knowledge, have been done to explore the use of mixed-effects model for biomass estimation using lidar data.

In this study, vegetation types derived from aerial photographs are used to stratify forest for biomass modeling. Aerial photography is a fundamental remote sensing data source that possesses fine spatial and temporal details for producing base maps and performing environmental analysis (Lillesand et al., 2008). It has been widely used for mapping vegetation types for decades (e.g., Avery, 1978; Colwell, 1946; Fensham and Fairfax, 2002; Morgan et al., 2010). The recent advances in digital imaging and analysis also make aerial photography a rapidly-evolving tool for environmental analysis and ecological management (Morgan et al., 2010). In the U.S., a number of national programs such as NHAP (National High Altitude Program), NAPP (National Aerial Photography Program), NAIP (National Agriculture Imagery Program), and NDOP (National Digital Orthophoto Program) have collected and delivered aerial photographs every 3–10 years that cover the conterminous states from the late 1980s. Besides their wide temporal and spatial coverage, the aerial photographs acquired through these programs are usually free or at low cost for public use, making them ideal for detailed vegetation type mapping (Davies et al., 2010; Higinbotham et al., 2004).

The main goal of this study is to investigate whether integrating airborne lidar data with traditional vegetation maps derived from aerial photographs can improve biomass estimation for forest landscape in California. We specifically explore the efficacy of mixed-effects modeling to integrate the two remotely sensed data sources. We also compare the performance of two common but different approaches to vegetation classification.

2. Study area and data

2.1. Study area

Our study area is located in the United States Forest Service Sagehen Creek Experimental Forest in California, which covers approximately 3925 ha and is on the eastern slope of the Sierra Nevada approximately 32 km north of Lake Tahoe (Fig. 1). Conifer species present include white fir (*Abies concolor*), red fir (*Abies magnifica*), mountain hemlock (*Tsuga mertensiana*), lodgepole pine (*Pinus contorta*), Jeffrey pine (*Pinus jeffreyi*), sugar pine (*Pinus lambertiana*), and western white pine (*Pinus monticola*) (Table 1). Non-forested areas include fens, wet and dry montane meadows and shrub fields. Elevation ranges from 1862 m to 2670 m with slopes averaging 18% but can reach 70% in parts of the watershed.

2.2. Field data collection

A systematic grid of geo-referenced 0.05 ha circular plots was installed with a random starting location (Fig. 2). The grid consists of three sampling densities, 500 m, 250 m, and 125 m spacing. The entire watershed was sampled by plots spaced on a 500 m interval. Areas not occupied by Jeffrey pine plantations were further sampled at 250 m spacing; 125 m spacing was used in 10 unique forest types to conduct high density sampling. A total of 523 plots were established in the field between 2004 and 2006. These field plots were located with a handheld Garmin eTrex recreational GPS with horizontal accuracy of 3 to 11 m, which are called RGPS plots hereinafter. Nine of the ten locations of 125 m plot spacing were revisited in 2006 and a Trimble® GeoXH™ handheld GPS with Zephyr Geodetic antenna was used to re-measure the center of 81 individual plots. The average horizontal accuracy of the new GPS measurements is 0.1 m with the majority <0.2 m and, at the worse case, 1.5 m. These plots are DGPS plots hereinafter.

At each plot, all trees greater than 5 cm in diameter at breast height (DBH, breast height = 1.37 m) were measured with a nested sampling design. Canopy trees (≥19.5 cm DBH) were tagged and measured in the whole plot; Understory trees (≥5 cm DBH to <19.5 cm DBH) were measured in a randomly selected third of the plot. Tree measurements include species, DBH, tree height, and vigor. Vigor was defined into six different classes: 1) healthy trees with no visible defects, 2) healthy trees with minimal damage or
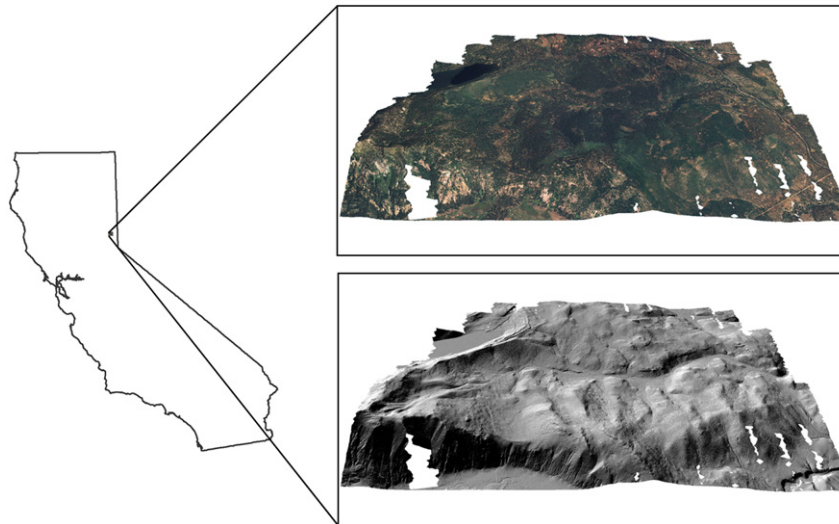
Fig. 1. Location of the study area. Top-right: Aerial photographs draped over the lidar DEM. Bottom-right: A hillshade of the lidar DEM.

defect (broken top/dead top, abnormal lean, etc.), 3) live trees that are near death or will be dead in the next five years, 4) recently dead trees with little decay and that retain their bark, branches and top, 5) trees that show some decay and have lost some bark, branches and may have a broken top, and 6) extensive decay and missing bark and most branches and have a broken top. The first three vigor classes are for live trees and the last three are for dead trees.

2.3. Lidar data

Lidar data were collected from September 14 to 17, 2005 for the study area using an Optech ALTM 2050 system on an airplane flying at an altitude of ~800 m and average velocity of 260 km per hour. The ALTM 2050 acquired up to three returns per pulse at a pulse frequency of 50 kHz, scan frequency of 38 Hz, and a maximum scan angle of 15°, creating a swath width of ~580 m. The point density is about 2–4 returns per square meter. Optech, Inc. rates the RMSE precision of individual point locations surveyed by the ALTM 2050 as ± 15 cm vertical and ± 50 cm horizontal.

2.4. Vegetation types from aerial photographs

USDA Forest Service (USFS) provided a vegetation type map, which was produced by visually interpreting 1 m NAIP (National Agricultural Imagery Program) Digital Orthophoto Quadrangles (scale 1:15,840, natural color) taken on September 16, 2005 and manually delineating the vegetation polygons. The vegetation polygons were initially typed using the CALVEG (Classification and Assessment with Landsat of Visible Ecological Groupings) classification system (USDA, 1981),

which is a provisional system that meets the floristically based level of the U.S. National Vegetation Classification Standard (NVCS) hierarchy. The CALVEG system was designed to classify California's existing vegetation communities and the CALVEG types are also called "Dominant Types" in accordance with the USFS Existing Vegetation Classification and Mapping Technical Guide (Brohman and Bryant, 2005). The CALVEG types were crosswalked to other classification systems including SAF (Society of American Forester) (Eyre, 1980), CWHR (California Wildlife-Habitat Relationships) (Meyer and Laudenslayer, 1988), and U.S. NVC (National Vegetation Classification) alliance-level vegetation types (FGDC, 2008).

In this study, the two national-wide vegetation classification systems, SAF and NVC alliance-level vegetation types, were chosen for biomass estimation due to their broad applicability (Fig. 3). The NVC alliance-level vegetation types are based on NVCS, which establishes national procedures for field plot records and classification of existing vegetation types for the United States. These procedures provide a dynamic and practical way to publish new or revised descriptions of vegetation types while maintaining a current, authoritative list of types for multiple users to access and apply (Jennings et al., 2009). The early efforts of NVC started in 1994 and the first NVCS was adopted in 1997 by FDGC (Federal Geographic Data Committee). As early as of April 1997, a total of 1571 NVC types had been identified at the alliance-level (Grossman et al., 1998). Since then, the vegetation classification has been continuously evolving and updated (Jennings et al., 2009). In contrast to NVC that uses all vascular plant species present in a community to help define vegetation classes, the SAF types emphasize dominant species of a stand. In many cases, the SAF types are more broadranging over both structural and environmental gradients than are the

Table 1
Common tree species in this study area and their allometric equations for calculating biomass. BAT = total above ground biomass; BST = biomass of stem with bark; BSW = biomass of stem without bark; CIR = stem basal circumference; DBH = diameter at breast height; HT = tree height.

Species	Abbr.	Common name	Equation	Units (biomass, DBH or CIR, height)	Source
Abies concolor	ABCO	White fir	$\ln(\text{BST}) = 3.011904 + 2.7727 \times \ln(\text{DBH})$	g, cm, –	Halpern and Means, 2004
Abies magnifica	ABMA	Red fir	$\ln(\text{BST}) = 3.020046 + 2.7590 \times \ln(\text{DBH})$	g, cm, –	Halpern and Means, 2004
Juniperus occidentalis	JUOC	Sierra juniper	$\ln(\text{BSW}) = -8.5802 + 2.6389 \times \ln(\text{CIR})$	kg, cm, –	Means et al., 1994
Pinus contorta	PICO	Lodgepole pine	$\ln(\text{BST}) = -9.10508 + 2.3363 \times \ln(\text{DBH})$	mg, cm, –	Means et al., 1994
Pinus jeffreyi	PIJE	Jeffrey pine	$\ln(\text{BST}) = 1.817891 + 2.952 \times \ln(\text{DBH})$	g, cm, –	Halpern and Means, 2004
Pinus lambertiana	PILA	Suger pine	$\ln(\text{BST}) = 3.229148 + 2.6863 \times \ln(\text{DBH})$	g, cm, –	Halpern and Means, 2004
Pinus monticola	PIMO	Western white pine	$\text{BAT} = 20,800 + 0.1544 \times (\text{DBH}^2 \times \text{HT})$	g, cm, cm	Halpern and Means, 2004
Populus tremuloides	POTR	Quaking aspen	$\ln(\text{BAT}) = -2.6224 + 2.4827 \times \ln(\text{DBH})$	kg, cm –	Jenkins et al., 2004
Tsuga mertensiana	TSME	Mountain hemlock	$\ln(\text{BAT}) = -10.1688 + 2.5915 \times \ln(\text{DBH})$	mg, cm, –	Jenkins et al., 2004
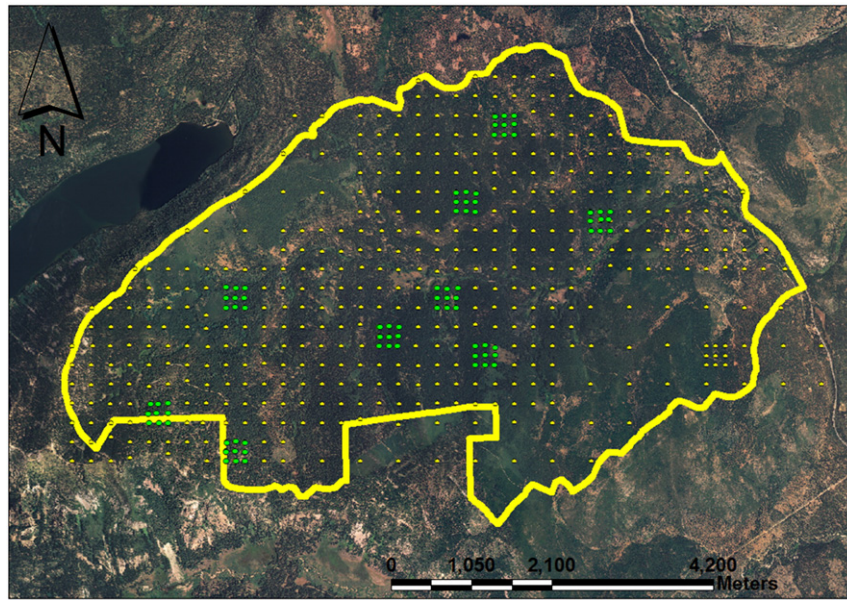
Fig. 2. Field plots of vegetation measurements. The smaller dots indicate the plots located with a recreational GPS. The larger dots indicate the plots located with both a recreational GPS and a differential GPS. The thick line is the boundary of the vegetation type map.

alliances recognized in NVC (Grossman et al., 1998), so in total a much smaller number of SAF types (86 forest types) have been identified for the whole United States.

3. Methods

3.1. Biomass calculation at the plot-level

Biomass can be most accurately calculated using species-specific allometric equations. A comprehensive review of the literature was conducted to search species-specific allometric equations and, during the selection process, preference was given to equations meeting all or most of the following criteria: 1) being derived from a high number (~40–100) of sample trees, 2) from DBH ranges similar to those in our dataset, 3) from geographical sites most similar to our study location, and 4) including all or the most relevant biomass components of a tree. The final equations we selected are from Halpern and Means (2004), Jenkins et al. (2004), and Means et al. (1994) (see Table 1). To derive the biomass at the plot level, we summed the biomass of live trees with DBH > 5 cm (the total biomass of understory trees was multiplied by three given that only a random third of each plot was measured for them) and converted the biomass total to density based on the area of each plot. We only consider live trees because the dead trees usually have few or no leaves and thus generate much fewer laser returns.

3.2. Lidar data processing

The first step of lidar data processing is to filter the raw lidar points and separate them into ground and non-ground returns (Chen et al., 2007). Then, the ground returns identified were interpolated to generate a Digital Elevation Model (DEM) of 1 m cell size. The canopy height of individual points was calculated as the difference between their original Z values and the corresponding DEM cell elevations. Based on the canopy height, the following statistics were calculated for all points within a given field plot: mean (h_u), standard deviation (h_{std}), skewness (h_{skn}), and kurtosis (h_{kurt}); proportion of lidar points within different height bins (0 to 5 m, 5 to 10 m, …, 45 to 50 m, and > 50 m, denoted as p_{0to5}, p_{5to10}, …, p_{45to50}, and $p_{>50}$, respectively); percentile heights (5, 10, …, 100 percentile, denoted as h_5, h_{10}, …, h_{100}, respectively; note that 100

percentile height corresponds to maximum height); and quadratic mean height (h_{qm}) (see Table 2). The quadratic mean height was calculated as Lefsky et al. (1999). Two sets of lidar metrics were generated: one is based on all lidar returns and the other is based on first returns since some studies have found that first returns may have better performance in predicting vegetation attributes (e.g., Kim et al., 2009). All of the above lidar data processing was conducted using the Tiffs (Toolbox for Lidar Data Filtering and Forest Studies) software (Chen, 2007).

3.3. Statistical analysis

The mixed-effects model used to predict plot-level biomass from lidar metrics and vegetation types is as follows:

$$
\begin{aligned}
\mathbf{Y} &= \mathbf{Xb} + \mathbf{Zb} + \mathbf{e} \\
\mathbf{b} &\sim N(\mathbf{0}, \mathbf{G}) \\
\mathbf{e} &\sim N(\mathbf{0}, \mathbf{R}) \\
\mathrm{cov}(\mathbf{b}, \mathbf{e}) &= 0
\end{aligned}
\tag{1}
$$

where $\mathbf{Y}$ is a vector of biomass for n field plots, $\mathbf{X}$ is the $n \times p$ design matrix for the p fixed effects, $\mathbf{Z}$ is the $n \times q$ design matrix for q random effects, β is a $p \times 1$ vector for the fixed effects, $\mathbf{b}$ is $q \times 1$ vector for the q random effects, and ε is the $n \times 1$ vector for the error random effects. Note that 1) the random effect vector $\mathbf{b}$ has Gaussian (Normal) distributions with zero means and variance–covariance matrix $\mathbf{G}$, which is called the G covariance structure; 2) the error vector ε could be correlated with variance–covariance matrix $\mathbf{R}$, which is modeled with variograms in this study; and 3) the random effects $\mathbf{b}$ and ε are independent. Given that there usually exist power–law relationships between biomass and other vegetation attributes such as DBH or height (Zianis and Mencuccini, 2004), biomass and all lidar metrics were log-transformed so that the developed models are linear at the log-scale.

We used stepwise regression to select the statistically significant lidar metrics for predicting biomass. Since both the response and predictor variables are at the log-scale, the developed models are *multiplicative* at the original scale. The multiplicative models served as the benchmark and starting point for developing mixed-effects models; in other words, we added and tested random effects only for the lidar metrics selected in the multiplicative models. Modeling variance structure is
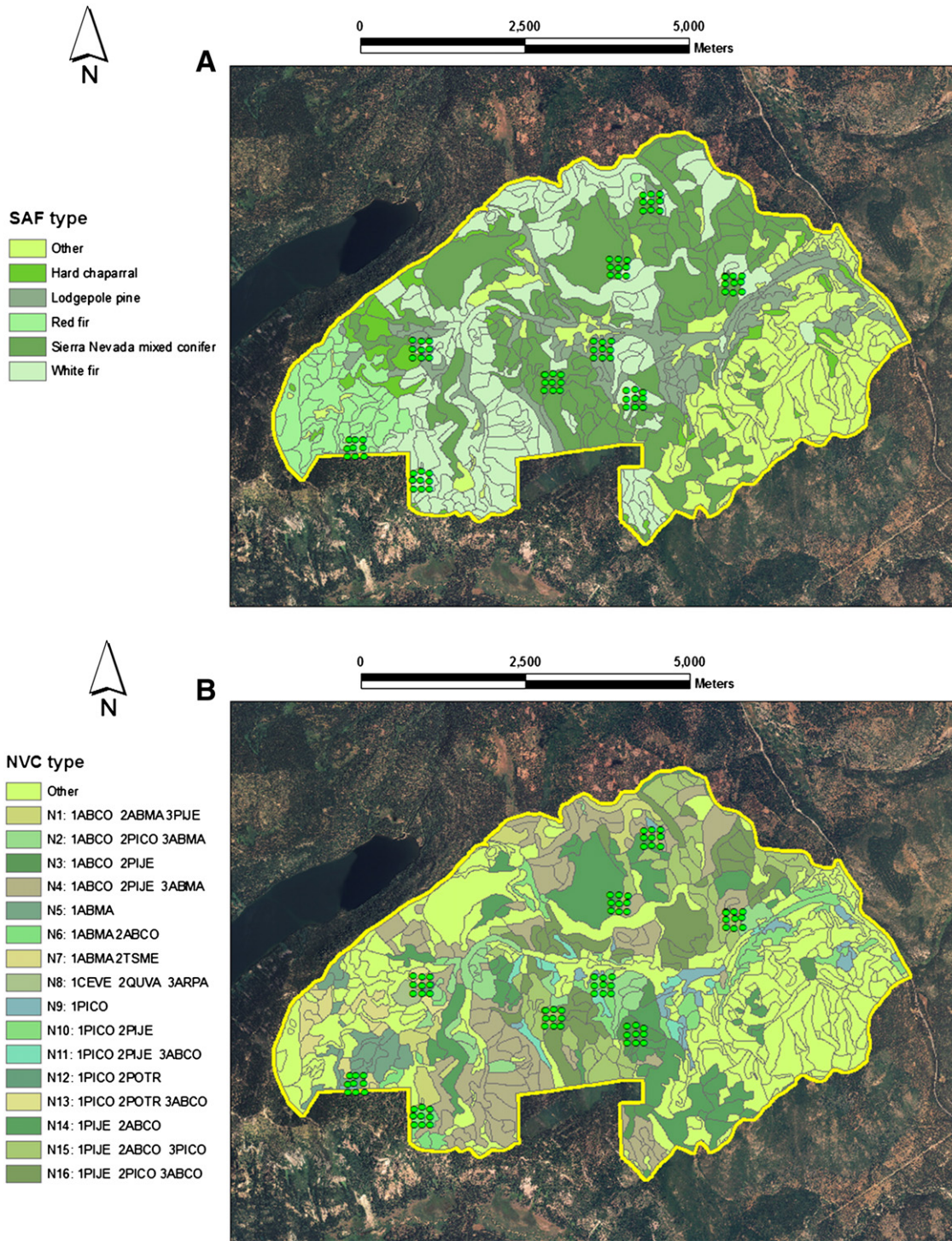
Fig. 3. Vegetation type maps of the study area. (a) SAF type, (b) NVC alliance-level type.

probably the most powerful and critical feature of mixed-effects models, which allows correlation among observations. To find the most parsimonious yet effective **G** covariance structure, we initially fit a model with all predictor variables having random effects and their covariance matrix being unstructured (UN), then fit models with a Variance Components (CV) covariance structure, which means that the individual random effects are independent and the off-diagonal elements of the covariance matrix are zeros. If the estimate of any random effects is statistically insignificant from zero across all different vegetation types, the random

effect was dropped from the model. A total of four different types of variogram models (exponential, spherical, Gaussian, and Matern) were tested to model the spatial dependence of the residuals and calculate the variance–covariance matrix **R**. AIC (Akaike Information Criteria) was used to help select the best models, which usually have the lowest AIC. However, if the AIC values of two models have a difference less than 2, such models are considered indistinguishable (Burnham and Anderson, 2002). Once the best models had been selected, leave-one-out cross-validation was used to calculate the model coefficient of

Table 2
Lidar metrics for predicting forest attributes.

Lidar metrics	Description
h_u, h_{std}, h_{skn}, h_{kurt}	Mean, standard deviation, skewness, kurtosis of height of lidar points
p_{0to5}, p_{5to10}, ..., p_{45to50}, $p_{>50}$	Proportion of lidar points within height bins (0 to 5 m, 5 to 10 m, ..., 45 to 50 m, and >50 m)
h_5, h_{10}, ..., h_{100},	Percentile height of lidar points
h_{qm}	Quadratic mean height of lidar points

determination (R^2) and RMSE so that a straightforward comparison can be made between the results from this study and those from others. We used SAS 9.1.3 (SAS Institute Inc.) to fit mixed-effects models.

Among the 81 DPGS plots, one has questionable GPS accuracy and another three plots are outside of the vegetation type map so they are excluded from our analysis. When stepwise regression was used to estimate biomass for the remaining 77 DGPS plots, it was found that four plots have large residuals (>3 standard deviations). After a careful examination of the tree characteristics of the four plots and inspection of their corresponding point clouds, it was found that there was obvious mismatch of tree information (e.g., tree density, size) between lidar point clouds and field data for three plots. It is suspected that there might be large errors of plot coordinates or vegetation measurements in the field data of these three plots, so they were excluded from our analysis as well. However, the remaining plot was kept since no distinct mismatch can be identified, resulting in a total of 74 DGPS plots in our ground truth data. Table 3 shows the cross-tabulation of the 74 DPGS plots in the NVC alliance-level and SAF vegetation type classification systems. We developed mixed-effects models based on two vegetation types (SAF vs. NVC alliance-level) and two sets of lidar metrics (derived from all returns vs. first returns), which lead to a total of four sets of mixed-effects models for the DGPS plots.

4. Results

When the lidar metrics from all returns were used for the 74 DGPS plots, the two-way stepwise regression (with an enter probability of

0.05 and leave probability of 0.1) selected two lidar metrics h_{qm} and p_{35to40} in the multiplicative model:

$$\ln(AGB) = 1.571\ln\left(h_{qm}\right) + 0.055\ln(p_{35to40}) + 2.066 \qquad (2)$$

where AGB is the aboveground live tree biomass in Mg/ha and h_{qm} is the quadratic mean height in meters, and p_{35to40} is the proportion of lidar points between 35 and 40 m. Starting with the two lidar metrics selected in Eq. (2) and using the SAF vegetation types, we follow the procedure described in Section 3.3 to develop and test mixed-effects models (see DGPS.A.SAF.M1-7 in Table 4 for the models developed). When both h_{qm} and p_{35to40} are modeled as random effects, it was found that the model with the variance components (VC) covariance structure of random effects (model DGPS.A.SAF.M2) produced much smaller AIC compared to the one with the unconstructed (UN) covariance matrix (model DGPS.A.SAF.M1, Table 4), indicating that model DGPS.A.SAF.M2 should be preferred. An examination of model DGPS.A.SAF.M2 revealed that (1) the estimates of the random effects of intercept and the metric p_{35to40} are zeros and (2) the fixed-effects p_{35to40} is not statistically significant. So, the lidar metric p_{35to40} was removed and no random effect for intercept was modeled, resulting in model DGPS.A.SAF.M3. This further reduced the AIC to 23.4 compared to the AIC of 26.3 from model DGPS.A.SAF.M2. Starting with model DGPS.A.SAF.M3, four different variogram models (exponential, spherical, Gaussian, and Matern) were used to model the variance-covariance matrix **R** (models DGPS.A.SAF.M4-7). It was found that these models have higher AICs (models DGPS.A.SAF.M5-7) or very small (=0.2) AIC differences (model DGPS.A.SAF.M4) compared to model DGPS.A.SAF.M3. This indicates that, after incorporating the fixed and random effects in model DPGS.A.SAF.M3, the residuals of AGB have no significant spatial autocorrelation at the scale of current minimal plot spacing (125 m) or larger. As a result, model DGPS.A.-SAF.M3 was chosen as the final mixed-effects model in this case of using SAF vegetation type and the lidar metrics from all returns for the 74 DGPS plots.

Similarly, we developed models for the cases of using 1) SAF vegetation types and lidar metrics from first returns (see models DPGS.F.SAF.M1-7 in Table 4), 2) NVC alliance-level vegetation types and lidar metrics from all returns (see models DGPS.A.NVC.M1-7 in Table 5), and 3) NVC alliance-level vegetation types and lidar metrics

Table 3
Cross-tabulation of 74 DGPS plots in two vegetation classification systems: NVC alliance-level type (rows N1–N16) and SAF type (HRC-Hard Chaparral, LPN-Lodgepole Pine, RFR-Red Fir, SMC-Sierra Nevada Mixed Conifer, WFR-White Fir). The name of each NVC alliance-level type lists 1–3 dominant tree species in that type. See Table 1 for the abbreviated species names in each NVC type.

NVC type	SAF type					
	HRC	LPN	RFR	SMC	WFR	Total
N1: 1ABCO 2ABMA 3PIJE	0	0	0	0	1	1
N2: 1ABCO 2PICO 3ABMA	0	0	0	0	4	4
N3: 1ABCO 2PIJE	0	0	0	0	5	5
N4: 1ABCO 2PIJE 3ABMA	0	0	0	0	13	13
N5: 1ABMA	0	0	4	0	0	4
N6: 1ABMA 2ABCO	0	0	0	0	7	7
N7: 1ABMA 2TSME	0	0	1	0	0	1
N8: 1CEVE 2QUVA 3ARPA	2	0	0	0	0	2
N9: 1PICO	0	3	0	0	0	3
N10: 1PICO 2PIJE	0	7	0	0	0	7
N11: 1PICO 2PIJE 3ABCO	0	1	0	0	0	1
N12: 1PICO 2POTR	0	1	0	0	0	1
N13: 1PICO 2POTR 3ABCO	0	3	0	0	0	3
N14: 1PIJE 2ABCO	0	0	0	8	0	8
N15: 1PIJE 2ABCO 3PICO	0	0	0	10	0	10
N16: 1PIJE 2PICO1 3ABCO	0	0	0	4	0	4
Total	2	15	5	22	30	74

Table 4
Different mixed-effects models of biomass estimation based on differential GPS plots (denoted as DGPS in the model no.), lidar metrics derived from all or first returns (A or F in the model no.), and SAF forest cover type. UN means that the G covariance matrix is unconstructed; VC means that the Variance Components matrix is used as the G covariance structure. The best model in each set is bolded.

Model no.	Fixed effects	Random effects	G cov. structure*	Variogram model	AIC
DGPS.A.SAF.M1	Intercept, h_{qm}, p_{35to40}	Intercept, h_{qm}, p_{35to40}	UN	None	49.2
DGPS.A.SAF.M2	Intercept, h_{qm}, p_{35to40}	Intercept, h_{qm}, p_{35to40}	VC	None	26.3
DGPS.A.SAF.M3	**Intercept, h_{qm}**	**h_{qm}**	**VC**	**None**	**23.4**
DGPS.A.SAF.M4	Intercept, h_{qm}	h_{qm}	VC	Exponential	23.2
DGPS.A.SAF.M5	Intercept, h_{qm}	h_{qm}	VC	Spherical	23.8
DGPS.A.SAF.M6	Intercept, h_{qm}	h_{qm}	VC	Gaussian	23.6
DGPS.A.SAF.M7	Intercept, h_{qm}	h_{qm}	VC	Matern	25.2
DGPS.F.SAF.M1	Intercept, h_{qm}, h_{40}, p_{35to40}	Intercept, h_{qm}, h_{40}, p_{35to40}	UN	None	50.7
DGPS.F.SAF.M2	Intercept, h_{qm}, h_{40}, p_{35to40}	Intercept, h_{qm}, h_{40}, p_{35to40}	VC	None	28.9
DGPS.F.SAF.M3	**Intercept, h_{qm}**	**h_{qm}**	**VC**	**None**	**26.0**
DGPS.F.SAF.M4	Intercept, h_{qm}	h_{qm}	VC	Exponential	26.0
DGPS.F.SAF.M5	Intercept, h_{qm}	h_{qm}	VC	Spherical	25.9
DGPS.F.SAF.M6	Intercept, h_{qm}	h_{qm}	VC	Gaussian	25.5
DGPS.F.SAF.M7	Intercept, h_{qm}	h_{qm}	VC	Matern	27.5

Table 5
Different mixed-effects models of biomass estimation based on differential GPS plots (DGPS in the model no.), lidar metrics derived from all or first returns (A or F in the model no.), and NVC alliance-level vegetation type. UN means that the G covariance matrix is unconstructed; VC means that the Variance Components matrix is used as the G covariance structure. The best model in each set is bolded.

Model no.	Fixed effects	Random effects	G cov. structure*	Variogram model	AIC
DGPS.A.NVC.M1	Intercept, h_{qm}, p_{35to40}	Intercept, h_{qm}, p_{35to40}	UN	None	104.0
DGPS.A.NVC.M2	Intercept, h_{qm}, p_{35to40}	Intercept, h_{qm}, p_{35to40}	VC	None	31.8
DGPS.A.NVC.M3	**Intercept, h_{qm}**	**h_{qm}**	**VC**	**None**	**25.8**
DGPS.A.NVC.M4	Intercept, h_{qm}	h_{qm}	VC	Exponential	25.2
DGPS.A.NVC.M5	Intercept, h_{qm}	h_{qm}	VC	Spherical	25.9
DGPS.A.NVC.M6	Intercept, h_{qm}	h_{qm}	VC	Gaussian	25.5
DGPS.A.NVC.M7	Intercept, h_{qm},	h_{qm}	VC	Matern	27.7
DGPS.F.NVC.M1	Intercept, h_{qm}, h_{40}, p_{35to40}	Intercept, h_{qm}, h_{40}, p_{35to40}	UN	None	161.4
DGPS.F.NVC.M2	Intercept, h_{qm}, h_{40}, p_{35to40}	Intercept, h_{qm}, h_{40}, p_{35to40}	VC	None	33.7
DGPS.F.NVC.M3	**Intercept, h_{qm}**	**h_{qm}**	**VC**	**None**	**28.8**
DGPS.F.NVC.M4	Intercept, h_{qm}	h_{qm}	VC	Exponential	28.3
DGPS.F.NVC.M5	Intercept, h_{qm}	h_{qm}	VC	Spherical	28.6
DGPS.F.NVC.M6	Intercept, h_{qm}	h_{qm}	VC	Gaussian	28.5
DGPS.F.NVC.M7	Intercept, h_{qm},	h_{qm}	VC	Matern	30.5

from first returns (see models DGPS.F.NVC.M1-7 in Table 5). With the same rationale as above, we selected the best models for these three cases, which are model DPGS.F.SAF.M3, DGPS.A.NVC.M3, and DGPS.F.NVC.M3, respectively. Note that the multiplicative models based on first returns include an additional lidar metric, h_{40}. However, in the corresponding mixed-effects models, this metric is not statistically significant any more. As a result, all best mixed-effects models have the same model structure in terms of the fixed effects, random effects, **G** covariance structure, and **R** covariance matrix (i.e., variogram models).

Table 6 summarizes the fitting statistics of multiplicative and the best mixed-effects models developed based on the 74 DGPS plots. The examination of fitting statistics of the mixed-effects models indicates that all mixed-effects models, based on either NVC alliance-level or SAF vegetation types, outperformed the corresponding multiplicative models. For example, when all returns were used, the R^2 increased from 0.77 to 0.83 for NVC alliance-level types and to 0.82 for SAF types. The RMSE decreased by about 10% for all returns and by about 5% for first returns. Among all models based on the DGPS

Table 6
Model fitting statistics calculated with leave-one-out cross validation for multiplicative and mixed-effects models.

	Multiplicative model		Mixed effects model			
			NVC alliance-level type		SAF type	
	R^2	RMSE (Mg/ha)	R^2	RMSE (Mg/ha)	R^2	RMSE (Mg/ha)
DGPS plots (n = 74)						
All returns	0.77	80.8	0.83	72.2	0.82	72.8
First returns	0.77	80.2	0.81	74.5	0.81	75.1
RGPS plots (n = 74)						
All returns	0.66	98.7	0.70	94.0	0.72	92.5
First returns	0.67	97.4	0.70	95.2	0.68	98.2

plots, the mixed-effects model based on all returns and NVC alliance-level vegetation type (model DGPS.A.NVC.M3) has the highest R^2 (0.83) and the lowest RMSE (72.2 Mg/ha). However, since the model based on all returns and SAF vegetation type (model DGPS.A.-SAF.M3) has almost the same R^2 (0.82) and RMSE (72.8 Mg/ha) as model DGPS.A.NVC.M3 while using a smaller number of vegetation classes (5 SAF classes instead of 16 NVC classes), it was considered as the best model from the aspects of both model parsimony and fitting statistics. Fig. 4 shows the biomass map of the study area based on model DGPS.A.SAF.M3.

5. Discussion

5.1. Comparison with previous studies

Our results indicate that the mixed-effects models have better performance than the corresponding fixed-effects models. This finding is consistent with previous studies that used other remotely-sensed data: Meng et al. (2007) used NDVI derived from Landsat ETM + imagery and forest inventory data to develop a linear fixed-effects model and linear mixed-effects models to estimate merchantable biomass for the state of Georgia. They found that the linear mixed-effects model with random effects in both intercept and slope best fits the data and achieved a R^2 of 0.57 while the fixed-effects model produced a R^2 of 0.31 only.

Some previous studies found that the integration of lidar data and optical or radar imagery does not necessarily produce better results in biomass modeling. For example, Hyde et al. (2006) found that the addition of Quickbird and SAR/InSAR structure metrics (such as NDVI and backscatter intensity) to LVIS (Laser Vegetation Imaging Sensor) resulted in no improvement for estimating biomass across 120 one-hectare circular plots in the Sierra Nevada of California. This was explained by the fact that the structure metrics from lidar, radar, and Quickbird are redundant (Hyde et al., 2006). Using different inputs (categorical vegetation types instead of continuous structure metrics such as NDVI) and statistical approaches (mixed-effects instead of fixed-effects models), we found that it is possible to improve biomass estimation by integrating lidar and optical remote sensing data. The difference between this study and Hyde et al. (2006) might be attributed to our different modeling strategy (i.e., considering the biomass dependence on vegetation types) and our use of mixed-effects models and vegetation types, but more research is needed to further investigate this issue.

The two multiplicative models have the same R^2 (0.77) and similar RMSE (80.8 Mg/ha for all returns; 80.2 Mg/ha for first returns). These fitting statistics are comparable to those from Gonzalez et al. (2010), which used the lidar data collected by the same lidar system (Optech ALTM 2050). They used field measurements of 39 plots collected by a modified FIA (Forestry Inventory and Analysis) design to develop stepwise regression model to estimate aboveground live tree biomass in North Yuba in the Tahoe National Forest in California, a site only ~50 km away from our study area. Their final model included lidar metrics such as quadratic mean height and five percentile heights (p10, p20, p30, p40, and p50), with R^2 of 0.80, slighter higher than our fixed-effects model. However, their model RMSE is 123 Mg/ha, much larger than 80.8 Mg/ha of our fixed-effects model (Eq. 2). The causes of the large RMSE difference between this study and Gonzalez et al. (2010) are multifaceted: besides using fixed- instead mixed-effects models, their study uses field plots consisting of four subplots of 17.95 m radius while we use single plots of 12.62 m radius: the larger field plot introduces more variability of canopy structure, making it more difficult to characterize using a single set of metrics; another reason for the larger RMSE in Gonzalez et al. (2010) is that they incorporated uncertainty in their field biomass using Monte Carlo simulations.
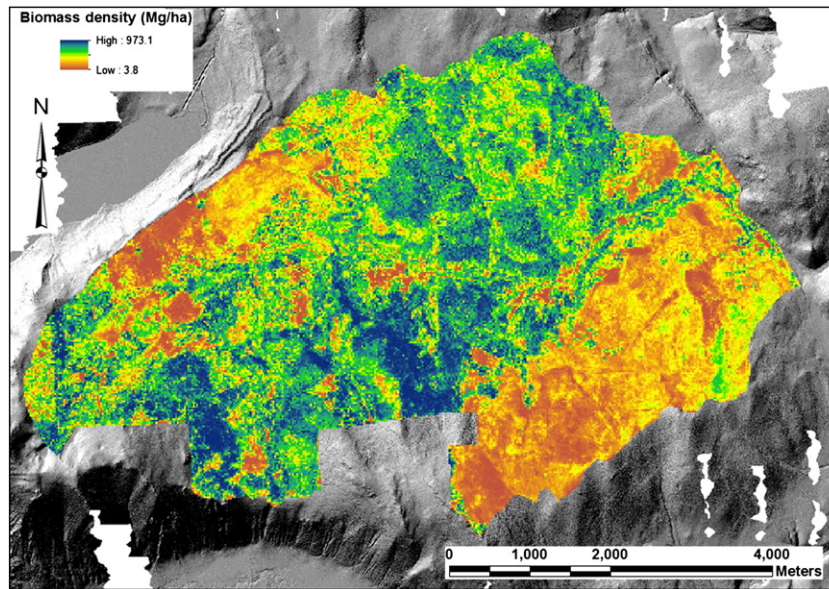
Fig. 4. Biomass map of the study area (based on model DGPS.A.SAF.M3).

5.2. Advantages of mixed-effects models

When vegetation types or other information are available to partition a study area into different strata, an alternative approach is to fit a statistical model for each stratum. Compared to the approach of fitting individual stratum-specific regression models, mixed-effects models have the advantage of using fewer parameters while possibly achieving comparable or even better performance. For instance, Meng et al. (2007) compared a mixed-effects model with the approach of fitting an individual regression model within each region (IRR). They divided the state of Georgia in USA into five eco-regions, each including 10 to 67 counties. The mixed-effects models were developed at the county-level, so the minimal sample size of each stratum (eco-region) is 10. They found that the mixed-effects model obtained slightly better performance than the IRR approach even though the mixed-effects model used much fewer parameters.

In our study, the need to use mixed-effects models instead of fitting individual vegetation type specific statistical models is obvious because we have very limited numbers of plots per vegetation type (see Table 3). For example, among the 16 NVC alliance-level vegetation types, 11 types have 5 or less field plots associated with each. Using the SAF classification system leads to fewer types and thus higher average number of plots per vegetation type. However, there are only 2 plots for the Hard Chaparral (HRC) and 5 plots for the Red Fir (RFR) types. Fitting statistical models for such small samples is clearly questionable from the statistical standpoint (Green, 1991).

Mixed-effects models deal with the biomass dependence on vegetation types from a different perspective: the coefficients of the biomass models for different vegetation types could be assumed to vary as random Gaussian variables. This assumption puts a constraint on the variability of model coefficients and prevents unreliable estimates of model coefficients from being produced even when the sample size is small. Take model DGPS.A.SAF.M3 as an example (see Table 4), which is essentially a random slope model with the coefficient of h_{qm} (at the log scale) varying among different vegetation types:

$$\ln AGB_{ij} = 1.6971 \ln\left(h_{qm,ij}\right) + b_i^* \ln\left(h_{qm,ij}\right) + 1.3860 \qquad (3)$$

where AGB_{ij} is the aboveground live tree biomass for plot j of vegetation type i; $h_{qm,ij}$ is the quadratic mean height of all lidar points for the plot j of vegetation type i. b_i is the random coefficient estimated with the empirical best linear unbiased predictions (EBLUPs) for vegetation type i and it represents the estimated deviation from the mean slope (i.e., 1.6971). Fig. 5 shows the estimated biomass models for different vegetation types. These regression lines could be much different from the ones derived from vegetation type specific regression models. For instance, a regular "least squares" regression model for the HRC vegetation type will create a line that passes through the two HRC plots; such a line will be highly sensitive to the small sample size problem and thus will have less generalization ability for prediction. Instead, in the mixed-effects models, the regression line of a given vegetation type is a combination of a) the coefficients of the fixed effects (1.3860 for intercept, and 1.6971 for slope for this example as shown in Eq. 3), and b) the estimated coefficients of the random effects (b_i in Eq. 3). The mean regression line (determined by the coefficients of the fixed effects) can be thought as an initial estimate for the regression model of a specific vegetation type, much like the *prior* estimate in Bayesian statistics. This is the essential reason why mixed-effects models could be less susceptible to the small sample size issue. Additionally, if more samples are available for a given vegetation type, mixed-effects modeling will take advantage of the available sample data and the estimated model will be closer to the one derived from the regular least square regression, exemplified by the models for the WFR, SMC, and LPN vegetation types shown in Fig. 5. This explains why mixed-effects models are effective in modeling biomass when vegetation types of a wide range of sample sizes exist.

5.3. NVC alliance-level versus SAF vegetation types

One of the interesting results from this study is the lack of differences between the mixed-effects models developed from the two vegetation types (NVC alliance-level vs. SAF). The differences are less than 0.01 for R^2 values and less than 0.6 Mg/ha for RMSE. As introduced in Section 2.4, NVC alliance-level types define a vegetation class based on all vascular plants present while SAF types are defined by the dominant species. Thus, NVC alliance-level types represent a finer scale of vegetation classification. However, from the perspective of biomass estimation, the more coarsely scaled SAF types perform nearly as well because the dominant species account for the vast majority of AGB. These results suggest that we can focus classification and mapping schemes on the dominant species for mixed-effects modeling of biomass.
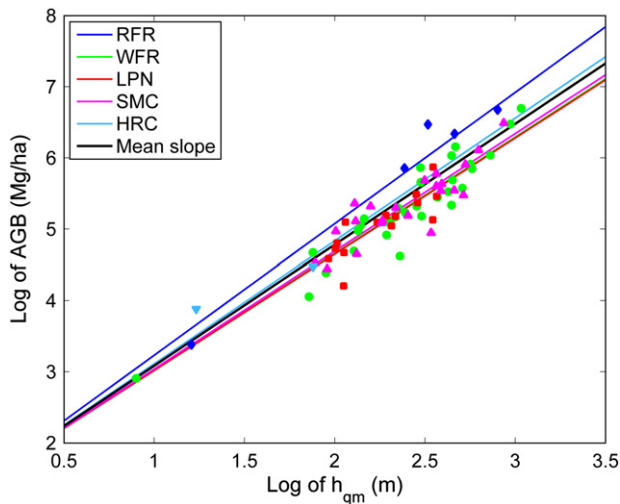
Fig. 5. Mixed-effects model of biomass estimation kmodel DGPS.A.SAF.M3). RFR-Red Fir, WFR-White Fir, LPN-Lodgepole Pine, SMC-Sierra Nevada Mixed Conifer, HRC-Hard Chaparral.

5.4. Lidar metrics from all returns vs. first returns

A few studies have also used lidar metrics derived from first returns only for predicting biomass. For example, Hall et al. (2005) found that a canopy cover metric derived from first returns was able to predict the foliage biomass and total aboveground biomass in a Ponderosa pine forest in Colorado with R^2 of 0.79 and 0.74, respectively. Kim et al. (2009) found that the model using first returns improved the R^2 by 0.1 for predicting the total aboveground biomass in a mixed coniferous forest in Arizona compared to the one using all returns.

Our results indicate that using first returns reduced RMSE by 0.6 Mg/ha compared to using all returns, which are in line with the findings from Kim et al. (2009). However, the improvement is too small to be considered statistically significant. It is interesting that, among the mixed-effects models, the ones based on all returns outperformed the ones based on first returns. The R^2 increased from 0.81 to 0.83 for NVC alliance-level type and, correspondingly, the RMSE decreased by 3%. The reasons for this contrasting pattern and the discrepancy between this study and previous studies are unclear, but it might be related to specific forest conditions and the way with which the specific lidar system generates individual returns (Wagner et al., 2007).

5.5. Biomass estimation using field data located with a recreational GPS

The accuracy of remote sensing based biomass maps is influenced not only by the specific earth observation data and the statistical approaches used, but also by the accuracy of the calibration data used for deriving models and estimates. One of the common problems in field data is the low geo-location accuracy caused by either the use of low-cost GPS or the existence of dense forests. Dominy and Duncan (2001) reported the difficulty of quality satellite reception beneath a dense forest canopy, with the degree of spatial error seriously affecting fine-scale vegetation mapping. Miura and Jones (2010) used a Garmin eTrex GPS (average ± 5.5 m horizontal error) to locate the centers of 25-m radius circular plots for field measurements and related to airborne lidar data. They had to manually shift the plots to achieve a better registration between lidar data and field measurements. However, few studies have evaluated the impacts of GPS accuracy on biomass estimation using lidar data.

The availability of both differential GPS coordinates and recreational GPS coordinates for the 74 plots in our study site made it

possible to directly assess the impacts of plot coordinate accuracy on biomass estimation. Table 6 reported the fitting statistics of the multiplicative and mixed-effects models based on RGPS plot coordinates. The use of recreational instead of differential GPS in our study site resulted in a decrease of R^2 by 0.10–0.13 and an increase of RMSE by about 21–31%. This degradation in performance due to GPS accuracy will likely vary depending on the site-specific conditions (e.g., canopy structure, spatial heterogeneity, and topography). Nevertheless, our results emphasize the value of differential GPS to locate field plots for vegetation measurements.

6. Conclusions

Lidar is a state-of-the-art technology for mapping biomass, which relies on the fundamental relationship between biomass and canopy structure metrics such as height. Motivated by the biomass dependence on vegetation types, this study uses an innovative method, mixed-effects models, to integrate airborne lidar and vegetation types derived from aerial photographs to map biomass over the Sagehen Creek Experimental Forest in the Sierra Nevada of California. It was found that mixed-effects models can effectively deal with the small samples associated with each vegetation type and can improve biomass estimation compared to the use of lidar data alone in multiplicative models.

The vegetation of our study site was classified based on two different systems: SAF and NVC alliance-level classes. We found that, despite its emphasis on dominant species, the SAF cover types are as powerful as the NVC alliance-level vegetation types in the mixed-effects modeling of biomass. This result suggests that vegetation classification for carbon assessment could focus on dominant species given the strong relationship between forest stand biomass and dominant species. The vegetation types of this study were visually interpreted from aerial photographs. For many places, especially those in developing countries, updated aerial photographs are not always available. Due to the increasing accessibility of high spatial resolution satellite imagery such as Worldview-2, further research should be done in the future to investigate the use of high spatial resolution satellite imagery, digital image classification, and airborne lidar data for biomass and carbon mapping.

References

Asner, G. P., Hughes, R. F., Varga, T. A., Knapp, D. E., & Kennedy-Bowdoin, T. (2009). Environmental and biotic controls over aboveground biomass throughout a tropical rain forest. *Ecosystems*, *12*, 261–278.

Avery, T. E. (1978). *Forester's guide to aerial photo interpretation.* Agriculture Handbook 308. : U.S. Department of Agriculture, Forest Service.

Banskota, A., Wynne, R. H., Johnson, P., & Emessiene, B. (2011). Synergistic use of very high-frequency radar and discrete-return lidar for estimating biomass in temperate hardwood and mixed forests. *Annals of Forest Science*, *68*(2), 347–356.

Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L., & Kimes, D. S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Quebec. *Remote Sensing of Environment*, *112*, 3876–3890.

Existing vegetation classification and mapping technical guide. Brohman, R., & Bryant, L. (Eds.). (2005). *Gen. Tech. Rep. WO–67.* Washington DC: U.S. Department of Agriculture Forest Service, Ecosystem Management Coordination Staff.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd edition). New York: Springer-Verlag Press.

Chave, J., Muller-Landau, H. C., Baker, T. R., Easdale, T. A., Ter Steege, H., & Webb, C. O. (2006). Regional and phylogenetic variation of wood density across 2456 neotropical tree species. *Ecological Applications*, *16*, 2356–2367.

Chen, Q. (2007). Airborne lidar data processing and information extraction. *Photogrammetric Engineering and Remote Sensing*, *73*(2), 109–112.

Chen, Q. (2010). Retrieving canopy height of forests and woodlands over mountainous areas in the Pacific coast region using satellite laser altimetry. *Remote Sensing of Environment*, *114*, 1610–1627.

Chen, Q., Gong, P., Baldocchi, D. D., & Xie, G. (2007). Filtering airborne laser scanning data with morphological methods. *Photogrammetric Engineering and Remote Sensing, 73* (2), 175–185.

Colwell, R. N. (1946). The estimation of ground conditions from aerial photographic interpretation of vegetation types. *Photogmmmetric Engineering, 12*(2), 151–161.

Davies, K. W., Petersen, S. L., Johnson, D. D., Davis, D. B., Madsen, M. D., & Zvirzdin, D. L. (2010). Estimating juniper cover from National Agriculture Imagery Program (NAIP) imagery and evaluating relationships between potential cover and environmental variables. *Rangeland Ecology & Management, 63*(6), 630–637.

De Jong, B., Anaya, C., Masera, O., Olguin, M., Paz, F., Etchevers, J., Martinez, R. D., Guerrero, G., & Balbontin, C. (2010). Greenhouse gas emissions between 1993 and 2002 from land-use change and forestry in Mexico. *Forest Ecology and Management, 260*(10), 1689–1701.

De Santis, A., Asner, G. P., Vaughan, P. J., & Knapp, D. E. (2010). Mapping burn severity and burning efficiency in California using simulation models and Landsat imagery. *Remote Sensing of Environment, 114*(7), 1535–1545.

Dominy, N. J., & Duncan, B. (2001). GPS and GIS methods in an African rain forest: Applications to tropical ecology and conservation. *Conservation Ecology, 5*(2), 537–549.

Drake, J. B., Knox, R. G., Dubayah, R. O., Clark, D. B., Condit, R., Blair, J. B., & Hofton, M. (2003). Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: Factors affecting the generality of relationships. *Global Ecology & Biogeography, 12*, 147–159.

Dubayah, R. O., Sheldon, S. L., Clark, D. B., Hofton, M. A., Blair, J. B., Hurtt, G. C., & Chazdon, R. L. (2010). Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *Journal of Geophysical Research, 115*, G00E09, doi:10.1029/2009JG000933.

Eyre, F. H. (1980). Forest cover types of the United States and Canada. Washington, D.C: Society of American Foresters (SAF) 148 p.

Fensham, R. J., & Fairfax, R. J. (2002). Aerial photography for assessing vegetation change: A review of applications and the relevance of findings for Australian vegetation history. *Australian Journal of Botany, 50*(4), 415–429.

FGDC (Federal Geographic Data Committee) (2008). *National Vegetation Classification Standard*, Version 2 FGDC-STD-005-2008 (version 2). *Vegetation Subcommittee, Federal Geographic Data Committee, FGDC Secretariat*. Reston, Virginia, USA: U.S. Geological Survey.

Gonzalez, P., Asner, G. P., Battles, J. J., Lefsky, M. A., Waring, K. M., & Palace, M. (2010). Forest carbon densities and uncertainties from LiDAR, QuickBird and field measurements in California. *Remote Sensing of Environment, 114*(7), 1561–1575.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research, 26*(3), 499–510.

Grossman, D. H., Faber-Langendoen, D., Weakley, A. S., Anderson, M., Bourgeron, P., Crawford, R., Goodin, K., Landaal, S., Metzler, K., Patterson, K. D., Pyne, M., Reid, M., & Sneddon, L. (1998). International classification of ecological communities: Terrestrial vegetation of the United States. Volume I. *The National Vegetation Classification System: development, status, and applications*. Arlington, Virginia, USA: The Nature Conservancy.

Guo, Z., Hong, C., & Sun, G. (2010). Estimating forest aboveground biomass using HJ-1 Satellite CCD and ICESat GLAS waveform data. *Science in China Series D: Earth Sciences, 53*(Supplement:1), 16–25.

Hall, S. A., Burke, I. C., Box, D. O., Kaufmann, M. R., & Stoker, J. M. (2005). Estimating stand structure using discrete-return lidar: An example from low density, fire prone ponderosa pine forests. *Forest Ecology and Management, 208*, 189–209.

Halpern, C., & Means, J. (2004). *Pacific Northwest Plant Biomass Component Equation Library*. Long-Term Ecological Research. Corvallis, OR: Forest Science Data Bank.

Higinbotham, C. B., Alber, M., & Chalmers, A. G. (2004). Analysis of tidal marsh vegetation patterns in two Georgia estuaries using aerial photography and GIS. *Estuaries, 27*(4), 670–683.

Hyde, P., Dubayah, R., Walker, W., Blair, B., HOfton, M., & Hunsaker, C. (2006). Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sensing of Environment, 102*, 63–73.

Jenkins, J. C., Chojnacky, D. C., Heath, L. S., & Birdsey, R. A. (2004). Comprehensive database of diameter-based biomass regressions for North American tree species. *General technical report NE-319*. Newtown Square, PA: U.S. Department of Agriculture, Forest Service.

Jennings, M. D., Faber-Langendoen, D., Loucks, O. L., Peet, R. K., & Roberts, D. (2009). Standards for associations and alliances of the U.S. National Vegetation Classification. *Ecological Monographs, 79*, 173–199.

Kim, Y., Yang, Z. Q., Cohen, W. B., Pflugmacher, D., Lauver, C. L., & Vankat, J. L. (2009). Distinguishing between live and dead standing tree biomass on the North Rim of Grand Canyon National Park, USA using small-footprint lidar data. *Remote Sensing of Environment, 113*, 2499–2510.

Lefsky, M. A., Harding, D., Cohen, W. B., Parker, G., & Shugart, H. H. (1999). Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of Environment, 67*, 83–98.

Lillesand, T., Kiefer, R., & Chipman, J. (2008). *Remote sensing and image interpretation* (6th edition). NY: John Wiley & Sons.

Lim, K., Treitz, P., Baldwin, K., Morrison, I., & Green, J. (2003). Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing, 29*, 658–678.

Loudermilk, E. L., Hiers, J. K., O'Brien, J. J., Mitchell, R. J., Singhania, A., Fernandez, J. C., Cropper, W. P., & Slatton, K. C. (2009). Ground-based LIDAR: A novel approach to quantify fine-scale fuelbed characteristics. *International Journal of Wildland Fire, 18*, 676–685.

MacLean, G. A., & Krabill, W. B. (1986). Gross-merchantable timber volume estimation using an airborne LiDAR system. *Canadian Journal of Remote Sensing, 12*, 7–18.

Malmsheimer, R. W., Bowyer, J. L., Fried, J. S., Gee, E., Izlar, R. L., Miner, R. A., Munn, I. A., Oneil, E., & Stewart, W. C. (2011). Managing forests because carbon matters: Integrating energy, products, and land management policy. *Journal of Forestry, 109*, S7–S48.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. A., Cohen, W. B., Harmon, M. E., & McKee, W. A. (1999). Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment, 67*, 298–308.

Means, J. E., Hansen, H. A., Koerper, G. J., Alaback, P. B., & Klopsch, M. W. (1994). Software for computing plant biomass — BIOPAK users guide. *Gen. Tech. Rep. PNW-GTR-340*. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Meng, Q., Cieszewski, C. J., Madden, M., & Borders, B. (2007). A linear mixed effects model of biomass and volume of trees using Landsat ETM+ images. *Forest Ecology and Management, 244*(1–3), 93–101.

Meyer, K. E., & Laudenslayer, W. F. (1988). *A guide to wildlife habitats of California*. Sacramento: California Department of Fish and Game.

Miura, N., & Jones, S. D. (2010). Characterizing forest ecological structure using pulse types and heights of airborne laser scanning. *Remote Sensing of Environment, 114*, 1069–1076.

Morgan, J. L., Gergel, S. E., & Coops, N. C. (2010). Aerial photography: A rapidly evolving tool for ecological management. *BioScience, 60*(1), 47–59.

Morris, D. L., Western, D., & Maitumo, D. (2009). Pastoralist's livestock and settlements influence game bird diversity and abundance in a savanna ecosystem of southern Kenya. *African Journal of Ecology, 47*(1), 48–55.

Næsset, E., & Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment, 112*, 3079–3090.

Nelson, R., Boudreau, J., Gregoire, T. G., Margolis, H., Næsset, E., Gobakken, T., & Stahl, G. (2009). Estimating Quebec provincial forest resources using ICESat/GLAS. *Canadian Journal of Forest Research, 39*, 862–881.

Nelson, R., Krabill, W., & Tonelli, J. (1988). Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment, 24*, 247–267.

Niklas, K. J. (1995). Size-dependent allometry of tree height, diameter and trunk-taper. *Annals of Botany, 75*, 217–227.

Ni-Meister, W., Lee, S. Y., Strahler, A. H., Woodcock, C. E., Schaaf, C., Yao, T. A., Ranson, K. J., Sun, G. Q., & Blair, J. B. (2010). Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. *Journal of Geophysical Research-Biogeosciences, 115*, doi:10.1029/2009JG000936 Article No.: G00E11.

Salas, C., Ene, L., Gregoire, T. G., Næsset, E., & Gobakken, T. (2010). Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models. *Remote Sensing of Environment, 114*, 1277–1285.

Stahl, G., Holm, S., Gregoire, T. G., Gobakken, T., Næsset, E., & Nelson, R. (2011). Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research, 41*(1), 96–107.

Ursino, N. (2007). Modeling banded vegetation patterns in semiarid regions: Interdependence between biomass growth rate and relevant hydrological processes. *Water Resources Research, 43*(4), W04412.

USDA Forest Service (1981). CALVEG: A classification of California vegetation. San Francisco CA: Pacific Southwest Region, Regional Ecology Group 168 pp.

Wagner, W., Roncat, A., Melzer, T., & Ullrich, A. (2007). Waveform analysis techniques in airborne laser scanning. *IAPRS, 39*(Part 3/W52), 413–418.

Waring, R. H., Way, J. B., Hunt, E. R., Morrissey, L., Ranson, K. J., Weishampel, J. F., Oren, R., & Franklin, S. E. (1995). Imaging radar for ecosystem studies. *Bioscience, 45*, 715–723.

Zianis, D., & Mencuccini, M. (2004). On simplifying allometric analyses of forest biomass. *Forest Ecology and Management, 187*(2–3), 311–332.

Chapter 4

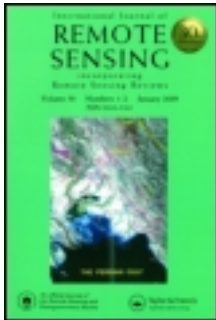# Discrimination of vegetation types in alpine sites with ALOS PALSAR, RADARSAT-2, and lidar-derived information

Discrimination of vegetation types in alpine sites with ALOS PALSAR-, RADARSAT-2-, and lidar-derived information

Gaia Vaglio Laurin [a] [d] , Fabio Del Frate [a] , Luca Pasolli [b] , Claudia
Notarnicola [b] , Leila Guerriero [a] & Riccardo Valentini [c] [d]

[a] Department of Civil Engineering and Computer Science
Engineering , Tor Vergata University of Rome 00133 , Italy

[b] EURAC Research Institute for Applied Remote Sensing , Viale
Druso , 1 I-39100 , Bolzano , Italy

[c] Department of Forest Resources and Environment , University of
Tuscia , Viterbo , I-01100 , Italy

[d] CMCC – Centro Euro-Mediterraneo per i Cambiamenti Climatici
(Euro- Mediterranean Center for Climate Change), via Augusto
Imperatore , Lecce , 73100 , Italy
Published online: 25 Jun 2013.

PLEASE SCROLL DOWN FOR ARTICLE

instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Taylor & Francis
Taylor & Francis Group

Discrimination of vegetation types in alpine sites with ALOS PALSAR-, RADARSAT-2-, and lidar-derived information

Gaia Vaglio Laurin[a,d]*, Fabio Del Frate[a], Luca Pasolli[b], Claudia Notarnicola[b], Leila Guerriero[a], and Riccardo Valentini[c,d]

[a]Department of Civil Engineering and Computer Science Engineering, Tor Vergata University of Rome 00133, Italy; [b]EURAC Research Institute for Applied Remote Sensing, Viale Druso, 1 I-39100 Bolzano, Italy; [c]Department of Forest Resources and Environment, University of Tuscia, Viterbo I-01100, Italy; [d]CMCC – Centro Euro-Mediterraneo per i Cambiamenti Climatici (Euro-Mediterranean Center for Climate Change), via Augusto Imperatore, Lecce 73100, Italy

Natural vegetation monitoring in the alpine mountain range is a priority in the European Union in view of climate change effects. Many potential monitoring tools, based on advanced remote sensing sensors, are still not fully integrated in operational activities, such as those exploiting very high-resolution synthetic aperture radar (SAR) or light detection and ranging (lidar) data. Their testing is important for possible incorporation in routine monitoring and to increase the quantity and quality of environmental information. In this study the potential of ALOS PALSAR and RADARSAT-2 SAR scenes' synergic use for discrimination of different vegetation types was tested in an alpine heterogeneous and fragmented landscape. The integration of a lidar-based canopy height model (CHM) with SAR data was also tested. A SPOT image was used as a benchmark to evaluate the results obtained with different input data. Discrimination of vegetation types was performed with maximum likelihood classification and neural networks. Six tested data combinations obtained more than 85% overall accuracy, and the most complex input which integrates the two SARs with lidar CHM outperformed the result based on SPOT. Neural network algorithms provided the best results. This study highlights the advantages of integrating SAR sensors with lidar CHM for vegetation monitoring in a changing environment.

1. Introduction

Mountain areas are vulnerable, heterogeneous, and dynamic regions continuously changed by human land use, hazard phenomena, and increased socio-economic competition. Monitoring the evolution of mountain areas at various scales in space and time is an urgent issue that can be addressed by remote sensing technology (Schneiderbauer, Zebisch, and Steurer 2007) responding to the European Union (EU) climate change monitoring needs. Endemic mountain plant species are threatened by the upwards migration of more competitive sub-alpine shrubs and tree species, leading to considerable loss of endemic species in mountain regions (Voigt et al. 2004) and severe loss in the economic value of EU forest land (Hanewinkel et al. 2013). Proper management action relies on natural vegetation monitoring, and should at least be sufficiently detailed to detect changes in the extent of main

---

*Corresponding author. Email: laurin@disp.uniroma2.it

vegetation types. Advanced remote sensing sensors are still not fully integrated in operational activities, such as those exploiting very high-resolution synthetic aperture radar (SAR), hyperspectral, or light detection and ranging (lidar) data. Their testing is important for possible incorporation in routine monitoring of forested areas and to increase the quantity and quality of environmental information.

SAR sensors can be of meaningful support in vegetation studies, being well suited to provide cover information (Rahman and Sumantyo 2010; Simard, Saatchi, and De Grandi 2000), distinguish forest from non-forest (Lehmann et al. 2012; Longepe et al 2011), or monitor regrowth (Minchella et al. 2009). SAR is an all-weather system (Lehmann et al. 2012; Lu, Batistella, and Moran 2007), recently made available at very high spatial resolution with increased polarimetric and revisiting capabilities.

In recent land cover applications, RADARSAT-2 (5.04 GHz centre frequency, 5.6 cm wavelength) has been used in China to identify forest area and deforestation (Zhang et al. 2012), and in the Brazilian Pantanal, in combination with Advanced Land Observation Satellite (ALOS) Phased Arrayed L-band SAR (PALSAR; 1.27 GHz centre frequency, 23 cm wavelength), to map land cover (Evans et al. 2010). SAR systems can successfully compete with optical ones for forest mapping: in a comparison between the radar vegetation index (RVI) extracted from ALOS PALSAR full polarimetric data and Landsat TM data, Ling et al. (2009) found that RVI is much better suited for this kind of application.

It is also recognized that SAR channels, especially at C and L bands and in single temporal acquisition, have a limited ability to distinguish different woody vegetation with high biomass content, such as forest types or dense shrubs (Kurvonen and Hallikainen 1999; Lee et al. 2005; Touzi, Landry, and Charbonneau 2004; Yatabe and Leckie 1996). Thus there is great interest in understanding the effectiveness of advanced SAR systems in woody vegetation type discrimination, especially in those areas requiring operational and frequent monitoring.

Besides classification purposes, SAR data can be exploited for the estimate of forest above-ground biomass (AGB), which is fundamental information for climate change modelling activities and mitigation policies. Indeed, the dependence of SAR backscattering on AGB has been broadly documented (Dobson et al. 1992; LeToan et al. 1992; Kasischke, Christensen, and Bourgeau-Chavez 1995), especially at longer wavelengths such as L and P bands (Ferrazzoli and Guerriero 1995), and several retrieval methods have been proposed (Rignot et al. 1994; Del Frate and Solimini 2004; Englhart, Keuck, and Siegert 2012) which take advantage of this peculiarity of microwave remote sensing.

Several studies have illustrated the usefulness of SAR texture (the spatial arrangement of the intensity or backscattering of pixels) in improving classification accuracy. Most exploited texture from both optical and SAR data (Lu, Batistella, and Moran 2007; Wijaya, Marpu, and Gloaguen 2008; Vaglio Laurin et al. 2013), while others only based their analysis on SAR and derived texture: Dekker (2003) and Del Frate, Pacifici, and Solimini (2008) experimented with SAR texture data in an urban area; in Gabon, Simard, Saatchi, and De Grandi (2000) used the texture from the Japanese Earth Resources Satellite (JERS-1) SAR data to refine the classification of the flooded forest class; Li et al. (2012) used ALOS PALSAR and RADARSAT-2 textures for the classification of a tropical moist region, proving that by using data combination, results are considerably more accurate than when data sets are considered individually.

When using SAR operating at different frequencies, such as at L and C bands, the backscattered energy can return from different portions of the vegetation along the vertical

profile, due to the varying penetration capabilities of the bands and the scattering mechanisms peculiar to foliar strata or branches. Information on the textural arrangement of these vegetation portions can represent an additional value in classification and discrimination of natural vegetation.

Vegetation and forestry analyses have enormously benefitted from the advent of lidar technology over the last two decades: penetrating the canopy, this system can provide information on the vertical canopy structure. The lidar backscattered signal can be used as punctual height information or it can be interpolated to generate digital terrain models (DTMs) from ground returns, and digital surface models (DSMs) from above-ground returns. Subtraction of DTM from DSM corresponds to the canopy height model (CHM), which provides a measure of the height of the upper canopy for each pixel of vegetation in the surveyed area (Kraus and Pfeifer 1998).

Lidar data have been used in the alpine area to gather structural information on forests (Chauve et al. 2008; Hollaus et al. 2006; Jochem et al. 2011). In other environments, the effort of integrating lidar with other data types for classification of vegetation types has been proved successful. For instance, Dowling and Accad (2003) joined the height information generated by lidar with digital video to map vegetation types and height classes in a riparian zone in Australia; Bork and Su (2007) compared the classifications obtained by lidar, multispectral, and the two combined data types in Canadian rangelands, finding that data integration resulted in accuracy improvements of 16% to 20%. Similarly, Geerling et al. (2007) found that the fused Compact Airborne Spectrographic Imager (CASI) spectral and lidar information produced better results than single dataset use in a natural floodplain classification in The Netherlands; while Onojeghuo and Blackburn (2011) combined hyperspectral imagery and textural information with lidar CHM for the effective mapping of reed bed habitats in the UK.

Lidar data most commonly used for forestry and mapping applications are not produced by dedicated flights: technicians often exploit the raster CHM available at either a low cost or even for free from surveys carried out for purposes other than vegetation applications (Corona et al. 2012). The increased availability of these Lidar-derived products, released by local administrations due to widespread lidar use in topographic mapping, can be very valuable in the support of vegetation monitoring efforts. These data sets could be integrated in routine monitoring activities if their use is proven to be beneficial in increasing the ability to discriminate and map natural vegetation.

The innovation of this study is twofold. Initially, it is to explore the potential of ALOS PALSAR and RADARSAT-2 SAR scenes' synergic use for discrimination of different vegetation types – with a focus on woody vegetation – in an untested alpine heterogeneous and fragmented landscape, thus assessing the effectiveness of advanced SAR systems to provide detailed mapping in natural environments. Four vegetation types were considered in the study sites, which are covered by two RADARSAT-2 Standard Quad-Pol (SQP) and one ALOS PALSAR Fine Beam Dual (FBD) polarimetric scenes. The discrimination ability of these two SAR types was explored singularly and simultaneously, also using textural variables. A second innovative feature is the integration of a lidar-based CHM with SAR data to evaluate the advantages offered by the addition of this frequently available data type. Vegetation types were discriminated by means of neural networks (NNs), comparing the results with those obtained by a maximum likelihood classification (MLC). The performances obtained by a SPOT 5 optical sensor were used as a benchmark.

2. Materials and methods

2.1. Study area

An area of approximately 30 × 30 km in the Autonomous Province of Bolzano (Bozen), South Tyrol, in Northern Italy (Figure 1), with elevations from 224 to 3343 m, was selected for this study according to the availability of ground truth in its northeastern (NW limit 700000 E, 5174000 N) and southwestern ranges (SW limit 670500 E, 5142300 N).

In the mountain ranges, the floristic composition is shaped according to altitudinal planes along an elevation gradient: the planes' extent is flexible and changes according to many variables (i.e. morphology, orientation, soil type, orography, human influence, etc.). A rough division identifies the *basal* (<600 m), the *sub-mountain* (400–1200 m), the *mountain* (800–2200 m), the *alpine* (2000–3000 m), and the *nival* planes (2600–3000 m).

In the study sites several vegetation associations, based on floristic criteria, are found in the different planes. The ground truth generation allowed differentiation of four main
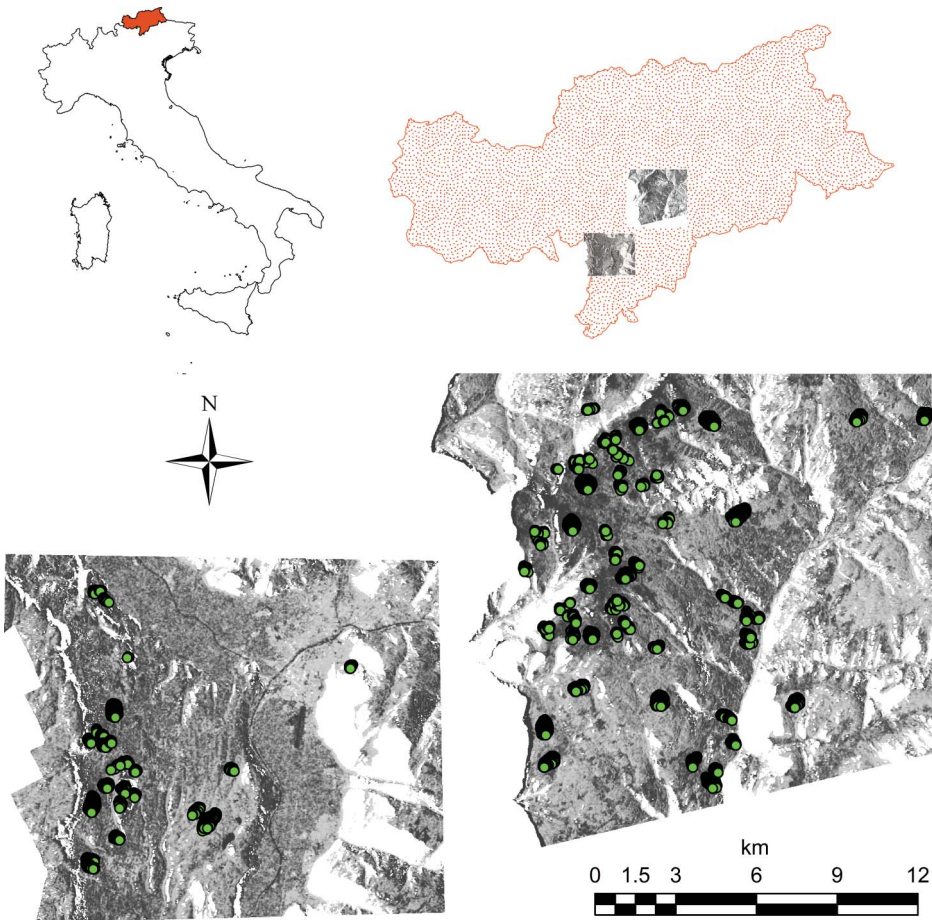
Figure 1. The study area, covering regions where PALSAR and RADARSAT-2 data are overlapped, is illustrated by the two grey-level main images (RADARSAT-2 HV channels in both scenes, in white areas masked for distortion) with respect to South Tyrol (in the upper right) and sites of ground truth availability (green spots over the images). The scale bar refers to SAR images. The location of South Tyrol with respect to Italy is in the upper left in orange.

vegetation types according to structural criteria, which are those detected by remote sensing: needle-leaved forest, broadleaved forest, shrubs and dwarf pines, and grasslands.

According to information provided by the SEA (Scuola Educazione Ambiente) Dobbiaco centre of Alta Pusteria (accessed December 21, 2012: http://www.sea-dobbiaco. it/Default.aspx), the vegetation of the area can be characterized as follows. Broadleaved forest (Figure 2(*a*)) is typical of sub-mountain and basal planes, the former preferred by mesophilous species (i.e. *Fagus sylvatica*) and the latter by thermophilous or thermo-mesophilous examples (i.e. *Quercus*, *Carpinus*, *Castanea* genera). Needle-leaved forest (Figure 2(*b*)) is found in the upper part of the mountain plane, with trees from different genera such as *Picea*, *Larix*, and *Pinus*. Historically, human influence tended to expand grassland for livestock grazing, lowering the natural upper limit of these forests by 200–300 m. Needle-leaved species can also be found in the cooler areas of the sub-mountain plane mixed with broadleaved ones. Grasslands (Figure 2(*c*)) are frequently composed of tens of genera and are found in both the alpine plane (intermixed with short woody vegetation) and the lower planes, in plateaux with flat terrain and deeper soil, where they are grazed by livestock. Few very resistant species (e.g. *Saxifraga caesia*) with highly patchy distribution can also be found in the the nival plane, which is dominated by lichens and mosses.

Shrubs and dwarf pines are (Figure 2(*d*)) characteristic of an alpine plane – due to high winds, the vegetation has horizontal growth and its height is reduced, usually below 1 m.

Figure 2.   Vegetation types in the study area, from orthophotographs: (*a*) broadleaved forest, (*b*) needle-leaved forest, (*c*) grasslands, (*d*) shrubs, and dwarf pines. All images here are illustrated at 1:2000 m scale.

Pinus mugo formations are prevalent in the study sites, as are *Rhododendron* and *Carex* genera; at lower ranges, *Alnus viridis* and *Salix* spp. can be found.

2.2. *Remote sensing data*

One ALOS PALSAR Fine Beam Double Polarization (HH, HV) scene was acquired on 27 June 2010, and two RADARSAT-2 Standard Quad Polarization (HH, HV, VH, VV) scenes on 11 and 28 September 2010, with incidence angles of 28.8°, 32.1°, and 34.3°, respectively. Original images were provided in single-look complex (SLC) format with pixel sizes of 4.93 m and 17.48 m in azimuth and ground range directions, respectively. All SAR scenes were multi-looked, radiometrically calibrated, and geocoded with the help of a high-geometrical resolution (2.5 m) digital elevation model and filtered with a Frost filter with a 5×5 window size for further speckle reduction. The final resolution was set to 20×20 m.

Masks over areas of distortion from layover, shadowing, and foreshortening effects were generated and used to exclude some areas during ground truth selection. All preprocessing was carried out with SARscape software (http://www.sarmap.ch/). To compensate geolocation inaccuracies, all scenes were co-registered with a SPOT 5 image dated 5 September 2006 (4 VIS-NIR bands), with a RMSE of 0.86 (of pixel size) for both RADARSAT-2 scenes, and of 0.69 for ALOS PALSAR.

The SPOT 5 image was orthocorrected, atmospherically corrected with the Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm in ENVI version 4.7 (Exelis Visual Information Solutions, Boulder, Colorado), and resampled to 20 m with bilinear interpolation, for use in tests as a benchmark for comparison with other results.

SAR amplitude values were extracted for pixels outside masked areas and used for further analysis. The values of two cross-polarized channels from RADARSAT-2 scenes were averaged, thus obtaining a single channel hereafter denoted as HV. Similarly, CHM heights and SPOT reflectance values were extracted for ground truth pixels.

A lidar survey, requested by the South Tyrol government, was carried out between October 2006 and December 2007 over the Autonomous Province of Bolzano using a first–last return Optech ALTM3100 sensor operating at a frequency of 100.000 kHz (25° of maximum field of view; minimum 12 points per 2.5×2.5 m area; 30 cm error in standard deviation for height values; and 90% of pulses penetrating the canopy in broadleaved forests). The DTM and DSM data sets generated from this lidar survey were provided for the present study by the Autonomous Province of Bolzano – Office of Territorial Coordination, and were then algebraically subtracted to generate a CHM. The CHM, originally provided at 2.5 m resolution, was resampled with bilinear interpolation to 20 m. This step was needed to perform the integration of different data types, harmonizing their spatial resolution to the coarser one. SAR and lidar data geolocation accuracies were compared in correspondence of visible features (roads, bridges, large buildings, etc.) in the scenes, indicating no need for further co-registration of these different data sets.

Eight textures based on the Grey-Level Co-Occurrence Matrix (GLCM) (Haralick, Shanmugan, and Dinstein 1973) were generated with ENVI version 4.7 for each of the polarizations of ALOS PALSAR and RADARSAT-2 scenes: mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation. We used 64 greyscale quantization levels, 1 pixel shift, and different window sizes: 3×3, 5×5, 7×7, 9×9, and 15×15. A wrapper-based approach for texture variable selection was adopted to reduce the number of features entering in classification (see Kohavi and John 1997 for a description of

feature selection methods). With this approach, the feedback from a classifier was used to guide the search, in this case Kappa coefficient and overall accuracies obtained with MLC. We based the selection of features only on MLC results, since classification tests including those textures were satisfactory for both MLC and NN. This allowed us to compare the two classifiers considering the same inputs. Furthermore, the selection based on MLC was automated and successfully applied during previous research (Vaglio Laurin et al. 2013): this adds one texture at a time to SAR data, repeating the tests with increasing window size; one variable is selected for each polarization, retaining the one that when added to the original data produces the best classification. The selection process ends when any further increase in window size does not correspond to accuracy improvement. When applied to this data set, the procedure resulted in the choice of a 9×9 window size for both SAR types, with RADASAT-2 mean textures selected for HH and HV polarizations and homogeneity for VV polarization, and ALOS PALSAR mean textures selected for both HH and HV channels.

2.3. Ground truth data

An automated sampling method, performed over an orthorectified and georeferenced spring 2006 SPOT 5 image, with the support of field and auxiliary data, was used to generate an original ground truth data set for this area; the procedure is fully described in Notarnicola et al. (2009). Part of these ground truth samples were located in areas of strong geometric distortion of the SAR data used for this research, and some classes missed sufficient sampling. Therefore, the masks of distortion were first applied to select the exploitable ground truth; then a visual interpretation procedure was used to increase the number of pixels and sites, and thus the representativeness of all of the vegetation classes of interest.

The procedure used the Corine Land Cover 2000 (CLC2000) database and ortho-corrected aerial photographs at 0.5 m spatial resolution, collected during summer 2006 (provided by the Autonomous Province of Bolzano – Office of Territorial Coordination) as information layers additional to the original ground truth data. Areas classified in the CLC2000 vector layer as corresponding to our vegetation classes were selected and identified over the orthophotographs, on which the original ground truth areas were also visualized. This allowed us to enlarge the original areas, thus gathering enough ground truth for each vegetation type, with a minimum of 21 regions of interest (ROIs) for the less diffused grasslands type. Pixels within these ROIs were randomly selected in order to obtain the same number of pixels per vegetation type (i.e. 1450 for each type). Although the ROIs were collected as evenly distributed as possible over the entire study area, a certain degree of spatial correlation can occur due to the distribution of vegetation according to altitudinal plans. We confirmed the absence of significant differences in the different ground truth used for training, validation, and testing with statistical tests (not reported here).

2.4. Classification algorithms

Tests were conducted by both MLC and NN. The network architecture, defined after a trial-and-error approach, was composed of two hidden layers, each with a doubled number of neurons with respect to the number of inputs in tests, and a scale conjugate gradient (Moller 1993) training algorithm.

The ROIs were randomly divided into three sets: training (60%, 870 pixels), validation (15%, 232 pixels) exclusively for NN usage, and test (25%, 348 pixels) for both NN and MLC.

The division of ground truth into three sets is required in NN to evaluate the network generalization capabilities and perform cross-validation, avoiding overfitting and falling into local minima of error (Bishop 1995). Rather, the MLC only employs training and testing sets.

The MLC and NN algorithms were chosen as being very widespread and suited to solving complex tasks. The MLC used to select texture features is a robust algorithm available in all image analysis software, and is still used in SAR-based forest mapping and classification (Lehmann et al. 2012; Rahman and Sumantyo 2010). It can serve as a benchmark for results obtained with more complex algorithms such as NN, which belong to the machine learning-algorithm family, thus being able to learn from empirical data provided as input with a non-parametric approach. NN is broadly used in remote sensing research (Del Frate et al. 2003; Lu and Weng 2007), and we used it to assess the extent of improvement in vegetation discrimination by machine learning, with respect to MLC.

2.5. *Data analysis*

We conducted different tests with the aim of assessing the best data combination for the discrimination of the considered vegetation types, starting with the testing of SAR data singularly and in multiples, then progressively adding the selected texture features and CHM, as in Table 1.

This progressive approach showed how the contribution of different data sets influenced the results, which are expressed by the overall accuracy (OA) and Kappa (K) coefficient of agreement (Cohen 1960). The overall accuracy is the percentage of correctly classified pixels; the Kappa coefficient represents the difference between the actual agreement and chance agreement.

The statistical differences in MLC and NN results were determined at the 95% level of significance using the Z-test (Congalton and Green 1999), which is performed for a pairwise comparison of the proposed methods and takes into account the ratio between the difference value of two Kappa coefficients and the difference in their respective variance.

Table 1.   List of data sets used for tests carried out with MLC and NN.

Inputs	Description
(1) RS2	RADARSAT-2 (HH, VV, averaged HV/VH polarizations)
(2) PSR	ALOS PALSAR, with HH and HV polarizations
(3) RS2+PSR	Combination of (1) and (2)
(4) RS2+TX	As in (1) with the addition of selected RADARSAT-2 texture features
(5) PSR+TX	As in (2) with the addition of PALSAR selected texture features
(6) RS2+PSR+2TX	As in (3) with the addition of two PALSAR and RADARSAT-2 selected texture features
(7) RS2+CHM	As in (1) with the addition of CHM
(8) PSR+CHM	As in (2) with the addition of CHM
(9) RS2+PSR+CHM	As in (3) with the addition of CHM
(10) RS2+TX+CHM	As in (4) with the addition of CHM
(11) PSR+TX+CHM	As in (5) with the addition of CHM
(12) RS2+PSR+2TX+CHM	As in (6) with the addition of CHM
(13) SPOT	SPOT 5, 4 VIS-NIR bands

3. Results

3.1. Separability analysis

The Jeffries–Matusita (J–M) separability measurement (Richards 1999) was computed for training ROIs for all possible pairs of vegetation types (Figure 3). The value of the J–M measurement ranges from 0 to 2.0 and indicates how well the selected ROI pairs are statistically separated; values close to 1.9 indicate that the ROI pairs have good separability.

The good separability of 'needle-leaved forest vs. grassland' pair in all tests, except in the case of test 1 based on RADASAT-2 channels only, is evident from Figure 1. Similarly, it is possible to observe a good separability for the pair 'shrubs and dwarf pines vs. grassland' in all tests featuring PALSAR data; RADARSAT-2, even with the addition of texture features, is not able to differentiate between these two vegetation types, and the information from lidar cannot help either. The behaviour of the two pairs 'broadleaved vs. shrubs and dwarf pines' and 'needle-leaved vs. shrubs and dwarf pines' is quite similar: the separability largely increases with the addition of height information. Only in the case of test 13 is there a marked difference between the separability of these two pairs: the optical inputs alone are not able to differentiate the 'needle-leaved vs. shrubs and dwarf pines' pair. The 'broadleaved vs. grassland' pair has very good separability in all inputs, including RADARSAT-2 texture and where PALSAR channels and CHM are both present. The least separable pair is the 'broadleaved vs. needle-leaved', which has a good score – but does still not indicate full separability – only for SPOT inputs; nonetheless, it is important to be aware that for this pair the inputs including PALSAR texture features are those producing higher J–M values.

3.2. MLC results

Results obtained with MLC are reported in Table 2. The OA from SAR channels inputs, singularly or combined, show the lowest OA values among the tests (always <70%), evidencing a low capability to discriminate the four different forest/vegetation types, with lowest values obtained for shrubs and dwarf pines and broadleaved vegetation.
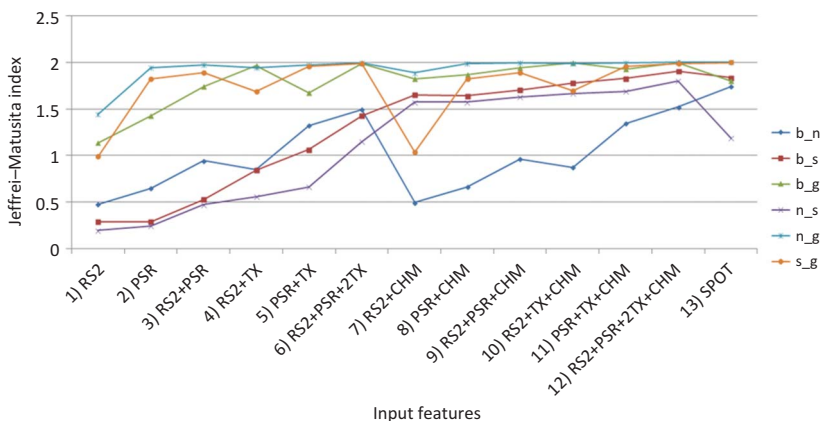

Figure 3. Jeffrei–Matusita index for inputs as in Table 1, and vegetation type pairs as follows: b_n = broadleaved vs. needle-leaved; b_s = broadleaved vs. shrubs and dwarf pines; b_g = broadleaved vs. grassland; n_s = needle-leaved vs. shrubs and dwarf pines; n_g = needle-leaved vs. grassland; s_g = shrubs and dwarf pines vs. grassland.

Table 2. Test result obtained with MLC.

Inputs	OA	K
(1) RS2	61.3	0.48
(2) PSR	67.6	0.56
(3) RS2+PSR	68.9	0.58
(4) RS2+TX	74.0	0.65
(5) PSR+TX	80.2	0.73
(6) RS2+PSR+2TX	85.4	0.80
(7) RS2+CHM	82.4	0.76
(8) PSR+CHM	84.1	0.78
(9) RS2+PSR+CHM	86.9	0.82
(10) RS2+TX+CHM	87.6	0.83
(11) PSR+TX+CHM	92.0	0.89
(12) RS2+PSR+2TX+CHM	93.6	0.91
(13) SPOT	92.8	0.90

Note: OA, overall accuracy; K, K coefficient.

RADARSAT-2 fails completely to separate shrubs and dwarf pines from the rest, but it identifies all other types with >70% accuracy. ALOS PALSAR, even with a higher OA with respect to RADARSAT-2 (67.6% vs 61.3%), shows broad confusion among types, with lowest accuracy for the broadleaved forest (excepting grasslands).

When the CHM is added as input to previous channels, the OA values strongly increase (from 16.5% to 21.1%), with OA always >80% and noticeable improvement in the identification of shrubs and dwarf pines. More specifically, RADARSAT-2 improvement is due only to the correct identification of shrubs and dwarf pines and grasslands, while other classes showed practically no change. For PALSAR, the largest improvement is found for shrubs and dwarf pines discrimination, but with a remarkable positive change occurring also for the two forest types. The combined use of the two SAR types and CHM produces higher OA but does not yield much improvement with respect to PALSAR and CHM inputs only (OA, respectively, 86.9% and 84.1%): the positive changes are in favour of the forest types.

When the texture features are added to channels, the improvement in OA is also remarkable (from 12.6% to 16.5%) in regard to all vegetation types. Unreported tests where the textures from the two SARs were added separately showed that ALOS PALSAR yielded higher accuracy. In any case, the best result (OA >85%) was obtained using textures at both frequencies, with shrubs and dwarf pines the most difficult types to discriminate (72.9%), and accuracy >83% for all other types.

When both texture features and CHM are used in input together with SAR channels, OA are the highest obtained (from 87.6% to 93.6%). The best result occurs using both SAR types, and it is higher than that obtained with SPOT optical bands, which are used as a benchmark (OA 92.8%); singularly, PALSAR outcomes RADARSAT-2 for accuracy, with confusion still slightly present in forest classes. The resulting user's and producer's accuracies are described in Table 3.

3.3. NN results

The results obtained with NN follow an identical pattern with respect to those obtained by MLC, with respect to both overall results (Table 4) and single-class accuracies: classes behave very similarly regardless of algorithm. All NN overall accuracies were higher than

Table 3. Producer's (PA) and user's (UA) accuracies obtained for the different inputs with MLC listed as in Table 1, and expressed in percentages.

	Test 1		Test 2		Test 3		Test 4		Test 5	
	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
BL	77.0	53.8	48.0	63.3	62.9	65.8	85.3	67.6	76.15	84.4
NL	70.7	55.4	65.5	66.7	79.0	59.0	74.1	67.2	66.67	81.4
SD	12.4	49.4	61.5	48.8	38.8	52.3	42.2	64.5	83.62	63.5
GS	85.3	81.8	95.7	96.0	95.1	98.8	94.5	96.5	94.54	98.2

	Test 6		Test 7		Test 8		Test 9		Test 10	
	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
BL	83.6	86.9	77.3	78.4	67.5	79.1	71.8	87.1	84.5	83.0
NL	86.5	76.0	70.7	87.2	74.4	78.0	81.6	81.6	76.1	92.7
SD	73.0	80.4	89.4	76.4	97.1	83.9	97.1	83.2	95.1	80.7
GS	98.6	99.4	92.5	89.4	97.4	94.2	97.4	96.6	97.4	96.6

	Test 11		Test 12		Test 13	
	PA	UA	PA	UA	PA	UA
BL	82.8	91.1	88.8	92.0	93.7	92.6
NL	89.4	90.7	89.9	92.0	91.7	87.6
SD	98.6	91.0	97.4	91.1	90.8	93.2
GS	98.0	95.8	98.6	99.7	95.1	98.2

Note: BL, broadleaved forest; NL, needle-leaved forest; SD, shrubs and dwarf pines; GS, grasslands.

Table 4. Test results obtained with NN.

Inputs	OA	Z
(1) RS2	65.9	2.57*
(2) PSR	70.3	1.52
(3) RS2+PSR	74.3	3.15*
(4) RS2+TX	79.0	3.14*
(5) PSR+TX	82.1	1.23
(6) RS2+PSR+2TX	88.1	2.13*
(7) RS2+CHM	84.5	1.43
(8) PSR+CHM	85.8	1.23
(9) RS2+PSR+CHM	89.8	2.32*
(10) RS2+TX+CHM	92.1	3.97*
(11) PSR+TX+CHM	94.3	2.22*
(12) RS2+PSR+2TX+CHM	97.7	5.20*
(13) SPOT	93.0	0.14

Notes: OA, overall accuracy; Z, scores from Z-test.
∗ denotes when the result, to be considered in absolute values, is significantly different from that obtained by MLC.

those of MLC, but not all differences were found to be significant. In fact, to be significantly different at the 95% confidence level, the absolute value of the Z score should be >1.96, and this happened for eight inputs out of 13.

4. Discussion and conclusions

First, this study evaluated the performance and integration of PALSAR and RADARSAT-2 for discriminating vegetation in the alpine mountain range, and subsequently tested the addition of lidar-derived information (CHM) for improving the results. The tests were conducted using two different algorithms, MLC and NN; results were evaluated by means of ground truth data, and were compared with those obtained by optical data, which are usually employed in routine monitoring activities.

The results underline how different input combinations are able to distinguish vegetation types, with variations due to inputs used and the vegetation type. Six combinations obtained >85% in overall accuracy (Table 4), over the value generally considered as a quality threshold. The most complex input (Test 12), which integrates the two SARs with CHM, outperformed the result based on SPOT data used here as a benchmark.

The synergic use of two SAR frequencies, C and L bands, improved vegetation type discrimination. As expected and confirmed by the separability analysis, discriminating grassland type is a simple task due to the absence of woody structure (Figure 3). Inputs based solely on the C band are less effective for this purpose due to the low penetration ability at this frequency. As evidenced in other research (Ferrazzoli, Guerriero, and Schiavon 1999), the simultaneous use of different frequencies is beneficial to improving SAR discrimination power, especially if texture is also added.

Woody vegetation distinction, the most difficult task, was fully obtained with Tests 11 and 12, both also including CHM, producing values included in the 80–90% range for all user's and producer's accuracies (Table 3). In this case the synergic use of SARs and derived textures is essential but insufficient to perform complete separation, and height information provided by lidar is required as discussed below. It is worth noting that even the separability obtained with SPOT optical bands (Test 13, Figure 3), interacting primarily with the leaf stratum, is low for the pair 'needle-leaved vs. shrubs and dwarf pines'. In this case the separability is higher with synergic SAR use (Test 6) than with optical bands (Test 13). This is attributed to the similarity of the spectral responses between *P. mugo* leaves (main species in this vegetation type) and conifer tree leaves in the visible and near infrared bands.

This research evidenced that between the two SARs, PALSAR was better suited for vegetation distinction, especially for the woody vegetation types: as expected, lower frequencies having higher penetration characteristics are more sensitive to difference in biomass contents (Ferrazzoli et al. 1997; Ranson and Sun 1994). When analysing SAR scenes the texture elements can be as important as backscatter data (Kourglia et al. 2010; Nelson, Ward, and Bauer 2006), and in this context the textural information was very helpful in the discrimination of woody vegetation types, especially for those having similar and high biomass, such as broadleaved and needle-leaved forests. The improvement obtained with the addition of texture features is similar to that reported by other studies, both for the level of improvement (Vaglio Laurin et al. 2013; Simard, Saatchi, and De Grandi 2000) and for being differently related to each land cover class (Nyoungui, Tonye, and Akono 2002). Texture highlights differences in the spatial arrangement of backscattered values, which in this environment correspond to differences in the structural and spatial arrangement of leaves and branches. The broadleaved and needle-leaved forest types show obvious differences in leaves, but also in the arrangement of the branches, with needle-leaved species characterized by an ordering of branches in the horizontal plane around the stem. Therefore, for the discrimination of this pair the use of textural information, better if derived from PALSAR, accounts for the improvements obtained in Tests 6 and 12.

The integration of CHM with SAR data was successfully performed. CHM always provided increased separability and accuracies, but this was especially relevant for discrimination between shrubs and dwarf pines *versus* both forest types. All three types cover the areas densely (Figure 2) and look similar when SAR penetration is limited. Similarities are found also in leaf shape and in the structural arrangement of the branches. Furthermore, the soils at those mountain elevations are rarely dry in the months of SAR acquisition (June and September) and, due to the limited height of shrubs and dwarf pines, part of the signal received returns from the ground. Thus, the backscattered values for this type can be high due to this wet ground component, becoming comparable to the values of forest or vegetation with higher biomass. In this study we found that only the information on height differences provided by CHM is able to fully solve this ambiguity. Rather, CHM does not help in those cases where vegetation heights are similar, such as in the distinction of grassland from shrubs and dwarf pines.

Adoption of the NN algorithm improved the results in most of the cases analysed, and based on a Z-test it did so significantly in 8 cases out of 13. The comparison – not reported here – of the confusion matrices obtained with NN with those obtained with MLC showed that the improvement obtained by NN was equally distributed among all vegetation types. The effectiveness of NN and SAR for complex forest analysis tasks has already been suggested (Del Frate and Solimini 2004), and it is confirmed here, as this machine learning-algorithm suffers less from noise and saturation effects. It allows the efficient integration of different inputs, exploits non-linear relationships, and thus optimizes the combination of data characterized by different ranges and dynamics.

With climate change recognized as a driving force for vegetation species distribution (Ruiz-Labourdette, Fe Schmitz, and Pineda 2013), vegetation monitoring is required to tackle present climate challenges, which are severely impacting the alpine flora in Europe (Pauli et al. 2012), besides the day-to-day management of natural resources. The frequency of vegetation monitoring will have to be raised to provide prompt information for adaptation strategies and management needs. The availability of SAR data has greatly increased in recent years and is going to grow even more with forthcoming missions such as the planned ALOS-2 and Sentinel-1, and lidar data availability has also increased considerably, making these sensors suitable for introduction in operational monitoring. This study confirms the advantages of integrating SAR sensors with lidar CHM for vegetation monitoring in a changing environment, suggesting their increased use to complement or replace optical sensors, while also considering the all-weather capabilities of SAR sensors. More importantly, SAR allows the retrieval of vegetation AGB through the analysis of backscattering levels, a task that is not achievable using optical data alone. Multifrequency SAR data are already available at the global level, and the interferometric capabilities of the forthcoming ALOS2 mission may allow the acquisition of forest height information, removing the need for lidar coverage. Here, AGB retrieval was not attempted due to the unavailability of ground truth data from the National Forest Inventory, but the possibility of exploiting SAR for different purposes and its global coverage represent strong points in favour of the adoption of this data type for forest studies worldwide.

The high heterogeneity and fragmentation of the study site, resulting in vegetation areas intermixed with agricultural zones, urban areas, rocks, and other cover types, made the production of a classification map unadvisable here and is beyond the scope of this research. Nevertheless the users' community can often access local high-resolution maps, ground truth for all cover types present, and masks of vegetated and non-vegetated areas. Users could easily introduce into their operational activities SAR and lidar CHM advanced tools to monitor specific vegetation types and to classify land cover and its changes.

Acknowledgements

References

Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.

Bork, E. W., and J. G. Su. 2007. "Integrating LIDAR Data and Multispectral Imagery for Enhanced Classification of Rangeland Vegetation: A Meta Analysis." *Remote sensing of Environment* 111: 11–24.

Chauve, A., C. Vega, F. Bretar, S. Durrieu, T. Allouis, M. Pierrot-Deseilligny, and W. Puech. 2008. "Processing Full-Waveform LiDAR Data in an Alpine Coniferous Forest: Assessing Terrain and Tree Height Quality." *International Journal of Remote Sensing* 30: 5211–5228.

Cohen, J. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20: 37–46.

Congalton, R., and K. Green. 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL: CRC/Lewis Press.

Corona, P., R. Cartisano, R. Salvati, G. Chirici, A. Floris, P. Di Martino, M. Marchetti, G. Scrinzi, F. Clementel, D. Travaglini, and C. Torresan. 2012. "Airborne Laser Scanning to Support Forest Management." *European Journal of Remote Sensing* 45: 27–37.

Dekker, R. J. 2003. "Texture Analysis and Classification of ERS SAR Images for Map Updating of Urban Areas in the Netherland." *IEEE Transactions on Geoscience and Remote Sensing* 41: 1950–1958.

Del Frate, F., F. Pacifici, and D. Solimini. 2008. "Monitoring Urban Land Cover in Rome, Italy, and its Changes by Single-Polarization Multi-temporal SAR Images." *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing* 1: 87–97.

Del Frate, F., G. Schiavon, D. Solimini, M. Borgeaud, D. H. Hoekman, and M. A. M. Vissers. 2003. "Crop Classification Using Multiconfiguration C-Band SAR Data." *IEEE Transactions on Geoscience and Remote Sensing* 41: 1611–1619.

Del Frate, F., and D. Solimini. 2004. "On a Neural Network Algorithm for Retrieving Forest Biomass from SAR Data." *IEEE Transactions on Geoscience and Remote Sensing* 42: 24–34.

Dobson, M. C., F. T. Ulaby, T. LeToan, A. Beaudoin, E. S. Kasischke, and N. Christensen. 1992. "Dependence of Radar Backscatter on Coniferous Forest Biomass." *IEEE Transactions on Geoscience and Remote Sensing* 30: 412–415.

Dowling, R., and A. Accad. 2003. "Vegetation Classification of the Riparian Zone along the Brisbane River, Queensland, Australia, Using Light Detection and Ranging (LiDAR) Data and Forward Looking Digital Video." *Canadian Journal of Remote Sensing* 29: 556–563.

Englhart, S., V. Keuck, and F. Siegert. 2012. "Modeling Aboveground Biomass in Tropical Forests Using Multi-Frequency SAR Data — A Comparison of Methods." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5: 298–306.

Evans, T., M. Costa, K. Telmer, and T. Silva. 2010. "Using ALOS/PALSAR and RADARSAT-2 to Map Land Cover and Seasonal Inundation in the Brazilian Pantanal." *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing* 3: 560–575.

Ferrazzoli, P., and L. Guerriero. 1995. "Radar Sensitivity to Tree Geometry and Woody Volume: A Model Analysis." *IEEE Transactions on Geoscience and Remote Sensing* 33: 360–371.

Ferrazzoli, P., L. Guerriero, and G. Schiavon. 1999. "Experimental and Model Investigation on Radar Classification Capability." *IEEE Transactions on Geoscience and Remote Sensing* 37: 960–968.

Ferrazzoli, P., S. Paloscia, P. Pampaloni, G. Schiavon, S. Sigismondi, and D.Solimini. 1997. "The Potential of Multifrequency Polarimetric SAR in Assessing Agricultural and Arboreous Biomass." *IEEE Transactions on Geoscience and Remote Sensing* 35: 5–17.

Geerling, G. W., M. Labrador-Garcia, J. Clevers, A. Ragas, and A. J. M. Smits. 2007. "Classification of Floodplain Vegetation by Data-Fusion of Spectral (CASI) and LiDAR data." *International Journal of Remote Sensing* 28: 4263–4284.

Hanewinkel, M., D. A. Cullmann, M. J. Schelhaas, G. J. Nabuurs, and N. E. Zimmermann. 2013. "Climate Change may cause Severe Loss in the Economic Value of European Forest Land." *Nature Climate Change* 3: 203–207.

Haralick, R. M., K. Shanmugan, and I. Dinstein. 1973. "Textural Features for Image Classification." *IEEE Transactions on Systems, Man, and Cybernetics* 3: 610–621.

Hill, R. A., and A. G. Thomson. 2005. "Mapping Woodland Species Composition and Structure Using Airborne Spectral and LiDAR Data." *International Journal of Remote Sensing* 17: 3763–3779.

Hollaus, M., W. Wagner, C. Eberhöfer, and W. Karel. 2006. "Accuracy of Large-Scale Canopy Heights Derived from LiDAR Data under Operational Constraints in a Complex Alpine Environment." *ISPRS Journal of Photogrammetry and. Remote Sensing* 60: 323–338.

Jochem, A., M. Hollaus, M. Rutzinger, and B. Höfle. 2011. "Estimation of Aboveground Biomass in Alpine Forests: A Semi-Empirical Approach Considering Canopy Transparency Derived from Airborne LiDAR Data." *Sensors* 11: 278–295.

Kasischke, E. S., N. L. Christensen Jr., and L. L. Bourgeau-Chavez. 1995. "Correlating Radar Backscatter with Components of Biomass in Loblolly Pine Forests." *IEEE Transactions on Geoscience and Remote Sensing* 33: 643–659.

Kohavi, R., and G. H. John. 1997. "Wrappers for Feature Subset Selection." *Journal of Artificial Intelligence* 97: 273–324.

Kourglia, A., M. Ouarzeddinea, Y. Oukilb, and A. Belhadj-Aissaa. 2010. "Texture Modelling for Land Cover Classification of Fully Polarimetric SAR Images." *International Journal of Image and Data Fusion* 3: 129–148.

Kraus, K., and N. Pfeifer. 1998. "Determination of Terrain Models in Wooded Areas with Airborne Laser Scanner Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 53: 193–203.

Kurvonen, L., and M. T. Hallikainen. 1999. "Textural Information of Multitemporal ERS-1 and JERS-1 SAR Images with Applications to Landand Forest Type Classification in Boreal Zone." *IEEE Transactions on Geoscience and Remote Sensing* 37: 680–689.

Lee, J. S., M. R. Grunes, T. Ainsworth, I. Hajnsek, T. Mette, and K. P. Papathanassiou. 2005. "Forest Classification Based on L-Band Polarimetric and Inteferometric SAR Data." In *POLinSAR 2005 Workshop*, Frascati, January 17–21.

Lehmann, E. A., P. A. Caccetta, Z. Zhou, S. J. McNeill, X. Wu, and A. L. Mitchell. 2012. "Joint Processing of Landsat and ALOS-PALSAR Data for Forest Mapping and Monitoring." *IEEE Transactions on Geoscience and Remote Sensing* 50: 55–67.

LeToan, T., A. Beaudoin, J. Riom, and D. Guyon. 1992. "Relating Forest Biomass to SAR Data." *IEEE Transactions on Geoscience and Remote Sensing* 30: 403–411.

Li, G., D. Lu, E. Moran, L. Dutra, and M. Batistella. 2012. "A Comparative Analysis of ALOS PALSAR L-band and RADARSAT-2 C-band Data for Land-Cover Classification in a Tropical Moist Region." *ISPRS Journal of Photogrammetry and Remote Sensing* 70: 26–38.

Ling, F., Z. Li, E. Chen, and Q. Wang. 2009. "Comparison of ALOS PALSAR RVI and Landsat TM NDVI for Forest Area Mapping." In *Proceedings of the 2nd Asian-Pacific Conference on Synthetic Aperture Radar* (APSAR 2009), 132–135, Xi'an, October 26–30.

Longepe, N., P. Rakwatin, O. Isoguchi, M. Shimada, Y. Uryu, and K. Yulianto. 2011. "Assessment of ALOS PALSAR 50 m Orthorectified FBD Data for Regional Land Cover Classification by Support Vector Machines." *IEEE Transactions on Geoscience and Remote Sensing* 49: 2135–2150.

Lu, D., M. Batistella, and E. Moran. 2007. "Land-Cover Classification in the Brazilian Amazon with the Integration of Landsat ETM+ and RADARSAT Data." *International Journal of Remote Sensing* 28: 5447–5459.

Lu, D., and Q. Weng. 2007. "A Survey of Image Classification Methods and Techniques for Improving Classification Performance." *International Journal of Remote Sensing* 28: 823–870.

Minchella, A., F. Del Frate, F. Capogna, S. Anselmi, and F. Manes. 2009. "Use of Multitemporal SAR Data for the Monitoring of the Vegetation Recovery in Burned Areas." *Remote Sensing of Environment* 113: 588–597.

Moller, M. F. 1993. "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning." *Neural Networks* 6: 525–533.

Nelson, M. D., K. T. Ward, and M. E. Bauer. 2006. "Forest-Cover–Type Separation Using RADARSAT-1 Synthetic Aperture Radar Imagery." In *Proceedings of the Eighth Annual Forest Inventory and Analysis (FIA) Symposium*, 303–306, Monterey, CA, October 16–19.

Notarnicola, C., A. Frick, S. Kass, P. Rastner, G. Pulighe, and M. Zebisch. 2009. "Semi-automatic Classification Procedure for Updating Landuse MAPS with High Resolution Optical Images." In *Proceedings of International Geoscience and Remote Sensing Symposium*, III-975–III-978, Cape Town, July 12–17.

Nyoungui, A. N., E. Tonye, and A. Akono. 2002. "Evaluation of Speckle Filtering and Texture Analysis Methods for Land Cover Classification from SAR Images." *International Journal of Remote Sensing* 23: 1895–1925.

Onojeghuo, A. O., and G. A. Blackburn. 2011. "Optimising the Use of Hyperspectral and LiDAR Data for Mapping Reedbed Habitats." *Remote Sensing of Environment* 115: 2025–2034.

Pauli, H., M. Gottfried, S. Dullinger, O. Abdaladze, M. Akhalkatsi, J. L. Benito Alonso, G. Coldea, J. Dick, B. Erschbamer, R. Fernández Calzado, D. Ghosn, J. I. Holten, R. Kanka, G. Kazakis, J. Kollár, P. Larsson, P. Moiseev, D. Moiseev, U. Molau, J. Molero Mesa, L. Nagy, G. Pelino, M. Puşcaşş, G. Rossi, A. Stanisci, A. O. Syverhuset, J. P. Theurillat, M. Tomaselli, P. Unterluggauer, L. Villar, P. Pacal Vittoz, and G. Grabherr. 2012. "Recent Plant Diversity Changes on Europe's Mountain Summits." *Science* 336: 353–355.

Rahman, M. M., and J. T. S. Sumantyo. 2010. "Mapping Tropical Forest Cover and Deforestation Using Synthetic Aperture Radar (SAR) Images." *Applied Geomatics* 2: 113–121.

Ranson, K. J., and G. Sun. 1994. "Mapping Biomass of a Northem Forest Using Multifrequency SAR Data." *IEEE Transactions on Geoscience and Remote Sensing* 32: 388–396.

Richards, J. A. 1999. *Remote Sensing Digital Image Analysis*. Berlin: Springer-Verlag.

Rignot, E., J. Way, C. Williams, and L. Viereck. 1994. "Radar estimates of Aboveground Biomass in Boreal Forests of Interior Alaska." *IEEE Transactions on Geoscience and Remote Sensing* 32: 1117–1124.

Ruiz-Labourdette, D., M. Fe Schmitz, and F. D. Pineda. 2013. "Changes in Tree Species Composition in Mediterranean Mountains under Climate Change: Indicators for Conservation Planning." *Ecological Indicators* 24: 310–323.

Schneiderbauer, S., M. Zebisch, and C. Steurer. 2007. "Applied Remote Sensing in Mountain Regions." *Mountain Research and Development* 27: 286–287.

Simard, M., S. Saatchi, and G. F. De Grandi. 2000. "The Use of Decision Tree and Multiscale Texture for Classification of JERS-1 SAR Data over Tropical Forest." *IEEE Transactions on Geoscience and Remote Sensing* 38: 2310–2321.

Touzi, R., R. Landry, and F. J. Charbonneau. 2004. "Forest Type Discrimination Using Calibrated C-band Polarimetric SAR Data." *Canadian. Journal of Remote Sensing* 30: 543–551.

Vaglio Laurin, G., V. Liesenberg, Q. Chen, L. Guerriero, F. Del Frate, A. Bartolini, D. Coomes, B. Wilebore, J. Lindsell, and R. Valentini. 2013. "Optical and SAR Sensor Synergies for Forest and Land Cover Mapping in a Tropical Site in West Africa." *International Journal of Applied Earth Observation and Geoinformation* 21: 7–16.

Voigt, T., J. V. Minnen, M. Erhard, D. Viner, R. Koelemeijer, and M. Zebisch. 2004. *Indicators of Europe's Changing Climate*. Copenhagen: EEA.

Wijaya, A., P. R. Marpu, and R. Gloaguen. 2008. "Geostatistical Texture Classification of Tropical Rainforest in Indonesia." In *Quality Aspect in Spatial Data Mining*, edited by J. S. Stein and W. Bijker, 199–210. Boca Raton, FL: CRC Press.

Yatabe, S. M., and D. G. Leckie. 1996. "Clearcut and Forest-Type Discrimination in Satellite SAR Imagery." *Canadian Journal of Remote Sensing* 21: 455–467.

Zhang, F., M. Xu, C. Xie, Z. Xia, K. Li, and X. Wang. 2012. "Forest and Deforestation Identification Based on Multitemporal Polarimetric RADARSAT-2 Images in Southwestern China." *Journal of Applied Remote Sensing* 6: 063527.

Chapter 5

Optical and SAR sensor synergies for forest and land cover mapping in a tropical site in West Africa

Optical and SAR sensor synergies for forest and land cover mapping in a tropical site in West Africa

Gaia Vaglio Laurin [a,g,*], Veraldo Liesenberg [b], Qi Chen [d], Leila Guerriero [a], Fabio Del Frate [a], Antonio Bartolini [a], David Coomes [c], Beccy Wilebore [c], Jeremy Lindsell [e], Riccardo Valentini [f,g]

[a] Tor Vergata University, Dipartimento di Informatica, Sistemi e Produzione, Via del Politecnico 1, 00133 Rome, Italy
[b] Remote Sensing Group, Institute for Geology, TU-Bergakademie Freiberg, Bernhard von Cotta Str. 2, 09599 Freiberg, Germany
[c] Forest Ecology and Conservation Group, Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK
[d] Department of Geography, University of Hawai'i at Mānoa, 422 Saunders Hall, 2424 Maile Way, Honolulu, HI 96822, USA
[e] The Royal Society for the Protection of Birds, The Lodge, Sandy, Beds. SG19 2DL, UK
[f] Department of Forest Resources and Environment, University of Tuscia, Viterbo I-01100, Italy
[g] CMCC – Centro Mediterraneo per i Cambiamenti Climatici, via Augusto Imperatore (Euro-Mediterranean Center for Climate Change), Lecce 73100, Italy

ARTICLE INFO

ABSTRACT

The classification of tropical fragmented landscapes and moist forested areas is a challenge due to the presence of a continuum of vegetation successional stages, persistent cloud cover and the presence of small patches of different land cover types. To classify one such study area in West Africa we integrated the optical sensors Landsat Thematic Mapper (TM) and the Advanced Visible and Near Infrared Radiometer type 2 (AVNIR-2) with the Phased Arrayed L-band SAR (PALSAR) sensor, the latter two on-board the Advanced Land Observation Satellite (ALOS), using traditional Maximum Likelihood (MLC) and Neural Networks (NN) classifiers. The impact of texture variables and the use of SAR to cope with optical data unavailability were also investigated. SAR and optical integrated data produced the best classification overall accuracies using both MLC and NN, respectively equal to 91.1% and 92.7% for TM and 95.6% and 97.5% for AVNIR-2. Texture information derived from optical images was critical, improving results between 10.1% and 13.2%. In our study area, PALSAR alone was able to provide valuable information over the entire area: when the three forest classes were aggregated, it achieved 75.7% (with MCL) and 78.1% (with NN) overall classification accuracies. The selected classification and processing methods resulted in fine and accurate vegetation mapping in a previously untested region, exploiting all available sensors synergies and highlighting the advantages of each dataset.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The classification of moist tropical areas is a challenging task for a number of reasons. Firstly, the atmospheric conditions cause persistent cloud cover and limit the use of optical data. Secondly, the landscape is often fragmented into small patches of different land use and land cover types. Furthermore, the peculiar characteristics of natural vegetation cause a continuous transition among vegetation types, making it difficult to spectrally differentiate various successional stages, as exemplified by Lu et al. (2007) over Amazonian moist forested regions.

The Upper Guinean forests of West Africa have experienced a dramatic decrease of their original extent. Logging, mining, hunting and human population growths are still placing extreme stress on this biodiversity hotspot (CEPF, 2003). The few classification studies available for West Africa report problems in mapping forest and vegetation classes (Igue et al., 2006; Judex et al., 2006). Nevertheless, mapping activities are extremely important for conservation and planning issues and with respect to the emerging REDD+ (Reducing Emissions from Deforestation and Forest Degradation in Developing Countries) program (Gibbs et al., 2007). Detailed mapping is also important for national planning in many tropical countries, where local communities rely on woody vegetation as a primary source of products and energy (Avitabile et al., 2012).

Optical sensors have been the primary data sources for land cover classification since the launch of the Landsat satellite series in early 1970s. In recent years, Synthetic Aperture Radar (SAR) sensors have emerged as important tools for vegetation studies, being well suited to detect volumetric scattering, especially at L- and P-bands (Rahman and Sumantyo, 2010; Santos et al., 2004; Simard et al., 2000). Since radar can penetrate clouds, it is a supplementary data source when atmospheric conditions hamper optical data use (Lehmann et al., 2012; Lu et al., 2007; Mitchard et al., 2011). In general, the combination of the two data types is considered beneficial

(Lefsky and Cohen, 2003) because optical data allows the measurement of the reflectance of the topmost layer of the canopy and SAR data deliver useful geometric information without being affected by weather conditions. However, the benefits of data integration can vary according to landscape and specific sensors characteristics, making it useful to explore the synergistic use of different datasets in the West African region where, according to our knowledge, no classification study exists at a fine scale.

In addition to reflectance or backscattering coefficient, texture (the spatial arrangement of color or intensities of pixels) has been proved useful to improve classification accuracy (Berberoglu et al., 2007; Chica-Olmo and Abarca-Hernandez, 2000; Dekker, 2003; Dorigo et al., 2012; Lu et al., 2007). Textural differences can potentially be observed in optical data between the uneven canopy 'topography' of secondary and mixed semi-deciduous forests and the more uniform mature evergreen forests, classes which are usually hard to distinguish. In fact, forests of different ages and types are known to structural differences, i.e. in canopy gaps and/or in tree height composition (Yavitt et al., 1995), resulting in shadows and different reflectance values that influence the texture variables calculated over near infrared (NIR) or short wave infrared highly reflecting bands. In the case of semi-deciduous forest, bare trees are seen at these wavelengths as low reflecting objects and therefore produce less smooth textures. It is known that shadows are a function of canopy structure and cover and Asner (1998) reports that both red and NIR wavelength regions are highly sensitive to sub-pixel shadow fractions in Amazon tropical forests. A number of studies have used texture from multiple remotely sensed data such as Landsat and Advanced Land Observation Satellite (ALOS) Phased Arrayed L-band SAR (PALSAR; 1.27 GHz center frequency, 23 cm wavelength) for mapping tropical environments (Erasmi and Twele, 2009; Kuplich, 2006; Longepe et al., 2011; Wijaya et al., 2008). For instance, Chan et al. (2003) used textures from Landsat data and machine learning techniques to improve discrimination of logged versus non-logged semi-deciduous forest classes in Central Africa. In Gabon, Simard et al. (2000) used texture from Japanese Earth Resources Satellite (JERS-1) SAR data to refine the classification of the flooded forest class. The permutation of different bands, window sizes, and texture variables could potentially generate a large number of texture features: given a training dataset of a certain size, including too many variables may decrease classification accuracy and a feature selection method can be beneficial (Price et al., 2002). In our research, a relatively simple and automatic wrapper based approach for feature selection is proposed, which is considered optimal for supervised learning problems (Talavera, 2005).

In this study, we primarily investigate the potential of combining optical and radar sensors for discriminating land cover classes for a moist tropical area which has never been classified by high resolution remote sensing before. The optical sensors to be tested are Landsat TM (Thematic Mapper) and ALOS AVNIR-2 (Advanced Visible and Near Infrared Radiometer type 2); the SAR sensor used is the dual polarimetric ALOS PALSAR. The specific research objectives are multifold: (1) to assess the results obtained by sensor integration; (2) to explore the degree SAR can support land cover and vegetation mapping when atmospheric conditions affect optical data availability and quality, (3) to assess the value of textures for improving vegetation classification.

2. Materials and methods

2.1. Study area

The study area (covering a total of 7749 km^2) spans the border of Sierra Leone and Liberia and includes the recently established Gola
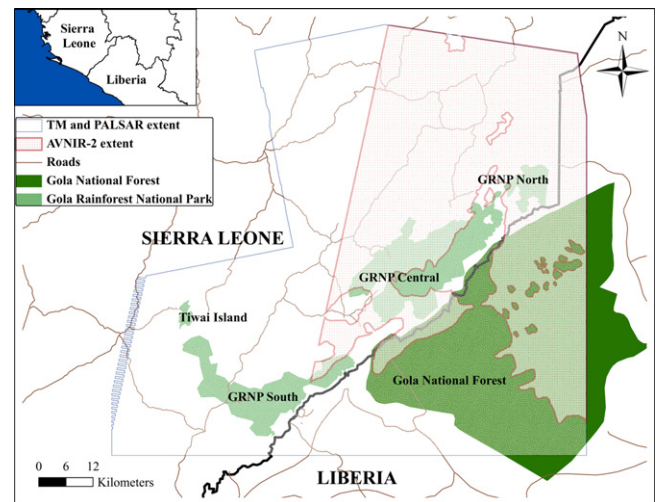
Fig. 1. Location of the study area; protected areas are shown in green. Area covered by TM and ALOS PALSAR data is shown in blue and by AVNIR-2 in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Rainforest National Park (GRNP) which is composed of three disjunct areas covering about 710 km^2 (North, Central and South), and most of the Liberian Gola National Forest (about 2300 km^2). In the north-west of the study area there is a fragmented landscape comprising small patches of disturbed forest, farmbushes, plantations, active agriculture areas, bare lands and settlements; in the south-east the landscape is dominated by forests (Fig. 1). Forests of the region are classified mainly as lowland moist evergreen with some drier parts occurrence, are dominated by *Fabaceae*, *Euphorbiaceae* and *Sterculiaceae* families (Cole, 1993).

Within Sierra Leone, the protected areas comprise the GRNP and Tiwai Island Wildlife Sanctuary. Inside GRNP, recent field surveys (Klop et al., 2008) identified a range of vegetation types: moist evergreen forest, moist semi-deciduous forest, freshwater inland swamp forest, forest regrowth and secondary/disturbed forest, farmbush, herbaceous swamps and floodplains. Commercial logging in Gola South was carried out in the periods 1963–1965 and 1975–1989. Since 1989, Gola has been subject to a conservation program. Outside of the protected areas, land cover is largely influenced by human activities and includes: disturbed and secondary forest, farmbush and shrubland/savanna, plantation, agriculture, bare soil, water courses and ponds.

In Liberia, the study area is sparsely populated and more forested. Verschuren (1983) indicated that the proposed Lofa-Mano Park (now referred to as Gola National Forest) covers a large area of evergreen rainforest in the south range, with semi-deciduous moist forest gradually taking over to the north, and with patches of low bush, marshes and some savannah on lateritic soil.

Overall, the area is characterized by a moist tropical climate with annual rainfall around 2500–3000 mm, a wet season lasting from May to October, and an altitude in the range 70–410 m with no abrupt elevation changes. The dry period occurs between December and March, and corresponds to the semi-deciduous phenological stage of vegetation in the moist forest.

2.2. Remote sensing data

We collected Landsat TM, ALOS AVNIR-2, and PALSAR scenes for image classification. The analysis of Landsat imagery from 1986 to the present day revealed cloud cover and haze over the area of interest in all seasons, as often reported for tropical forested regions (Asner, 2001). The best image, with respect to the time of field data

Table 1

Summary of remote sensing data used in the research.

| Sensor	Dates	Data characteristics
Landsat TM	15 Jan. 2007	30 × 30 m spatial resolution, 7 spectral bands Band 1: 0.45–0.52 m (Blue); Band 2: 0.52–0.60 m (Green); Band 3: 0.63–0.69 m (Red); Band 4: 0.76–0.90 m (Near infrared); Band 5: 1.55–1.75 m (Short wave infrared); Band 6: 10.4–12.5 m (Thermal infrared); Band 7: 2.08–2.35 m (Short wave infrared). L1T product. Thermal band neglected
ALOS PALSAR	22 June 2007, 24 Aug. 2007	SAR L-band in HH and HV polarization. FBD Level 1.1 product: Single Look Complex provided in slant range geometry
ALOS AVNIR-2	9 Dec. 2009	10 × 10 m spatial resolution, 4 spectral bands, Band 1: 0.42–0.50 m; Band 2: 0.52–0.60 m; Band 3: 0.61–0.69 m; Band 4: 0.76–0.89 m; Level 1B2 product
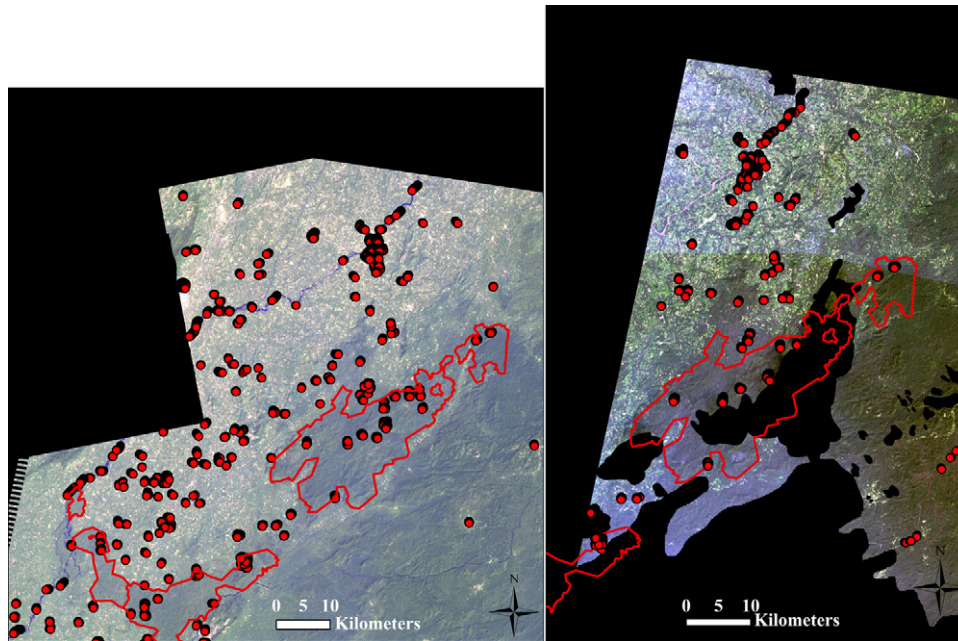
Fig. 2. Left: Landsat TM 2007 false color image (Red: band 7, Green: band 4, and Blue: band 3) cut to match the area in common with ALOS PALSAR. Right: AVNIR-2 cloud-masked true color image (Red: band 3, Green: band 2, and Blue: band 1), where the black areas are affected by cloud and thus manually masked out. The Gola Rainforest National Park and Tiwai Island Wildlife Reserve boundaries are shown in red. The locations where ROIs have been generated are shown as red points in both images. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
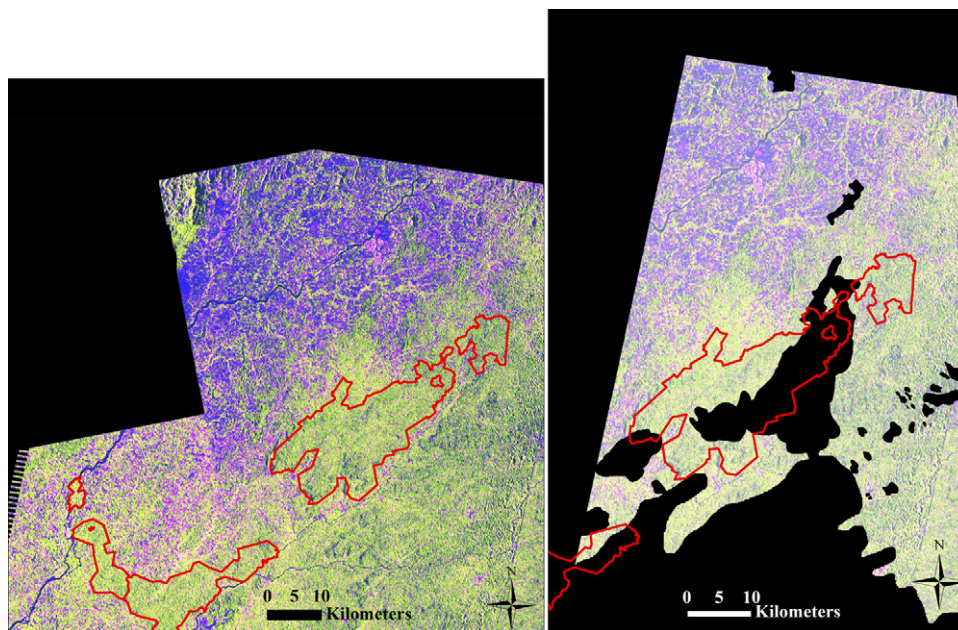
Fig. 3. Left: Three ALOS PALSAR mosaicked scenes (two dated 22 June 2007 and one 24 August 2007) 30 m pixel size, masked over the area common to the Landsat image. Right: Three ALOS PALSAR mosaicked scenes, 10 m pixel size, masked over the area common to the AVNIR-2 image. Both are false color composites (Red: HH, Green: HV, Blue: difference HH-HV). The Gola Rainforest National Park and Tiwai Island Wildlife Reserve boundaries are shown in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Characteristics of the selected land use classes and forest physiognomies.

Secondary (degraded) forest (SF)	Characterized by trees with height > 5 m, sometimes with presence of understory crops (coffee, cocoa), canopy cover > 20%, evidence of human disturbance, logging history, >15–20 years from regeneration if cleared
Farmbush/Shrubland/ Savanna (FB)	Dominated by non-tree vegetation such as shrubs and/or grasses, canopy cover < 20%, agriculture can be present but should not occupy more than 25% of the area, trees usually less than 5 m tall. Can be land which has been abandoned after shifting cultivation occurred few years before (<15–20)
Plantation (OP)	Dominated by oil palm or rubber, evidence of human activity, canopy cover >20%, tree height > 5 m
Agriculture (AG)	Dominated by specific crops such as rice on irrigated soil, usually found in small patches, evidence of human activity
Bare soil/Urban areas (BS)	Represented by villages, open areas not yet cultivated or exposed substrate with little or no vegetation, including recently burned land
Evergreen forest (EF)	Tree height > 5 m and canopy cover > 60%, no evidence of human disturbance
Semi-deciduous forest (DF)	Tree height > 5 m and canopy cover > 60%, no evidence of human disturbance, presence of semi-deciduous species
Water (W)	Rivers, water courses and ponds

collection was a Landsat TM scene from 15 January 2007, unevenly affected by haze and provided at level L1T (terrain corrected and projected in UTM zone 29N, Table 1, Fig. 2). This is coincident with the dry season and is therefore better suited for semi-deciduous vegetation detection. We acquired two AVNIR-2 scenes, dated 9 December 2009 (Table 1, Fig. 2), and three ALOS PALSAR FBD (Fine Beam Dual) scenes in slant range single look complex (SLC) format (Level 1.1), with pixel spacing 9.3 m in range and 2.7–4.5 m in azimuth, two being dated 22 June 2007 and one 24 August 2007 (Table 1, Fig. 3).

2.3. Ground truth data acquisition

The ground truth was derived from field surveys realized in Sierra Leone and it was linked to the remotely-sensed images with the aid of visual interpretation.

Inside GRNP, 656 forest plots realized in 2006 were classified as having prevalence of evergreen or semi-deciduous tree species using TWINSPAN (Hill, 1979), a clustering algorithm for classifying species and samples. Outside the Park, a total of 872 land cover validation points were collected from 2008 to 2011, including: agriculture (including slash-and-burn, farmbush, and rice fields), plantations (oil palm and rubber), secondary/disturbed forest, undisturbed forest, burnt and urban areas, mining and water courses. These land cover/land use classes were re-organized as described in Table 2, thus incorporating the phenological information derived from the 2006 field survey (Klop et al., 2008).

The surveys information allowed the selection of locations around which spectrally homogeneous Regions of Interest (ROIs) were digitized on screen. To mitigate any error introduced by temporal mismatching between remote sensing data and field surveys, and spatial mismatching caused by GPS inaccuracy, we selected only locations for which full agreement between available optical imagery and the field information was observed. We double-checked the selection of locations using also Google Earth very high resolution data, two Kompsat-2 (dated 17 April 2009) and three Ikonos (22 December 2009, 19 March 2006 and 19 February 2003), taking into account the differences on date acquisition between scenes. Cloud cover presence over the images seriously limited the use of field information: a total of 208 locations were selected in

the area covered by the TM image and 82 in the area covered by AVNIR-2, with most of the AVNIR-2 locations (70) coincident with those selected for TM data (Fig. 2). For Liberia, few forest locations were selected thanks to the data provided by GRNP staff and local experts (Lindsell, unpublished data). Different studies (Bayol and Chevalier, 2004; Shearman, 2009; World Bank, 2011) have indicated the similarity between the Gola North block and the Liberian forests, with a majority cover of undisturbed forest.

The TM ROIs were subdivided into three parts: training (70%), validation (15%) and test (15%), for the purposes of classifying the TM and PALSAR scenes at 30 m spatial resolution, with the validation dataset used only in NN implementation. The dataset, initially different in amount for the various classes, was reduced in size by a random selection to obtain 2500 pixels for each land cover class. The ROIs are as evenly distributed as possible over the entire image; anyway a certain degree of spatial correlation occurs for water and oil palm plantation classes, due to their aggregated distribution in the study area and to limited accessibility.

Prior to define the 82 ROIs for the AVNIR-2 images, we checked over TM the coincident locations and ROIs digitized: TM ROIs were retained if they covered a spectrally and visually homogeneous area in AVNIR-2 image. When the finer spatial resolution of AVNIR-2 allowed for identification of more than one land cover type inside a ROI, that region was either discarded, or else only the part of the ROI matching the indicated land cover was retained. As for TM, the ROIs were subdivided into three sets and randomly reduced to obtain 10,000 pixels per land cover class, for the purposes of classifying the AVNIR-2 and PALSAR scenes at 10 m spatial resolution.

Statistical tests (not reported here) confirmed the absence of significant differences in the TM and AVNIR-2 ground truth for the various land cover classes, as well as in the different data sets used for training, validation and testing. With the described procedure we derived an AVNIR-2 ground truth as compatible as possible to the TM one while maintaining spectral purity, allowing for better understanding the performances of each sensor in classification tasks with respect to SAR integration.

2.4. Remote sensing image preprocessing

Atmospheric corrections were performed on the TM and AVNIR-2 images to obtain the hemispherical directional reflectance factor (HDRF) using the Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm in ENVI version 4.7 (Exelis Visual Information Solutions, Boulder, Colorado). This is based on a MODTRAN4 approach for path scattered radiance, absorption, and adjacency effects (Felde et al., 2003). After atmospheric correction, haze was still detected on the TM image over the whole Sierra Leone area, while clearer conditions characterized most of the Liberian side. Haze removal was attempted using the Haze Tool developed by Hu et al. (2009), which is a refinement of the haze optimized transformation (HOT) algorithm (Zhang et al., 2002; Zhang and Guindon, 2003). The corrected scene was less affected by uneven haze distribution, but still generated errors in classification results. Since the haze problem occurred in the visible bands, we excluded bands 1–3 of the TM image in our analysis.

An orthorectification based on a digital elevation model (DEM) from the ASTER derived Global Digital Elevation Map mission (GDEM, Tachikawa et al., 2011) was also performed on AVNIR-2 images. A mask was applied in order to exclude cloudy areas and cloud shadows, manually digitized on screen (Fig. 2). Then, the two AVNIR-2 scenes were mosaicked and co-registered to the TM scene (with Root Mean Square (RMS) of 0.51 at 30 m pixel size), and eventually resampled to the original spatial resolution of 10 m. The AVNIR-2 mosaic covers only a portion of the study area due to cloud-masked zones and unavailability of an adjacent scene.

The FBD PALSAR images were multi-looked, terrain-corrected and geo-coded to a 15 m spatial resolution using ASF (Alaska Satellite Facility) MapReady 3.0.6 and the aforementioned ASTER GDEM. SAR scenes were calibrated according to Shimada et al. (2009), mosaicked, co-registered to a TM image (RMS of 0.96 pixels) and then resampled twice to produce two scenes: one at 30 m to match Landsat spatial resolution, and one at 10 m to match AVNIR-2. A Frost adaptive filter with a moderate window size of 5×5 was applied to the two SAR datasets (Fig. 3) to reduce speckle.

Our image analysis strategy was to combine either Landsat TM or AVNIR-2 with ALOS PALSAR in classification. Since the TM image and the PALSAR scenes do not overlay perfectly, we created subset images for TM and PALSAR corresponding to their area of overlap. The whole AVNIR-2 image falls within PALSAR, so there was no need to create a subset for it. Although the TM and AVNIR-2 images do not have equal area coverage, the ground truth for all classes is well represented in both images (Figs. 2 and 3).

2.5. Texture variables

Five texture variables including mean, entropy, correlation, variance, and second moment based on Grey-Level Co-Occurrence Matrix (GLCM) (Haralick et al., 1973) were created with ENVI version 4.7. Dorigo et al. (2012) describe extensively the GLCM texture characteristics and the need for selection of appropriate features and kernel size. We generated texture for the bands useful for vegetation detection (Asner, 1998), namely TM bands 4, 5, AVNIR-2 bands 3,4, plus for both ALOS PALSAR polarizations, using 64 grayscale quantization levels, 1 pixel shift and using 3×3, 5×5, 7×7, 9×9 and 15×15 window sizes. To reduce the number of features entering in classification a wrapper-based approach for texture variable selection was here adopted. Several features selection methods exist, such as wrapper- and filter-based (Kohavi and John, 1997); filter-based methods are algorithms that use only training data to calculate the feature weights, whereas wrapper-based algorithms use feedback from a classifier to guide the search. We automated the comparison of the overall accuracies and Kappa coefficients obtained with MLC, using the training and validation ROIs specific for each dataset. We added one texture at a time to optical and SAR data, repeating the tests with increasing window size. We selected one variable for each band or channel, choosing the one that when added to the original data, produced the best classification result.

The automated procedure selected the following texture variables: for TM, entropy on band 4 and mean on 5 (9×9 window); for AVNIR-2, mean on bands 3 and 4 (15×15 window); for ALOS PALSAR, mean on both HH and HV polarizations (9×9 window). Any further increase in window size did improved classification accuracy. The selected window size of 9×9 for TM is well in agreement with other studies based on the same sensor (Li et al., 2011; Zhu et al., 2012).

2.6. Classification algorithms and selected tests

The Maximum likelihood classifier (MLC) is probably the most widely used parametric classifier (Richards and Jia, 1999), considered robust and still employed in recent forestry research (Herold et al., 2004; Lehmann et al., 2012; Lu et al., 2007; Rahman and Sumantyo, 2010). It was here used to select texture features and as benchmark classification algorithm for all tests. Among the most commonly used non-parametric classifiers is neural networks (NN), which has several advantages including its non-parametric nature, easy adaptation to different input data, learning and generalization capabilities (Lu and Weng, 2007), and is a rather competitive classifier even if compared with other sophisticated approaches (Pacifici et al., 2002; Licciardi et al., 2009). Atkinson and Tatnall (1997)

Table 3
Datasets used as classification inputs and tested with MLC and/or NN approaches.

Optical data
  1) TM bands 4,5,7
  2) TM bands 4,5,7 + texture (entropy b4, mean b5; 9×9 window)
  3) AVNIR-2 all bands
  4) AVNIR-2 all bands + texture (mean b3, mean b4; 15×15 window)
SAR data
  5) ALOS PALSAR HH-HV bands
  6) ALOS PALSAR HH-HV bands + texture (mean HH, mean HV; 9×9 window)
Combined SAR + optical data
  7) TM bands 4,5,7 + texture + PALSAR HH HV
  8) TM bands 4,5,7 + TM texture + PALSAR HH HV + PALSAR texture
  9) AVNIR-2 + texture + PALSAR HH HV
  10) AVNIR-2 + texture + PALSAR HH HV + PALSAR texture

reported that neural networks usually overcome typical problems experienced using parametric classifiers for land targets and are suitable for fusing multisource spatial data. Del Frate and Solimini (2004) critically analyzed NN in forestry problems, also in view of collinearity of inputs, observing that NN less suffers with respect to other algorithms from noise and saturation in L-band SAR data.

Here NN were used only for selected input combinations, with the aim of assessing the extent of improvement that a machine learning algorithm can bring in classification tasks. The NN were trained and tested with the same ground truth sets used in MLC classifications, while the validation dataset was used exclusively for NN implementation, avoiding the overfitting that results from the termination of the training phase when the error over the test dataset reaches a minimum (Bishop, 1995). The Conjugate Gradient (SGC; Moller, 1993) method was employed in the feed-forward neural networks. NN architecture was selected for each input dataset with a trial and error approach.

In our test strategy we first compared the classification accuracies obtained with MLC and NN algorithms on progressively complex inputs: (1) optical reflectance alone, (2) optical reflectance with its texture variables, (3) optical reflectance with its texture features, and SAR backscatter, and (4) optical reflectance, SAR backscatter, and texture variables from both. We repeated the tests over the TM and the AVNIR-2 datasets, thus exploring data integration with different optical datasets.

Finally, we performed additional tests using SAR data alone, and with its derived textures, to assess its potential to support land cover mapping in areas persistently affected by clouds or haze and replace optical data. In these tests, we used two additional classification schemes by (1) merging two forest classes (secondary and semi-deciduous), and (2) merging all three classes (including evergreen) to reduce the complexity of the classification task and the number of land cover classes to be seven and six, respectively. Merged ROIs were randomly sampled to obtain again 2500 pixels per class. A list of the classification tests is reported in Table 3.

Classification accuracies were assessed using overall accuracy, error matrix and Kappa coefficient of agreement (Cohen, 1960). The overall accuracy is the percentage of the correctly classified pixels in the validation dataset whereas the Kappa coefficient is a measure of the difference between the actual agreement and the chance agreement. The statistical differences in the Kappa coefficients of MLC and NN classifications at a 95% level of significance was determined using the Z-test (Congalton and Green, 1999), which is performed for a pairwise comparison of the proposed methods and takes into account the ratio between the difference value of two Kappa coefficients and the difference of their respective variance.

3. Classification results

Optical data only: the MLC classification based on optical bands reached an overall accuracy of 80.4% for AVNIR-2 and 73.1% for TM

Table 4
Classification results for optical data obtained with MLC.

Land cover class	TM bands 4,5,7		TM bands 4,5,7 + TM texture		AVNIR-2 bands		AVNIR-2 bands + AVNIR-2 texture	
	PA%	UA%	PA%	UA%	PA%	UA%	PA%	UA%
AG	59.6	73.6	76	84.3	93.8	80.3	96.6	95.1
BS	99.6	99.3	100	99.6	82.2	96.7	98.3	99.1
EF	65.3	72.6	89.3	87	81.4	78.7	92.5	87
FB	81	68.9	86.6	81	85.2	87.1	95.4	94
OP	84.9	65.3	94.6	95.3	88.8	70.7	96.2	96.3
SF	38.6	57.6	65	70.3	36.6	56	80.8	88.9
DF	62.9	54.3	78	71.6	76.3	71.3	90.4	89
W	97.6	100	98.3	100	99	100	99.2	100
OA%	73.1		85.7		80.4		93.7	
KC	0.69		0.83		0.77		0.92	

PA, producer accuracy; UA, user accuracy; OA, overall accuracy; KC, kappa coefficient.

Table 5
Classification results for SAR data obtained with MLC.

Land cover class	PALSAR HH HV		PALSAR HH HV + texture		Land cover class	PALSAR HH HV + texture 7 classes		Land cover class	PALSAR HH HV + texture 6 classes	
	PA%	UA%	PA%	UA%		PA%	UA%		PA%	UA%
AG	43.3	47.9	58.6	57.6	AG	58.0	59.0	AG	57.6	58.9
BS	56.6	69.6	87.0	75.9	BS	75.9	87.0	BS	75.0	87.0
EF	52.3	32.6	43.3	61.3	EF	90.0	49.6	EF + SF + DF	76.0	68.9
FB	42.0	48.6	63.3	48.3	FB	49.6	66. 3	FB	89.0	76.0
OP	72.0	50.9	72.6	86.0	OP	86.3	71.9	OP	53.3	73.9
SF	8.6	38.6	37.6	17.3	SF + DF	4.6	25.9	W	86.3	97.3
DF	43.6	34.6	45.6	60.9	W	91.9	74.3			
W	87.3	95.3	97.3	95.3						
OA%	49.9		62.1		OA%	64.5		OA%	75.7	
KC	0.42		0.56		KC	0.58		KC	0.70	

PA, producer accuracy; UA, user accuracy; OA, overall accuracy; KC, kappa coefficient.

(Table 4). The addition of texture selected variables to each type of image resulted in a noticeable increase in overall accuracy, with AVNIR-2 obtaining 93.7% and TM 85.7%. For TM, the increase in user accuracy was over 10% in all classes except bare soil/urban areas, already well discriminated by bands only (Table 4). The most significant increases with texture use are for oil palm plantation (30%), semi-deciduous forest (17%) and evergreen forest (14%), with respect to user's accuracy. For AVNIR-2 the increase in user accuracy was more than 10% for agriculture (14%), oil palm plantation (25%), secondary (32%) and semi-deciduous forest (15%) classes.

SAR data only: the overall accuracy of the MLC classification with SAR inputs (Table 5) was 49.9% and improved to 62.1% when using mean texture variables derived from both polarizations. Individual classes obtained a user accuracy increase between 6.3% and 35.1%, except in the case of water and farmbush (no change) and secondary forest (decrease due to confusion with other forest classes). According to users accuracies, texture was especially beneficial in detecting evergreen forest (+28.7%), deciduous forest (+26.3%) and oil palm plantation (+35.1%). We obtained an overall accuracy of 64.5% when two forest types were merged, and of 75.7% when all forests were merged in a single class, thus with an increase of over 11% for a six classes land cover task.

Optical and SAR integrated data: SAR channels and their derived textures were concatenated to TM and AVNIR-2 datasets in two steps, first adding the channels only and then the selected textures parameters. The joint use of optical and SAR sensors increased overall accuracies (Table 6). When the TM bands 4–7 and their texture are used for MLC, the overall accuracy is 85.7% (Table 4); the addition of two PALSAR channels increases the accuracy to 89.5% and the further addition of PALSAR texture features increases it to 91.1%. For AVNIR-2, the pattern was similar, as the MLC of bands and texture leads to an overall accuracy of 93.7% (Table 4); the addition of PALSAR bands increases the accuracy to 94.1% and the further addition of PALSAR texture increases it to 95.6%. The use

Table 6
Classification results for joined optical and data obtained with MLC.

Land cover class	TM b4,5,7 + TM texture + PALSAR		TM b4,5,7 + TM texture + PALSAR + PALSAR texture		AVNIR-2 + AVNIR-2 texture + PALSAR		AVNIR-2 + AVNIR-2 texture + PALSAR + PALSAR texture	
	PA%	UA%	PA%	UA%	PA%	UA%	PA%	UA%
AG	91.6	92.9	94.6	93.9	97.4	95.3	97.6	96.6
BS	100.0	100.0	100.0	99.6	97.8	99.4	98.8	99.5
EF	89.3	88.6	90.0	91.0	92.9	87.8	94.6	92.1
FB	93.0	93.3	93.6	95.0	95.8	94.6	96.4	94.8
OP	97.9	96.0	99.6	98.9	96.8	97.1	98.0	98.3
SF	69.6	73.6	70.9	77.6	81.7	89.4	85.7	93.1
DF	78.0	74.3	82.3	75.0	91.2	89.4	94.0	90.5
W	98.6	95.3	99.3	95.3	99.4	100.0	99.8	100.0
OA%	89.5		91.1		94.1		95.6	
KC	0.88		0.89		0.93		0.95	

PA, producer accuracy; UA, user accuracy; OA, overall accuracy; KC, kappa coefficient.
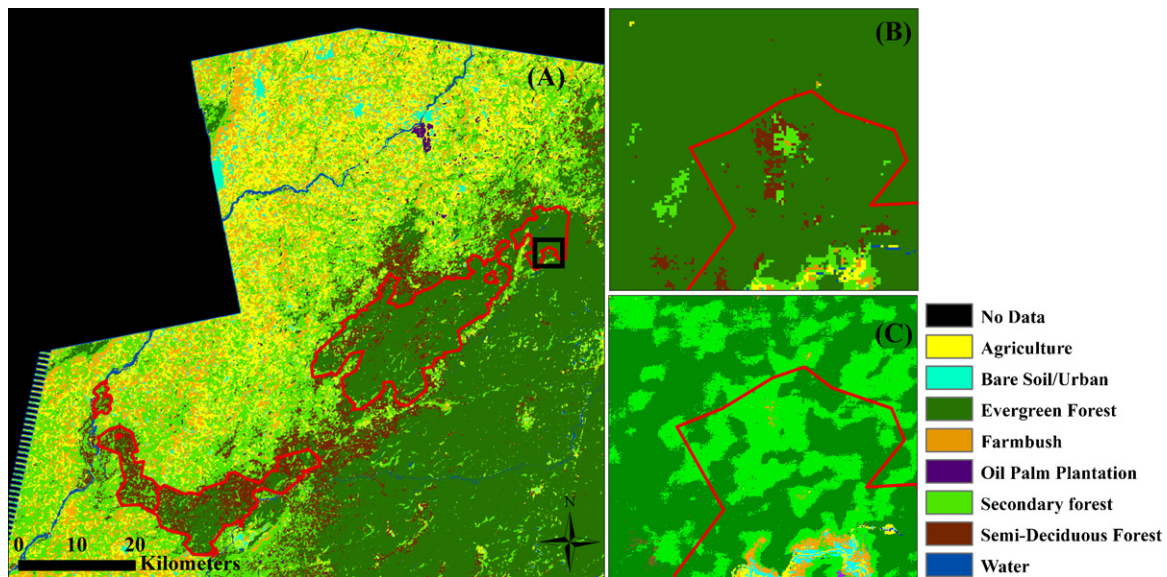
Fig. 4. Land cover classification map obtained with MLC and input 'TM bands 4,5,7 + TM Textures + PALSAR HH HV + PALSAR Textures'. The black box in the main figure (A) delimits a small area of the TM classification map shown in close up in (B); the same area as seen in the classification based on AVNIR-2 (C) (input 'AVNIR-2 + AVNIR-2 textures + SAR + SAR textures'). The Gola Rainforest National Park and Tiwai Island Wildlife Reserve boundaries are shown in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of SAR improved especially the agriculture (+8.6%) and farmbush classes (+12.3%) when added to TM data, as shown by users accuracies; when added to AVNIR-2 data, the increase was lower and distributed in all classes (Table 4 and Table 6). The use of SAR texture slightly changed the user accuracies from −0.4% to 4.3% with major benefits obtained for the semi-deciduous forest, evergreen forest and oil palm plantation classes.

Classification algorithms: the use of NN significantly increased accuracy of the results of optical data in the range 0.7–5.1% with respect to MLC according to *Z* statistics. With combined optical and SAR data, NN significantly increased results accuracy in the range 1.6–1.9% compared to MLC. For radar data as input, the use of NN increased overall accuracies from 1.1% to 3.3%, but the difference was significant only in one case. The increases brought by NN were not observed in any specific class but instead they were evenly distributed across all the land cover classes.

4. Discussion and conclusion

In the present study TM and AVNIR-2 optical datasets were used together with joined dual polarization SAR data to accurately classify a tropical area in West Africa, showing the effectiveness of integrating different data types in this complex region. The MLC classification map using TM near and shortwave infrared bands and SAR inputs is presented in Fig. 4 as the best result obtained with a freely available optical dataset and a widespread classification algorithm.

The distinction of different vegetation types is a task considered difficult due to the smooth transition between successional stages, such as farmbush (characterized by varying amounts of vegetation re-growth) and different forest types (Lucas et al., 2002; Vieira et al., 2003), and for the spectral similarity between evergreen and the semi-deciduous classes, the latter often characterized by a leaf-drop period that could last only a few weeks. Vegetation-oriented classification efforts are especially needed in the region, where a trans-boundary peace park is planned by Liberia and Sierra Leone governments and conservation decisions are frequently taken. The integrated SAR-optical land cover map (Fig. 4) illustrates GRNP effectiveness, with the undisturbed forest classes mainly confined

inside the park and toward Liberia. But it also shows the threat over the Gola South block, which is internally more fragmented than the other units and surrounded by anthropic pressure, and thus deserves special management attention.

We found that AVNIR-2 achieved a better result in classifying the eight land cover types, and especially the forest classes, with respect to spectrally reduced TM (bands 4, 5, 7). AVNIR-2 has a fine spatial resolution which allowed for capturing natural vegetation details (Table 4 and Fig. 4 (B)) and thus better represents environments characterized by mosaics of different vegetation and land cover types. Nevertheless, we still found a low separability of secondary and semi-deciduous classes (Table 4). Semi-deciduous forest mainly occurs in Gola South, where most selective logging took place until 1989. Accurate logging records are not available, but the indications are that these areas are still characterized by signs of old disturbances and thus share traits of secondary forest (Lindsall et al., in review). In our case study, as well as in other tropical environments, the distinction of forest types is useful for conservation while the mapping of the small forest patches outside reserves is also valuable for management, as their change in extension is a tool to assess the value of community forests as carbon sinks and the effectiveness of financial compensation schemes.

While the discrimination of the eight land cover/land use types of the study area was a difficult task for the PALSAR sensor alone, when the three forests classes were merged (thus reducing the task to six classes) the SAR classification accuracy reached 75.7% (Table 5). The SAR based map allows for clear detection of the forested areas (Fig. 5). Confusion between forest and later-stage farmbush and overestimation of the oil palm plantation class is observable, due to similarity of their vertical structure at L-band. In a similar way we observed confusion between early-stage farmbush and agriculture, both having lack of woody components. Anyway, this result goes beyond the usual SAR-based forest/non-forest mapping and the obtained accuracy value is similar or higher than other classifications based on SAR in African environments (Haack and Bechdol, 2000; Herold et al., 2004). As in our case, with both optical images impacted by atmospheric conditions (clouds or haze), is frequent to experience optical data unavailability in tropical regions. The loss of information on forests details, with

Table 7
Comparison of results obtained with ML and NN, network architecture (hidden layers and neurons used) and Z statistics scores.

	Overall accuracy		Hidden layers	Neurons	Z score
	ML	NN			
Landsat TM					
TM 4,5,7 bands	73.1	74.0	2	32	2.16
TM 4,5,7 bands + TM texture	85.7	90.8	2	54	−6.03
TM 4,5,7 bands + TM texture + PALSAR	89.5	91.4	2	32	−2.34
TM 4,5,7 bands + TM texture + PALSAR + PALSAR texture	91.1	92.7	2	54	2.2
AVNIR					
AVN bands	80.4	81.5	2	32	−1.98
AVN bands + texture	93.7	94.4	2	32	−2.16
AVN bands + texture + PALSAR	94.1	95.8	1	24	6.05
AVN bands + texture + PALSAR + PALSAR texture	95.6	97.5	2	32	−5.33
ALOS PALSAR					
HH HV	49.9	51.1	2	18	−0.89*
HH HV + texture	62.1	63.8	2	32	−1.35*
HH HV + texture 7 classes	64.5	67.8	2	32	2.54
HH HV + texture 6 classes	75.7	78.1	2	32	−1.89*

The asterisk indicates results which are not significant at the 5% confidence level.

respect to classification based on integrated optical and SAR data, was counterbalanced by full SAR coverage of the area. The use of SAR is suggested as a strategy to cope with persistent cloud cover affecting tropical regions: in this study we showed that SAR alone can still provide important landscape information. Furthermore, it is the SAR combination with optical data – TM or AVNIR-2 – that always produces the best outputs (Table 6). This result confirms SAR role in forest and vegetation mapping of tropical regions, and suggests that in areas affected by optical data loss different optical resolutions can be exploited and joined with SAR. If available, finer resolution sensors such as AVNIR-2 can help to detail specific sub-areas of interest, while larger areas or zones covered by clouds can be filled with SAR or optical lower resolution data.

With regards to the techniques here tested here to improve the mapping accuracy, we found that the addition of the optical texture features was very effective in improving the classification results in all cases, with an increase between 10.1% and 13.2%. The result shows that the reflectance differences and the shadow effects occurring in different forest types can be partially captured by optical texture measures. The addition of SAR texture variables increased the accuracies significantly in SAR-only tests, similarly to what has been found by Longepe et al. (2011) in tropical forests, again confirming the value of textural information when this data type is used singularly, and especially for discriminating classes of dense and tall vegetation (Table 5). On the other hand, SAR texture was much less effective when included as the last input to optical,

optical texture and SAR datasets (Table 7), perhaps due to the fact that very high values of accuracy were already reached and thus the margins of improvement were limited.

The use of the NN algorithm brought in an improvement in the results with respect to MLC. This improvement is limited when SAR and optical data are integrated (1.6–1.9%) and not significant in most classifications based on SAR data only (Table 7). The decision to adopt machine learning should take into consideration the need for expert knowledge, evaluating its cost against the amount of improvement obtained in classification. In this respect we found that the NN accuracy increase was beneficial but limited in amount, partially due to the high values already obtained by MLC, thus suggesting the machine learning adoption as an added tool only when resources allows for it.

A key constraint for implementing monitoring systems at a fine scale is the availability of finer resolution data (Achard et al., 2010). In this view, this research showed the improvement observed with the spatial resolution of AVNIR-2 sensors and its successful integration with a coarser SAR data. The atmospheric conditions in the moist tropical regions pose another serious challenge for data availability. Until new sensors are available, with increased revisiting capabilities, it is fundamental to exploit and integrate the few available data to accurately classify the landscape of those remote areas. This research tested a possible approach in a complex environment, and similar efforts can be easily replicated in the region to first map and then monitor the growing number of conservation units.

Techniques to mosaic multi-date images at a pixel level are an option for helping to solve the problem of cloud cover (Roy et al., 2010). The use of multi-temporal SAR datasets is also known to improve accuracy of results (Ranson and Sun, 1994; Chen et al., 1996) and another promising approach is the integration of microwave frequencies (e.g. at L- and C-band; Lardeux et al., 2011), which should be able to provide information both on foliage strata, helpful in the distinction of different canopies structure and vegetation water content, and on forest volumes, which are related to disturbance and age. While waiting for new data from future generation optical and SAR space missions, the adoption of these existing techniques and datasets could offer improved mapping capabilities and will represent our future research step for tropical complex regions.
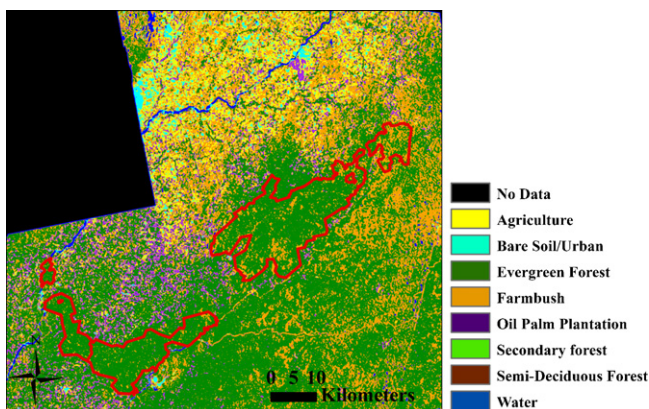
Fig. 5. Land cover classification map obtained with MLC and input 'ALOS PALSAR HH HV + texture 6 classes. The Gola Rainforest National Park and Tiwai Island Wildlife Reserve boundaries are shown in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Acknowledgements

References

Achard, F., Stibig, H-J., Eva, H.D., Lindquist, E.J., Bouvet, A., Arino, O., Mayaux, P., 2010. Estimating tropical deforestation from Earth observation data. Carbon Management 1 (2), 271–287.

Asner, G.P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. Remote Sensing of Environment 64 (3), 234–253.

Asner, G.P., 2001. Cloud cover in Landsat observations of the Brazilian Amazon. International Journal of Remote Sensing 22 (18), 3855–3862.

Atkinson, P.M., Tatnall, A.R.L., 1997. Neural networks in remote sensing. International Journal of Remote Sensing 18 (4), 699–709.

Avitabile, V., Baccini, A., Friedl, M.A., Schmullius, C., 2012. Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. Remote Sensing of Environment 117, 366–380.

Bayol, N., Chevalier, J.F., 2004. Current state of the forest cover in Liberia – forest information critical to decision making, Forest Resources Management. Study report for the World Bank, 25 June 2004.

Berberoglu, S., Lloyd, C.D., Atkinson, P.M., Curran, P.J., 2007. Texture classification of Mediterranean land cover. International Journal of Applied Earth Observation and Geoinformation 9 (3), 322–334.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, Oxford.

CEPF, 2003. Critical Ecosystem Partnership Fund. Guinean Forests of West Africa Hotspot Upper Guinean Forest Briefing Book. Prepared for: Improving Linkages Between CEPF and World Bank Operations, Africa Forum, Cape Town, South Africa, April 25–27, 2005.

Chan, J.C.-W., Laporte, N., Defries, R.S., 2003. Texture classification of logged forests in tropical Africa using machine learning algorithms. International Journal of Remote Sensing 24 (6), 1401–1407.

Chen, K.S., Huang, W.P., Tsay, D.H., Amar, F., 1996. Classification of multi-frequency polarimetric SAR imagery using a dynamic learning neural network. IEEE Transactions on Geosciences and Remote Sensing 34 (3), 814–820.

Chica-Olmo, M., Abarca-Hernandez, F., 2000. Computing geostatistical image texture for remotely sensed data classification. Computers and Geosciences 26 (4), 373–383.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37–46.

Cole, N.H.A., 1993. Floristic association in the Gola rain forests: a proposed biosphere reserve. Journal of Pure and Applied Science 2, 35–50.

Congalton, R., Green, K., 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. CRC/Lewis Press, Boca Raton, FL.

Del Frate, F., Solimini, D., 2004. On a neural network algorithm for retrieving forest biomass from SAR data. IEEE Transactions on Geoscience and Remote Sensing 42, 24–34.

Dekker, R.J., 2003. Texture analysis and classification of ERS SAR images for Map updating of urban areas in the Netherland. IEEE Transactions on Geoscience and Remote Sensing 41 (9), 1950–1958.

Dorigo, W., Lucieer, A., Podobnikar, T., Carni, A., 2012. Mapping invasive Fallopia japonica by combined spectral, spatial, and temporal analysis of digital orthophotos. International Journal of Applied Earth Observation and Geoinformation 19, 185–195.

Erasmi, S., Twele, A., 2009. Regional land cover mapping in the humid tropics using combined optical and SAR satellite data: a case study from Central Sulawesi, Indonesia. International Journal of Remote Sensing 30 (10), 2465–2478.

Felde, G.W., Anderson, G.P., Cooley, T.W., Matthew, M.W., Adler-Golden, S.M., Berk, A., Lee, J., 2003. Analysis of Hyperion Data with the FLAASH Atmospheric Correction Algorithm. Proceedings of the International Geoscience and Remote Sensing Symposium, Toulouse, France, (IGARSS'03), 90–92.

Gibbs, H.K., Brown, S., Nilesand, J.O., Foley, J.A., 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. Environmental Research Letters 2 (4), 1–13.

Haack, B., Bechdol, M., 2000. Integrating multisensor data and radar texture measures for land cover mapping. Computers and Geosciences 26 (4), 411–421.

Haralick, R.M., Shanmugan, K., Dinstein, I., 1973. Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics 3 (6), 610–621.

Herold, N.D., Haack, B.N., Solomon, E., 2004. An evaluation of radar texture for land use/cover extraction in varied landscapes. International Journal of Applied Earth Observation and Geoinformation 5 (2), 113–128.

Hill, M.O., 1979. TWINSPAN – A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. Ecology and Systematics. Cornell University, Ithaca, NY.

Hu, J., Chen, W., Li, X., He, X., 2009. A haze removal module for multispectral satellite imagery. Urban Remote Sensing Joint Event, Shanghai, China, (JURSE 2009), p. 4.

Igue, A.M., Houndgba, C.J., Gaiser, T., Stahr, K., 2006. Land Use/Cover Map and its Accuracy in the Oueme Basin of Benin (West Africa). Conference on International Agricultural Research for Development, Bonn, Germany, (Tropentag), p. 4.

Judex, M., Thamm, H. P., Menz, G., 2006. Improving land cover classification with knowledge based approach and ancillary data. Proceedings of the 2nd Workshop of the EARSeL SIG on Land Use and Land Cover.

Klop, E., Lindsell, J., Siaka, A., 2008. Biodiversity of Gola Forest, Sierra Leone. Royal Society for the Protection of Birds, Conservation Society of Sierra Leone, Government of Sierra Leone.

Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. Journal of Artificial Intelligence 97 (1–2), 273–324.

Kuplich, T.M., 2006. Classifying regenerating forest stages in Amazonia using remotely sensed images and a neural network. Forest Ecology and Management 234 (1–3), 1–9.

Lardeux, C., Frison, P.-L., Tison, C., Souyris, J.-C., Stoll, B., Fruneau, B., Rudant, J.-P., 2011. Classification of tropical vegetation using multifrequency partial SAR polarimetry. IEEE Geoscience and Remote Sensing Letters 8, 133–137.

Li, G., Lu, D., Moran, E.F., Hetrick, S., 2011. Land-cover classification in a moist tropical region of Brazil with Landsat Thematic Mapper imagery. International Journal of Remote Sensing 32 (23), 8207–8230.

Licciardi, G., Pacifici, F., Tuia, D., Prasad, S., West, T., Giacco, F., Thiel, C., Inglada, J., Christophe, E., Chanussot, J., Gamba, P., 2009. Decision fusion for the classification of hyperspectral data: outcome of the 2008 GRS-S data fusion contest. IEEE Transaction on Geoscience and Remote Sensing 47 (11), 3857–3865.

Lefsky, M.A., Cohen, W.B., 2003. Selection of remotely sensed data. In: Wulder, M.A., Franklin, S.E. (Eds.), Remote Sensing of Forest Environments: Concepts and Case Studies. Kluwer Academic Publishers, Boston, pp. 13–46.

Lehmann, E.A., Caccetta, P.A., Zhou, Z., McNeill, S.J., Wu, X., Mitchell, A.L., 2012. Joint Processing of Landsat and ALOS-PALSAR Data for Forest Mapping and Monitoring. IEEE Transactions on Geoscience and Remote Sensing 50, 55–67.

Longepe, N., Rakwatin, P., Isoguchi, O., Shimada, M., Uryu, Y., Yulianto, K., 2011. Assessment of ALOS PALSAR 50 m Orthorectified FBD Data for Regional Land Cover Classification by Support Vector Machines. IEEE Transactions on Geoscience and Remote Sensing 49 (6), 2135–2150.

Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing 28 (5), 823–870.

Lu, D., Batistella, M., Moran, E., 2007. Land-cover classification in the Brazilian Amazon with the integration of Landsat ETM+ and RADARSAT data. International Journal of Remote Sensing 28 (24), 5447–5459.

Lucas, R.M., Honzák, M., Amaral, I.D., Curran, P.J., Foody, G.M., 2002. Forest regeneration on abandoned clearances in central Amazonia. International Journal of Remote Sensing 23 (5), 965–988.

Mitchard, E.T.A., Saatchi, S.S., White, L.J.T., Abernethy, K.A., Jeffery, K.J., Lewis, S.L., Collins, M., Lefsky, M.A., Leal, M.E., Woodhouse, I.H., Meir, P., 2011. Mapping tropical forest biomass with radar and spaceborne LiDAR: overcoming problems of high biomass and persistent cloud. Biogeosciences 8 (4), 8781–8815.

Moller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 6 (4), 525–533.

Pacifici, F., Del Frate, F., Emery, W.J., Gamba, P., Chanussot, J., 2002. Urban mapping using coarse SAR and optical data: outcome of the 2007 GRS-S data fusion contest. IEEE Geoscience and Remote Sensing Letters 5 (3), 331–335.

Price, K., Guo, X., Stiles, J.M., 2002. Optimal Landsat TM band and vegetation indices for discrimination of six grassland types in eastern Kansas. International Journal of Remote Sensing 23 (23), 5031–5042.

Rahman, M.M., Sumantyo, J.T.S., 2010. Mapping tropical forest cover and deforestation using synthetic aperture radar (SAR) images. Applied Geomatics 2 (4), 113–121.

Ranson, K.J., Sun, G., 1994. Northern forest classification using temporal multifrequency and multipolarimetric SAR images. Remote Sensing of Environment 47 (2), 142–153.

Richards, J.A., Jia, X., 1999. Remote Sensing Digital Imaging Analysis: an Introduction, third ed. Springer, Berlin.

Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M., Loveland, T.R., Vermote, V., Zhang, C., 2010. Webenabled Landsat data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. Remote Sensing of Environment 114, 35–49.

Santos, J.R., Neeff, T., Dutra, L.V., Araujo, L.S., Gama, F.F., Elmiro, M.A.T., 2004. Tropical forest biomass mapping from dual frequency SAR interferometry (X and P-Bands). Twentieth International Society for Photogrametry and Remote Sensing (ISPRS) Congress: GeoImagery Bridging Continents, Istanbul, v. XXXV, pp. 1133–1136.

Shearman, P.H., 2009. An Assessment of Liberian Forest Area, Dynamics, FDA Concessions Plans, and their Relevance to Revenue Projections, Rights and Resources Initiative.

Shimada, M., Isoguchi, O., Tadono, T., Isono, K., 2009. PALSAR radiometric and geometric calibration. IEEE Transactions on Geosciences and Remote Sensing 47 (12), 3915–3932.

Simard, M., Saatchi, S., De Grandi, G.F., 2000. The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest. IEEE Transactions on Geoscience and Remote Sensing 38 (5), 2310–2321.

Tachikawa, T., Kaku, M., Iwasaki, A., Gesch, D., Oimoen, M., Zhang, Z., Danielson, J., Krieger, T., Curtis, B., Haase, J., Abrams, M., Crippen, R., Carabajal, C., 2011. ASTER

Global Digital Elevation Model Version 2–Summary of Validation Results. ASTER GDEM Validation Team: METI/ERSDAC NASA/LPDAAC USGS/EROS.

Talavera, L., 2005. An evaluation of filter and wrapper methods for feature selection in categorical clustering. 6th International Symposium on Intelligent Data Analysis (IDA05), pp. 440–451.

Verschuren, J., 1983. Conservation of Tropical Rain-forest in Liberia: Recommendations for Wildlife Conservation and National Parks. Report prepared for IUCN, Gland, Switzerland.

Vieira, I.C.G., Almeida, A.S., Davidson, E.A., Stone, T.A., Carvalho, C.J.R., Guerrero, J.B., 2003. Classifying successional forests using Landsat spectral properties and ecological characteristics in eastern Amazônia. Remote Sensing of Environment 87 (4), 470–481.

Wijaya, A., Marpu, P.R., Gloaguen, R., 2008. Geostatistical texture classification of tropical rainforest in Indonesia. In: Alfred Stein, J.S., Bijker, W. (Eds.), Quality Aspect in Spatial Data Mining. CRC Press Inc., pp. 199–210.

World Bank, 2011. Forest Resource Assessments in Liberia. Report on field verification. River Cess County, Liberia. Michael Abedilartey.

Yavitt, J.B., Battles, J.J., Lang, G.E., Knight, D.H., 1995. The canopy gap regime in a secondary Neotropical forest in Panama. Journal of Tropical Ecology 11 (3), 391–402.

Zhang, Y., Guindon, B., Cihlar, J., 2002. An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images? Remote Sensing of Environment 82 (2–3), 173–187.

Zhang, Y., Guindon, B., 2003. Quantitative assessment of a haze suppression methodology for satellite imagery: Effect on land cover classification performance. IEEE Transactions on Geoscience and Remote Sensing 41 (5), 1082–1089.

Zhu, Z., Woodcock, C.E., Rogan, J., Kellndorfer, J., 2012. Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data. Remote sensing of Environment 117, 72–82.

Chapter 6

Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data

Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data

Gaia Vaglio Laurin[a,f]*, Qi Chen[b], Jeremy A. Lindsell[c], David A. Coomes[d], Fabio Del Frate[f], Leila Guerriero[f], Francesco Pirotti[g], Riccardo Valentini[e,a]

[a]CMCC - Centro Mediterraneo per i Cambiamenti Climatici, via Augusto Imperatore (Euro-Mediterranean Center for Climate Change), IAFENT Division, via Pacinotti 5, Viterbo 01100, Italy

[b]Department of Geography, University of Hawai`i at Mānoa, 422 Saunders Hall, 2424 Maile Way, Honolulu, HI, 96822, USA

[c]The Royal Society for the Protection of Birds, The Lodge, Sandy, Beds. SG19 2DL, UK.

[d]Forest Ecology and Conservation Group, Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK.

[e]Department of Forest Resources and Environment, University of Tuscia, Viterbo I-01100 Italy

[f]Tor Vergata University, Department of Civil Engineering and Computer Sciences, Via del Politecnico 1, 00133 Rome, Italy

1

[g]CIRGEO – Interdepartmental Research Center of Geomatics, University of Padova, Via dell'Università 16, 35020 Legnaro, Italy

*Corresponding author: Email: gaia.vagliolaurin@cmcc.it, laurin@disp.uniroma2.it; Phone/Fax: 0039 06 72597710

Abstract:

The estimation of above ground biomass in forests is critical for carbon cycle modeling and climate change mitigation programs. Small footprint lidar provides accurate biomass estimates, but its application in tropical forests has been limited, particularly in Africa. Hyperspectral data record canopy spectral information that is potentially related to forest biomass. To assess lidar ability to retrieve biomass in an African forest and the usefulness of including hyperspectral information, we modeled biomass using small footprint lidar metrics as well as airborne hyperspectral bands and derived vegetation indexes. Partial Least Square regression (PLSR) was adopted to cope with multiple inputs and multicollinearity issues; the Variable of Importance in the Projection was calculated to evaluate importance of individual predictors for biomass. Our findings showed that the integration of hyperspectral bands ($R^2 = 0.70$) improved the model based on lidar alone ($R^2 = 0.64$); this encouraging result call for additional research to clarify the possible role of hyperspectral data in tropical regions. Replacing the hyperspectral bands with vegetation indexes resulted in a smaller improvement ($R^2 = 0.67$). Hyperspectral bands had limited predictive power ($R^2 = 0.36$) when used alone. This analysis proves the efficiency of using PLSR with small-footprint lidar and high resolution hyperspectral data in tropical forests for biomass estimation. Results also suggest that high quality ground truth data is crucial for lidar-based AGB estimates in tropical African forests, especially if airborne lidar is used as an intermediate step of upscaling field-measured AGB to a larger area.

1. Introduction

Remote sensing of forest aboveground biomass (AGB) has received increasing attention during the last decade due to its relevance to global carbon cycle modeling and to international programs aimed at reducing greenhouse gas emissions in tropical areas, such as the United Nations Reducing Emissions from Deforestation and Forest Degradation (REDD+). In particular, biomass mapping in tropical biomes is particularly important given the critical role of tropical forests in the global carbon cycle (Gibbs et al., 2007). Recent findings show that tropical forests store 21% more carbon than previously expected (Baccini et al., 2012).While the biomass of most temperate and boreal zones has been systematically inventoried at least once (Houghton et al., 2009), tropical regions suffer from operational limitations and consequent lack of data, which is especially marked in Africa (Baccini et al. 2009).

Airborne small-footprint Light Detection and Ranging (lidar) is considered the most accurate remote sensing technology for mapping biomass (Zolkos et al, 2013) and could be useful in filling this information gap. Discrete return (DRL) or full waveform (FWL) small-footprint lidar systems are now widespread and operated around the globe, enabling the collection of up to four returning energy pulses (as DRL) or all the returning energy (as FWL) from the forest vertical profile. The laser pulse returns are usually used to derive forest height metrics, which can then be related to field-observed AGB, with the latter obtained by means of field measures and allometric relationships. Due to the high operational costs, lidar-derived AGB estimates usually can only be obtained over limited areas. These local-scale or sub-national accurate estimates are crucial for REDD+ measuring, reporting, and verification (MRV), and for country level natural resources management and inventories (Naesset 2007; Peterson et al. 2007). Local AGB maps are also the basis for the extension of estimates to larger areas using remote sensing approaches

(Asner et al. 2010; De Sy 2012). However, to date there has been little research into mapping of biomass in tropical forests using airborne small-footprint lidar. Zolkos et al. (2013) conducted a comprehensive review and identified eight studies carried out with this system in tropical forests, with none in continental Africa.

The uncertainties associated with the current knowledge of the African ecosystems' carbon balance are rather high. A review of the most recent estimates of the net long-term carbon balance of African ecosystems, based upon observations, indicated a sink of the order of 0.3 Pg Cyr^{-1} with a very high uncertainty and a variable source; up to now many questions remain open, and it is unclear whether Africa is a net carbon source or a sink to the atmosphere (Ciais et al. 2011). Because of highly variable CO_2 fluxes and insufficiently studied ecosystems and ecosystem–human–climate interactions, there is a need for continued and enhanced observations of carbon stocks, fluxes and atmospheric concentrations to enable more precise assessments of Africa's carbon cycle (Justice et al. 2001), and its sensitivity to natural and anthropogenic pressures and future climate. Of primary importance is the need for continent-wide carbon cycle observations that support both bottom-up and top-down methods of estimating carbon sources and sinks (Lewis et al. 2011). An African integrated carbon-observing system is needed, encompassing both: (i) regional inventories and monitoring of soil and vegetation carbon stocks by forest and agricultural research stations; (ii) remote sensing-based estimates of forest biomass C stock distribution, at different scales and using active and/or passive sensors combined with field observations. In view of the above considerations regarding the contributions of African forests in the global carbon cycle, it is

clear how valuable it is to test biomass mapping by means of various sensors over different African forests.

Hyperspectral sensors, recording the reflectance of a large number of fine resolution spectral bands from visible to near infrared (NIR) or shortwave infrared (SWIR) range, are another frontier technology in remote sensing. Hyperspectral data can capture information regarding the biochemical composition of the upper canopy layer and have been used for forest type or species classification, estimation of biophysical and biochemical properties and health status (Asner and Martin, 2008; Koch, 2010; Goodenough et al., 2006).The ecosystem information recorded by hyperspectral data may relate to plant functional types –such as whether a species is light demanding -which could in turn affect wood density and thus biomass content (Baker et al., 2004; Chave et al. 2009). Hyperspectral data have been used to estimate grassland biomass directly (Cho et al. 2007; Psomas et al, 2011) and leaf canopy biomass (le Maire et al. 2008), while leaf area density, retrieved from fusion of hyperspectral and radar data, has been used in the estimation of forest AGB (Treuhaft et al. 2003). The few studies that have attempted to improve biomass estimates in boreal, temperate and tropical forests by combining hyperspectral imagery with lidar data have reported only modest or no improvement in model fit compared to the results from using lidar only (Anderson et al. 2008; Clark et al. 2012; Latifi et al. 2012; Swatantran et al. 2011). Despite these research efforts, the number of published studies on integrating lidar and hyperspectral data for biomass estimation is rather small. Further research is needed along this line, especially considering the opportunities from forthcoming hyperspectral missions, such as the Environmental Mapping and Analysis Program (EnMap), the PRecursore

IperSpettrale of the application mission (PRISMA), the Medium Resolution Imaging

Spectrometer (MERIS) and the Hyperspectral Infrared Imager (HyspIRI).

The main objectives of the present study are the following: (i) to test for the first time small

footprint lidar-based AGB retrieval in a West African tropical moist forest, (ii) to examine

whether the use of very high spatial resolution hyperspectral data in addition to lidar can improve

the biomass estimates.

2. Materials and methods

2.1 Study area and ground truth data

The study area is within the Gola Rainforest National Park (GRNP) in Sierra Leone, at the

westernmost end of the humid Upper Guinean Forest Belt, in West Africa (Fig.1).

INSERT FIGURE 1

The forests of this region are largely lowland moist evergreen forest with some areas

towards lowland dry evergreen and semi-deciduous forest types (Cole 1993). Within GRNP

Klop et al. (2008) identified moist evergreen, moist semi-deciduous, freshwater inland swamp

forest, forest regrowth and secondary/disturbed forest. The GRNP area has been protected

through conservation programs since 1989 but commercial logging, most intensively in the

southern block, was carried out in 1963–1965 and 1975–1989. Recent land cover mapping

(Vaglio Laurin et al., 2013) highlighted the importance of conserving this forest from

anthropogenic pressure in the surrounding areas. The climate is moist tropical, with annual rainfall around 2500–3000 mm, a dry season from November to April coincident with leaf-off condition of some semi-deciduous tree species, and an altitude of 70–410 m.

Field data collection carried out in 2006-2007 in the GRNP established over 600 0.125 ha circular plots across the whole park area, recording species information as well as structural and environmental forest parameters. In the plots, all trees with Diameter at Breast Height (DBH) > 30 cm were recorded, while trees with DBH included in the 10-30 cm range were measured in a 1/10 smaller subplot. Height measures were derived with a local DBH-height relationship and the AGB was obtained applying the Chave et al. (2005) general equation for moist tropical forest including DBH, height and wood density values. The data collection protocol and the allometric procedure are fully documented in Lindsell and Klop (2013). We selected all the plots surveyed by both lidar and hyperspectral sensors excluding some plots located less than 1 km from the park boundary where land cover changes were most likely to have occurred in the period between field and aerial data collection. We also excluded plots affected by cloud shadow in the hyperspectral data. We retained 70 ground truth plots, with an AGB range 0-586.9 Mg ha^{-1}(mean = 172.2 and standard deviation = 111.8 Mg ha^{-1}). These plots contained 136 species with DBH> 30 cm, and 86 occurring in the upper canopy layer.

2.2 Remote sensing data

The central and parts of the southern blocks of GRPN were surveyed by an airborne campaign in March 2012 over pre-defined flight lines covering part of the field, using a Pilatus PC-6 Porter aircraft equipped with lidar and hyperspectral sensors and a digital camera for aerial photographs.

INSERT FIGURE 2

The lidar sensor ALTM GEMINI (Optech Ltd), characterized by a 1064 nm laser wavelength and able to record up to 4 range measurements, was operated between 650-850m above ground level (AGL). The minimum laser density was set to 11 points/m^2. The lidar dataset was delivered as a point cloud of discrete returns, preprocessed in Terrascan (Terrasolid) software and adopted the ApplanixIN-Fusion^{TM}PPP Inertially-Aided Precise Point Positioning (IAPPP) to cope with the of absence of GPS base stations in the region. Positional error in x, y, z was always lower than 0.27 m for any axis. An additional check with points derived from the AUSPOS network of Geoscience Australia indicated a positional error lower than 0.2 m. The raw all-returns point cloud was processed using the Toolbox for Lidar Data Filtering and Forest Studies (TIFFS) (Chen, 2007) to derive a range of metrics for AGB estimation from each plot, including: mean height, quadratic mean height, skewness, kurtosis, height bins at 5 m intervals and 10% percentile heights. TIFFS used the ground returns identified by the data provider to generate a DTM (Digital Terrain Model) and calculated the relative height above terrain of each laser return by subtracting the corresponding DTM elevation from its original Z value. The lidar metrics were derived using the relative height of all laser points.

Hyperspectral data were acquired in 18 strips with an AISA Eagle sensor, with FOV equal to 39.7°, set to record 244 bands with 2.3 nm spectral resolution in the 400-970 nm range. The final spatial resolution was at 1 m after radiometric correction and orthorectification based on a lidar-derived Digital Elevation Model (DEM). A visual inspection of data from 30 randomly selected plots revealed a spatial mismatching between hyperspectral and lidar data within a range of 1 - 4 m.

Atmospheric correction of the hyperspectral images was performed using the Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm, which is based on a MODTRAN4 approach for path scattered radiance, absorption, and adjacency effects (Felde et al., 2003). Due to noise, all the bands outside the 450-900 nm range and four bands in the 759-766 nm range were removed, reducing the total number of bands to 186. Minimum Noise Fraction (MNF) transformation (Green et al. 1988) was used to further reduce noise in the dataset. For each image strip, 9 to 15 MNF components were selected by visual screening and used to compute the inverse MNF to transform back the bands in the original data space. Different strips have different noise levels and types: generally the useful information is included in the first 15 components, but this is a rule of thumb. Visual screening allowed to identify the correct number of components to be used for analysis (Williams and Hunt 2002; Underwood et al. 2003; Goodwin et al. 2005).

Eight vegetation indices (VIs) were calculated from the inverted MNF bands (Table 1): Normalized Difference Vegetation (NDVI) and Simple Ratio (SRI) (Sellers 1985), Atmospherically Resistant Vegetation (ARVI) (Kaufman and Tanre 1996), Red Edge Normalized Difference Vegetation (ReNDVI) (Sims and Gamon 2002), Vogelmann Red Edge

(VReI) (Vogelmann et al. 1993), Photochemical Reflectance (PRI) (Gamon et al. 1992), Red

Green Ratio (GRI) (Gamon and Surfus 1999), and Anthocyanin Reflectance 2 (AR2I) (Gitelson

et al. 2001). These indices were chosen for representing information from different portions of

the spectra of vegetation greenness, light use efficiency and leaf pigments and for being

relatively insensitive to shadow. For each plot we averaged the VIs and the 186 hyperspectral

bands after MNF inversion. Table 1 summarizes the lidar and hyperspectral inputs used in tests.

Aerial photographs were acquired simultaneously with lidar data using a Rollei H25 camera

equipped with a Phase One Digital Back. Images were georeferenced and orthorectified using the

lidar DEM. The orthophotos were acquired at 0.1 m spatial resolution, and used as reference for

visual screening during data analysis (i.e. to visualize plot edge effects).

Table 1. Description of remote sensing statistics used in biomass regression analyses

Input	Description
Lidar height metrics	-Mean of all plot returns -Standard deviation -Quadratic Mean -Skewness -Kurtosis -Proportion of points at height bins of 5m intervals -10% Percentiles from 10% to 100%
Hyperspectral bands	- 186 bands in the 450-900 nm interval, atmospherically corrected and noise minimized
Vegetation Indices	- Normalized Difference Vegetation Index (Sellers 1985) - Simple Ratio Index (Sellers 1985) - Atmospherically Resistant Vegetation Index (Kaufman and Tanre 1996) - Red Edge Normalized Difference Vegetation Index (Sims and Gamon 2002) - Vogelmann Red Edge Index (Vogelmann et al. 1993) - Photochemical Reflectance Index (Gamon et al. 1992) - Red Green Ratio (Gamon and Surfus 1999) - Anthocyanin Reflectance Index 2 (Gitelson et al. 2001)

2.3 Retrieval models and tests

The large number of often correlated metrics from airborne lidar and hyperspectral data pose challenges in statistical modeling of biomass due to the problems of multicollinearity and "curse of dimensionality" (Adam and Mutanga 2009; Dalponte et al. 2009). We used Partial Least Squares (PLS) regression to deal with these issues. PLS regression is closely related to principal component regression (PCR), but differs in that it uses the information from the response variable in addition to the predictors for feature transformation (Geladi and Kowalski 1986). PLS regression has been previously employed in spectral and chemical analysis of tropical forests (Asner and Martin 2008), for AGB estimation (Lei et al. 2012; Goodenough et al. 2005), and as a method for dealing with large hyperspectral datasets (Peerbay et al. 2013).

We modeled AGB from single and fused lidar and hyperspectral data, to understand the ability of our dataset to estimate AGB in an African rainforest, and assess the usefulness of these data integration. For hyperspectral data we tested both MNF-inverted bands and the derived VIs. The PLS regression results were compared with those obtained by a multiplicative power model (MPM), well suited to explain the usual power-law relationship occurring among biological parameters (Marquet et al. 2005). Inputs for both models were log transformed.

To develop the MPM, a forward stepwise regression of the log-transformed predictors and the AGB values was used to select the predictors; the initial model is then fitted using such predictors. Any of the selected predictors which were not significant from their p-value were

12

removed (p > 0.05) and the model is refitted; the procedure was iterated until all predictors are statistically significant. For PLS regression, the transformed features were selected by minimizing the 10-fold cross-validation prediction error. A traditional method like MPM or other common statistical techniques is often used as a benchmark in literature for AGB estimation (Chen et al. 2012) or vegetation type discrimination (Vaglio Laurin et al. 2013). Comparison between MPM and PLS regression is useful to illustrate the accuracy improvement.

We calculated the Variable of Importance in the Projection (VIP) to evaluate importance of individual predictors for biomass estimation; predictors with VIP scores > 1 are considered especially relevant for the mode l (Wold et al. 2001; Peerbhay et al. 2013).

3. Results

Based on lidar metrics alone, AGB was predicted with a coefficient of determination (R^2) equal to 0.64 and a RMSE of 67.8 Mg ha^{-1} using PLS; results obtained by MPM were less accurate ($R^2 = 0.57$). Hyperspectral bands had limited predictive power using PLS ($R^2 = 0.36$), and none with MPM. The VIs had very limited predictive power when entered into the models. Using PLS the addition of hyperspectral bands to lidar metrics increased the accuracy moderately ($R^2 = 0.70$, RMSE 61.7 Mg ha^{-1}), whilst replacing the hyperspectral bands with the VIs resulted in an even smaller improvement ($R^2 = 0.67$, RMSE 64.3 Mg ha^{-1}). No improvement of accuracy is obtained using MPM with combined lidar and hyperspectral dataset. In comparison to MPM, PLS produced improved accuracies in all models, except VIs alone. The AIC (Akaike's Information Criteria) was also calculated to compare different PLS models (Chen

13

et al. 2007). In general, compared to the model with the lowest AIC value, the models with an

AIC increase of 4-7 have considerable less support and the ones with an AIC increase of >10

have no support (Burnham and Anderson, 2002). Among our PLS models, the combination of

lidar metrics and hyperspectral bands has the lowest AIC value of 597 and thus the best

performance. The PLS model based on lidar metrics has an AIC value of 606, which corresponds

to an increment of 9 and indicates that such a model is at least considerably worse than the model

using both lidar metrics and hyperspectral bands. Table 2 illustrates the test results.

Table 2. Test results obtained with different combinations of lidar metrics and hyperspectral features and through two different statistical models.

Inputs	Multiplicative Power Model (MPM)		Partial Least Square Regression (PLS)		
	R2	*RMSE*	*R2*	*RMSE*	*AIC*
Lidar metrics	*0.57*	*72.7*	*0.64*	*67.8*	*606*
Hyperspectral bands	*0.00*	*111.0*	*0.36*	*91.2*	*646*
VIs	*0.08*	*106.2*	*0.02*	*116.8*	*668*
Lidar metrics + Hyperspectral bands	*0.57*	*72.7*	*0.70*	*61.7*	*597*
Lidar metrics + VIs	*0.57*	*72.7*	*0.67*	*64.3*	*601*

The scatterplots of the predicted vs. field observed AGB for different input combinations are

presented in Fig. 3.

INSERT FIGURE 3

Among lidar metrics, the inputs obtaining VIP scores >1 included all percentiles (except the 10th and the 100th), some low range height bins, mean height and quadratic mean height. Highest scores were obtained, in descending order, by the 40th height percentiles, 30th height percentiles, mean height, 50th and 60th height percentiles. Among hyperspectral inputs, the higher scores were assigned to bands in the green, and red-edge region of the spectra, and in the near infrared region close to the end of the available spectra. When using the combined dataset, all lidar metrics received scores >1 and greater than the hyperspectral bands. Fig.4 illustrates the most relevant input features selected by VIP procedure for the models based on single lidar and hyperspectral datasets.

INSERT FIGURE 4

To help understand the selection of relevant input feature by VIP for the lidar-based AGB model with respect to the forest structure, we graphically explored the distribution of trees and biomass in 7 classes of height, at 10 m intervals each (Fig. 5).

INSERT FIGURE 5

4. Discussion

4.1 Comparison to other studies of mapping tropical rainforest biomass

Our first aim was to test small footprint lidar for AGB estimation of an African tropical moist forest. The accuracy of our estimate is in within the range of those reported in other tropical studies that use small-footprint lidar (Asner et al. 2009; Asner et al. 2010; Asner et al. 2012b; Clark et al. 2011; d'Oliveira et al. 2012; Kennaway et al. 2008; Kronseder et al. 2012; Mascaro et al. 2011a). As far as we know, there are only two studies reporting usage of airborne lidar for mapping African tropical forests AGB (Asner et al. 2011, 2012a), both undertaken in Madagascar with a customized Optech 3100EA instrument (Carnegie Institution for Science, USA). The accuracy obtained in the current study is not far from that obtained in southeastern Madagascar (Asner et al. 2011), where 46 plots of 0.28 ha were used. After adopting improved allometric relationships to reflect regional variations in Madagascar, the authors reported a $R^2 =$ 0.68 for their AGB estimate. However in the second study, Asner et al. (2012a) reported a higher result ($R^2 = 0.88$) for three combined sites, including humid and dry forests and shrubland on the island, for which improved allometric relations and differential correction of GPS measures were used. Per site results were not reported, and it is not clear if the dataset from the 2011 study was incorporated into the 2012 one, for which the authors reported a lower mean AGB.

We note that most tropical studies using small footprint lidar, which achieve high accuracy of the estimates, are based on plots more than double the area (0.28 ha) of our plots (Asner et al. 2009, 2010, 2011, 2012a, 2012b; Mascaro et al. 2011b). Mascaro et al. (2011a and 2011b) demonstrated in a tropical moist forest that lidar prediction error, which is strongly related to the edge effect, scales with plot area with a RMSE decreasing from 63.2 to 11.1 Mg C ha^{-1} when increasing the plot size from 0.04 ha to one hectare. Similar conclusions are given by Kohler and Huth (2010) for another tropical site. Furthermore, the edge effect - responsible for disagreement

16

between remote sensing and field plot measures over which trees or parts of trees are inside the calibration plots - is more marked in small plots and in the presence of large tree crowns. In our study the plot size was less than the half of the size most commonly used in lidar calibration studies. Furthermore, since 25% of the measured trees had a DBH > 50 cm, the occurrence of edge effects was likely and indeed often observed, with plots frequently hosting very large crowns from neighboring mature trees (Fig. 2).

In contrast to studies which report that the higher lidar height percentiles explain most of the biomass variance (Patenaude et al. 2004; Skowronski et al. 2007), in our case the maximum VIP scores were assigned to the height percentiles included in the 30-60th range (Fig. 3a). This is an evidence of the multilayered structure of this mature forest, which possibly stores a large part of biomass in the subcanopy layer. This also indicates that the biomass within a plot is not primarily driven by the tallest trees, which despite having the individually largest biomass values are nonetheless far less abundant than the mid-size trees (Fig. 5).

4.2 Sources of uncertainties

The relatively low accuracy (the best R^2 was 0.70) we obtained in this study could be associated with different sources of uncertainty including: field measurement errors, plot locations errors, and errors introduced by the allometric model. These errors, together with error caused by geometrical and radiometric correction of remotely sensed data, are well known sources of uncertainty in remote sensing analysis (Lu et al. 2012).

In our datasets, there was a 5-6 years time lag between field and remote sensing data acquisitions. Even if the growth of a mature forest in a 5-year period is limited, this temporal mismatch can still cause errors in estimates due to natural mortality and regeneration. We exclude plots located close to GRNP border to limit the probability of abrupt forest changes, such as those resulting from illegal clearance or tree harvesting. It is known that forest biomass grows at different rates according to its successional stages (Hudak et al. 2012) and even in mature forests areas of regeneration are present due to natural tree mortality.

In our study site, plot locations were measured using a recreational Garmin GPS. Obtaining accurate GPS measures can be difficult in tall and dense forests; as well as in regions which lack base stations that allow for differential correction (Dominy and Duncan,2001), as was the case in our study area. Chen et al. (2012) reported that the use of plot locations measured by uncorrected GPS decreased the R^2 of the AGB estimates by 0.10–0.13 and increased the RMSE by about 21–31% in the mixed conifer forests of California.

Biomass mapping in Africa suffers from a major lack of regional specific allometric equations. According to studies conducted in the region (Henry et al. 2010; Djomo et al. 2010) the best available option is to use the Chave et al. (2005) general equations. Nevertheless, these equations were obtained without including African tree samples and the issue of their validity in Africa is still debated due to limited data for comparative research. The generic allometric equation that we used could be a major uncertainty. For example Henry et al. (2010) estimated a difference of approximately 40% in AGB values using site-specific versus generalized allometric equations in West Africa.

The control of uncertainty sources, such as those here mentioned, can be a bigger challenge in African forests than in other areas. Most of the African countries lack the technical and financial capacities for field measures extensive collection, and very limited infrastructure to support scientific research is available (Avitabile et al. 2001; Baccini et al. 2009). The establishment of a field network to collect quality ground truth for calibration of remote sensing data, and the development of regional allometric equations, are two major issues to which international programs should direct their support.

4.3 Lidar and hyperspectral data fusion

The addition of hyperspectral features to lidar resulted in an increase of R^2 values from 0.64 to 0.70 (Table 2), which is a slightly greater improvement than has been found in previous studies (Chen 2013). In northern biomes, Anderson et al. (2008) and Swatantran et al. (2011) obtained respectively modest and insignificant improvement using the Laser Vegetation Imaging Sensor (LVIS) and Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) fused datasets. Their results are difficult to compare with ours, due to the coarser resolution of those sensors and the difference in forests types. Swatantran et al. (2011) suggested that the predictive power of hyperspectral could be higher when lidar relationships with biomass are weaker, as observed by Anderson et al. (2008) and Roth (2009). This hypothesis is in part confirmed by our results, in which the lidar-AGB relationship is not as high as elsewhere and an increase in accuracy was brought by inclusion of hyperspectral data. Latifi et al. (2012) used very high spatial resolution sensors, namely a full waveform lidar and HyMap. They also reported minimal improvement in

19

AGB estimates from fused datasets, using PCR. The PLS regression used in this study is preferable to PCR, which might account for the difference. The only AGB estimation for a tropical area, carried out using a FLI MAP lidar and the hyperspectral HYDICE sensor, reported no improvement by the addition of hyperspectral VIs and spectral mixture fractions to lidar metrics (Clark et al. 2012), similar to the very small improvement we observed using our VIs. The increase in accuracy observed with hyperspectral original bands in our study can be explained by the ability of PLS to exploit information from the whole spectrum. The use of VIs, based on a limited subset of spectral bands, possibly excludes important bands for biomass estimation.

Our visual assessment indicates that there is ~1- 4 meters co-registration mismatching between hyperspectral and lidar data, which complicates our evaluation of hyperspectral data for biomass estimation. Given the geometric accuracy of airborne small-footprint lidar usually being sub-meter horizontally, it would be ideal if the georeferencing accuracy of hyperspectral is at the sub-meter level as well, to maximize the use of information from both sensors. This requires precise orthorectification of hyperspectral imagery, preferably based on a Digital Surface Model (DSM) instead of a Digital Terrain Model (DTM) because of the relief displacement caused by trees. This georeferencing accuracy issue has to be taken into account for prospective use of hyperspectral data in AGB estimation, for which the mismatch with reference or other data can be higher.

Latifi et al. (2012) and Papes et al. (2010) found that the most useful spectral ranges for estimating vegetation biomass are green and the NIR plateau, while Zhang et al. (2009) assumed that the greenness indices might have a positive potential toward predicting AGB. The VIP

scores in our study confirmed that the green and NIR portion of the spectra were useful for biomass estimation, but higher scores were obtained for bands in the red-edge (Fig 2b). It is widely recognized that the red-edge position relates to the health status of photosynthetic material in the vegetation (Horler et al. 1983) but it is unclear how this correlates with biomass variation, which calls for more future research along this line. The AISA Eagle sensor used in this study has a wavelength range of ~400 - 900 nm, which can be a limitation considering that other studies in literature have proved the significance of longer wavelengths (SWIR) for vegetation (Brown et al. 2000; Psomas et al. 2011); in particular Gong et al. (2003) proved that SWIR and NIR bands are most important for LAI estimation.

AGB-lidar modeling can be improved by stratifying the vegetation types using optical imagery or ancillary data (Clark et al. 2011; Garcia et al. 2010). Chen et al. (2012) illustrated the positive effect of integrating vegetation type maps derived by aerial photography in Sierra Nevada, using a mixed effects model. In temperate or boreal forests, dominated by few species and where vegetation type maps are often available, this approach can be feasible. However, it is less clear how hyperspectral-based stratification could be carried out in a tropical forest as our site, having very high diversity of tree species, often without marked dominance, and where detailed information on vegetation type is usually not present. The high number of tree species and thus variations in tree morphology, beside variations in spectral responses, can be a reason for explaining the fact that biomass estimation cannot be based on hyperspectral data alone. As a matter of fact literature shows that it is more useful in low-biomass scenarios like grasslands where some studies show that up to 61% of variation can be explained by VIs from hyperspectral data only (Clevers et al. 2007).

5. Conclusions

REDD+ advocates better documentation of performances of remote sensing data across ranges of biomes, vegetation cover, topography/ land forms, seasons, and land use patterns that occur across developing countries (Holmgren, 2008). Recently, the 17th Conference of the Parties (COP) to the United Nation Convention on Climate Change (UNFCCC) adopted the commitment that national REDD+ monitoring and reporting systems shall be based on a combination of field measurements and remote sensing data. Even if clear standards have not yet been established ,the Global Climate Observing System (GCOS, 2011) suggested some accuracy levels, driven by the need to quantify carbon stocks to initialize and test the carbon cycle and for national reporting, which are: <20% error for biomass values over 50Mg/ha, and 10Mg/ha for biomass values < 50t/ha. Houghton et al. (2009) also suggested a maximum of 18% AGB uncertainty. In the case of mature tropical forests, with mean AGB often over 200 Mg/ha, this translates to an error below 40 Mg/ha which is often difficult to achieve even with very advanced tools such as lidar systems. For large area AGB estimation in tropical environments, direct AGB retrieval based on radar sensors, which have full mapping and all weather capabilities, could thus be a cost-effective alternative to lidar sampling followed by further upscaling, especially if new dedicated missions will be launched, such as the European Space Agency Biomass.

Our research evidences that high quality ground truth data, especially in terms of geolocation accuracy and larger plot size, is needed when planning lidar-based AGB estimates in

22

tropical African forests. , Our results suggest that the quality of ground truth data can be even more important if airborne lidar is used as an intermediate step of upscaling field-measured AGB to a larger area or region, a procedure with associated additional uncertainty.

The methodology here presented includes an advanced retrieval algorithm (PLS), significant preprocessing with innovative techniques applied to hyperspectral data (MNF) and a method for testing different features from data fusion. Such a straightforward workflow can provide a robust method for evaluating the importance of spectral contributions and lidar metrics.

The findings related to hyperspectral and lidar data fusion presented in this research are encouraging, but call for additional research. The possible role of hyperspectral data in direct AGB estimation or stratification has to be clarified in different environments, and new VIs that can incorporate relevant biomass information could be developed. As vegetation characteristics strongly influence the sensors ability to retrieve information, additional research in various ecosystems is needed to be able to generalize conclusions about the usefulness of joint sensors use. Overall this study contributes to enlarge research on lidar and hyperspectral fused datasets applicability and provide interesting insight which could orient future sensors development and missions.

6. Acknowledgements

23

References

Adam E., Mutanga O., 2009. Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry. ISPRS Journal of Photogrammetry and Remote Sensing 64(6), 612–620.

Anderson J.E., Plourde L.C., Martin M.E., Braswell B.H., Smith M.L., Dubayah R.O., Hofton M.A., Blair J.B., 2008. Integrating waveform LiDAR with hyperspectral imagery for inventory of a northern temperate forest. Remote Sensing of Environment 112, 1856–1870.

Asner G. P., Martin R. E, 2008. Spectral and chemical analysis of tropical forests: scaling from leaf to canopy levels. Remote Sensing of Environment 112, 3958–3970.

Asner G., Flint Hughes R., Varga T., Knapp D., & Kennedy-Bowdoin T., 2009. Environmental and biotic controls over aboveground biomass throughout a tropical rain forest. Ecosystems 12(2), 261-278.

Asner G. P., Powell G.V.N., Mascaro J., Knapp D.E., Clark J.K., Jacobson J., Kennedy-Bowdoin T,, Balaji A., Paez-Acosta G., Victoria E., Secada L., Valqui M., & Hughes

R. F., 2010. High-resolution forest carbon stocks and emissions in the Amazon. Proceedings of the National Academy of Sciences of the United States of America 107(38), 16738–16742.

Asner G.P., Mascaro J., Muller-Landau H.C., Vieilledent G., Vaudry R., Rasamoelina M., Hall J.S., van Breugel M., 2011. A universal airborne LiDAR approach for tropical forest carbon mapping. Oecologia 168, 1147–1160.

Asner G., Clark J., Mascaro J., Vaudry R., Chadwick K. D., Vieilledent G., Rasamoelina M., Balaji A., Kennedy-Bowdoin T., Maatoug L., Colgan M., and Knapp D., 2012a. Human and environmental controls over aboveground carbon storage in Madagascar, Carbon Balance and Management 7, 2.

Asner G. P., Clark J. K., Mascaro J., Galindo Garcia G. A., Chadwick K. D., Navarrete Encinales D. A., Paez-Acosta G., Cabrera Montenegro E., Kennedy-Bowdoin T., Duque A., Balaji A., von Hildebrand P., Maatoug L., Phillips Bernal J. F., Knapp D. E., Garcia Davila M. C., Jacobson J., & Ordonez M.F., 2012b. High-resolution mapping of forest carbon stocks in the Colombian Amazon. Biogeosciences Discussion 9(3), 2445-2479.

Avitabile V., Herold M., Henry M., Schmullius C., 2011. Mapping biomass with remote sensing: a comparison of methods for the case study of Uganda. Carbon Balance and Management 6, 7.

Baccini A., Laporte N., Goetz S.J., Sun M., Dong H., 2008. A first map of Tropical Africa's above-ground biomass derived from satellite imagery. Environmental Research Letters 045011.

Baccini A., Goetz S. J., Walker W. S., Laporte N. T., Sun M., Sulla-Menashe D., Hackler J., Beck P. S. A., Dubayah R., Friedl M. A., Samanta S. & R. A. Houghton, 2012. Nature Climate Change 2, 182–185.

Baker T.R., Phillips O.L., Malhi Y., Almeida S., Arroyo L., Di Fiore A., Erwin T., Higuchi N., Killeen T.J., Laurance S.G., Laurance W.F., Lewis S.L., Lloyd J., Monteagudo A., Neill D.A., Patino S., Pitman N.C.A., Silva J.N.M., Vasquez Martinez R., 2004. Variation in wood density determines spatial patterns in Amazonian forest biomass. Global Change Biology 10, 545–562.

Brown L., Chen J.M., Leblanc S.G., Cihlar, J., 2000. A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: an image and model analysis. Remote Sensing of Environment 71, 16–25.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-theoretic Approach, Second edition, New York, Springer-Verlag Press.

Chave J., Andalo C., Brown S., Cairns M.A., Chambers J.Q., Eamus D., Fölster H., Fromard F., Higuchi N., Kira T., Lescure J., P., Nelson B. W., Ogawa H., Puig H., Riéra B., Yamakura T., 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia 145(1), 78-99.

Chave J., Coomes D., Jansen S., Lewis S.L., Swenson N.G., Zanne A.E., 2009. Towards a worldwide wood economics spectrum. Ecology Letters 12(4), 351–366.

Chen Q., 2007. Airborne lidar data processing and information extraction. Photogrammetric Engineering and Remote Sensing 73(2), 109–112.

Chen, Q., Gong, P., Baldocchi, D.D., Xie, G., 2007. Filtering airborne laser scanning data with morphological methods. Photogrammetric Engineering and Remote Sensing, 73(2), 175-185.

Chen, Q., Vaglio Laurin G., Battles J., and Saah D., 2012. Integration of airborne lidar and vegetation types derived from aerial photography for mapping aboveground live biomass. Remote Sensing of Environment 121, 108-117.

Chen, Q., 2013. Lidar remote sensing of vegetation biomass. In: Remote Sensing of Natural Resources, edited by Q. Weng and G. Wang, CRC Press: Taylor & Francis Group, pp 399-420.

Cho M.A., Skidmorea A., Corsi F., van Wieren S.E., Sobhana I., 2007. Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. International Journal of Applied Earth Observation and Geoinformation 9(4), 414–424.

Ciais P., Bombelli A., Williams M., Piao S. L., Chave J., Ryan, C. M., Henry M., Brender P., Valentini R., 2011. The carbon balance of Africa: synthesis of recent research studies. Philosophical Transactions of the Royal Society A 369, 2038–2057.

Clark M.L., Roberts D.A. Ewel J.J., Clark D.B., 2011. Estimation of tropical rain forest aboveground biomass with small-footprint LiDAR and hyperspectral sensors. Remote Sensing of Environment 115, 2931–2942.

Clevers, J.G.P.W., van der Heijden,G.W.A.M., Verzakov, S., Schaepman, M.E., 2007. Estimating grassland biomass using SVM band shaving of hyperspectral data. Photogrammetric Engineering & Remote Sensing 73(10), 1141–1148.

Cole, N. H. A., 1993. Floristic association in the Gola rain forests: a proposed biosphere reserve. Journal of Pure and Applied Science 2, 35-50.

Dalponte M., Bruzzone L., Vescovo L., Gianelle D., 2009. The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. Remote Sensing of Environment 133(11), 2345–2355.

De Sy V., Herold M., Achard F., Asner G.P., Held A., Kellndorfer J., Verbesselt J., 2012. Synergies of multiple remote sensing data sources for REDD+ monitoring. Current Opinion in Environmental Sustainability 4, 696–706.

d'Oliveira M.V.N., Reutebuch S.E., McGaughey R.J., & Andersen H. E., 2012. Estimating forest biomass and identifying low-intensity logging areas using airborne scanning lidar in Antimary State Forest, Acre State, Western Brazilian Amazon. Remote Sensing of Environment 124, 479-491.

Dominy N. J. and Duncan B., 2001. GPS and GIS methods in an African rain forest: applications to tropical ecology and conservation. Conservation Ecology 5(2), 6.

Djomo A.N., Ibrahima A., Saborowski J., Gravenhorst G., 2010. Allometric equations for biomass estimations in Cameroon and pan moist tropical equations including biomass data from Africa. Forest Ecology and Management 260, 1873–1885.

Felde G.W., Anderson G.P., Cooley T.W., Matthew M.W., Adler-Golden S.M., Berk A., Lee J., 2003. Analysis of Hyperion data with the FLAASH atmospheric correction algorithm. Proceedings of the International Geoscience and Remote Sensing Symposium, Toulouse, France, (IGARSS'03), 90-92.

Gamon J.A., Penuelas J., and Field C.B., 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sensing of Environment 41, 35-44.

Gamon J.A. and Surfus J.S., 1999. Assessing leaf pigment content and activity with a reflectometer. New Phytologist 143, 105-117.

Garcıa M., Riano D., Chuvieco E., Danson M., 2010. Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. Remote Sensing of Environment 114, 816-830.

GCOS, 2011. Report 154 - Systematic observation requirements for satellite-based products for climate supplemental details to the satellite-based component of the implementation plan for the global observing system for climate in support of the UNFCCC - 2011 Update, December 2011.

Geladi P. and Kowalski B. R., 1986. Partial least-squares regression: a tutorial. Analytica Chimica Acta 185, 1–17.

Gibbs H.K., Brown S., Niles J.O. and Foley J.A., 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality Environ. Res. Lett. 2 045023

Gitelson A.A., Merzlyak M.N., and Chivkunova O.B., 2001. Optical properties and nondestructive estimation of anthocyanin content in plant leaves. Photochemistry and Photobiology 71, 38-45.

Gitelson A.A., Zur Y., Chivkunova O.B., and Merzlyak M.N., 2002.Assessing carotenoid content in plant leaves with reflectance spectroscopy. Photochemistry and Photobiology 75, 272-281.

Gong P., Pu R., Biging G.S., Larrie M.R., 2003. Estimation of forest leaf area index using vegetation indices derived from hyperion hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 41(6), 1355-1362.

Goodenough D.G., Li J.Y., Dyk A., 2006. Combining hyperspectral remote sensing and physical modeling for applications in land ecosystems. IEEE International Geoscience And Remote Sensing Symposium (IGARSS 2006), Denver, Colorado, Vol. 1-8.

Goodenough D. G., Han T., Dyk A., Gour J., and Li J. Y., 2005. Mapping forest biomass with AVIRIS and evaluating SNR impact on biomass prediction, Natural Resources Canada, internal report, (presented at NASA JPL AVIRIS Workshop, May, Pasadena, California).

Goodwin N., Coops N.C., Stone C., 2005. Assessing plantation canopy condition from airborne imagery using spectral mixture analysis and fractional abundances. International Journal of Applied Earth Observation and Geoinformation 7(1), 11–28.

Green A. A., Berman M., Switzer P., and Craig M. D., 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal: IEEE Transactions on Geoscience and Remote Sensing 26(1), 65-74.

Henry M., Besnard A., Asante W. A., Eshun J., Adu-Bredu S., Valentini R., Bernoux M. & Saint-André, L., 2010. Wood density, phytomass variations within and among trees, and allometric equations in a tropical rainforest of Africa. Forest Ecology and Management 260, 1375–1388.

Hudak A.T., Strand E.K., Vierling L.A., Byrne J.C., Eitel J., Martinuzzi S. and Falkowski M.J., 2012. Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. Remote Sensing of Environment 123, 25-40.

Holmgren P., 2008. Role of satellite remote sensing in REDD. United Nations REDD Programme, MRV Working Paper series, 13 October 2008.

Horler D.N.H., Dockray M., Barber J, 1983. The red edge of plant leaf reflectance. International Journal of Remote Sensing 4(2), 273-278.

Houghton R. A., Hall F. G., and Goetz S. J., 2009. The importance of biomass in the global carbon cycle. Journal of Geophysical Research - Biogeosciences 114, Issue G2.

Justice, C., Wilkie, D., Zhang, Q., Brunner, J., Donoghue, C., 2001. Central African forests, carbon and climate change. Climate Research, 17(2), 229-246.

Kaufman Y.J. and D. Tanre, 1996. Strategy for direct and indirect methods for correcting the aerosol effect on remote sensing: from AVHRR to EOS-MODIS. Remote Sensing of Environment 55, 65-79.

Kennaway T. A., Helmer E. H., Lefsky M. A., Brandeis T. A., & Sherrill K. R., 2008. Mapping land cover and estimating forest structure using satellite imagery and coarse resolution lidar in the Virgin Islands. Journal of Applied Remote Sensing 2(1), 1-27.

Klop E., Lindsell J., Siaka A., 2008. Biodiversity of Gola Forest, Sierra Leone. Royal Society for the Protection of Birds, Conservation Society of Sierra Leone, Government of Sierra Leone.

Koch B., 2010. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. ISPRS Journal of Photogrammetry and Remote Sensing 65(6), 581–590.

Kohler, P. and Huth A., 2010. Towards ground-truthing of spaceborne estimates of above-ground life biomass and leaf area index in tropical rain forests. Biogeosciences 7, 2531–2543.

Kronseder K., Ballhorn U., Böhm V., Siegert F., 2012. Above ground biomass estimation across forest types at different degradation levels in Central Kalimantan using LiDAR data. International Journal of Applied Earth Observation and Geoinformation 18, 37-48.

Lei C., Ju C., Cai T., Jing X., Wei X., Di X., 2012. Estimating canopy closure density and above-ground tree biomass using partial least square methods in Chinese boreal forests. Journal of Forestry Research 23(2), 191-196.

le Maire G., François C., Soudani K., Berveiller D., Pontailler J.Y, Bréda N., Genet H., Davi H., Dufrêne E., 2008. Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. Remote Sensing of Environment 112, 3846–3864.

Latifi, H., Faßnacht F., Koch B., 2012. Forest structure modeling with combined airborne hyperspectral and LiDAR data. Remote Sensing of Environment 121, 10-25.

Lewis et al. 2009. Increasing carbon storage in intact African tropical forests. Nature 457, 1003-1006.

Lindsell J. A., Klop E., 2013. Spatial and temporal variation of carbon stocks in a lowland tropical forest in West Africa. Forest Ecology and Management 289, 10-17.

Lu D., Chen Q., Wang G., Moran E., Batistella M., Zhang M., Vaglio Laurin G., and Saah D., 2012. Aboveground forest biomass estimation with Landsat and LiDAR data and uncertainty analysis of the estimates. International Journal of Forestry Research 2012, Article ID 436537.

Lu D., 2006. The potential and challenge of remote sensing-based biomass estimation. International Journal of Remote Sensing 27(7), 1297–1328.

Marquet P.A., Quiñones R.A., Abades S., Labra F., Tognelli M., Arim M., Rivadeneira M., 2005. Scaling and power-laws in ecological systems. The Journal of Experimental Biology 208, 1749–1769.

Mascaro J., Asner G. P., Muller-Landau H. C., van Breugel M., Hall J., & Dahlin, K., 2011a. Controls over aboveground forest carbon density on Barro Colorado Island, Panama. Biogeosciences Discussions 7(6), 8817–8852.

Mascaro J., Detto M., Gregory P. A., Muller-Landau H.C., 2011b. Evaluating uncertainty in mapping forest carbon with airborne LiDAR. Remote Sensing of Environment 115, 3770–3774.

Næsset E., 2007. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. Scandinavian Journal of Forest Research 22(5), 433-442.

Patenaude G., Hill R. A., Milne R., Gaveau D.L.A., Briggs B.B. J. & Dawson, T. P., 2004. Quantifying forest above ground carbon content using LiDAR remote sensing. Remote sensing of environment 93(3), 368-380.

Papes M., Tupayachi R., Martinez P., Peterson A. T., & Powel G.V.N., 2010. Using hyperspectral satellite imagery for regional inventories: A test with tropical emergent trees in the Amazon Basin. Journal of Vegetation Science 21, 342–354.

Peerbhay K. Y., Mutanga O., Ismail R., 2013. Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu–Natal, South Africa. ISPRS Journal of Photogrammetry and Remote Sensing 79, 19–28.

Peterson B., Dubayah R., Hyde P., Hofton M., Blair J. B., Fites-Kaufman J., 2007. Use of LIDAR for forest inventory and forest management application. In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; McWilliams, William H., eds. Proceedings of the seventh annual forest inventory and

analysis symposium; October 3-6, 2005; Portland, ME. Gen. Tech. Rep. WO-77. Washington, DC: U.S. Department of Agriculture, Forest Service: 193-202.

Psomas A., Kneubuhler M., Huber S., Itten K., Zimmermann N.E., 2011. Hyperspectral remote sensing for estimating aboveground biomass and for exploring species richness patterns of grassland habitats. International Journal of Remote Sensing 32(24), 9007–9031.

Roth, K. L.,2009. A combined lidar and hyperspectral remote sensing analysis for mapping forest biomass. Unpublished master thesis. University of California, Santa Barbara, Dept of Geography.

Sims, D.A. and J.A. Gamon, 2002. Relationships Between Leaf Pigment Content and Spectral Reflectance Across a Wide Range of Species, Leaf Structures and Developmental Stages. Remote Sensing of Environment 81, 337-354.

Skowronski N., Clark K., Nelson R., Hom J., and Patterson M., 2007.Remotely sensed measurements of forest structure and fuel loads in the pinelands of New Jersey," Remote Sensing of Environment 108(2), 123–129.

Swatantran A., Dubayah R., Roberts D., Hofton M., & Blair J. B., 2011. Mapping biomass and stressin the Sierra Nevada using lidar and hyperspectral data fusion. Remote Sensing of Environment 115, 2917–2930.

Treuhaft R.N., Asner G.P., Law B.E., 2003. Structure-based forest biomass from fusion of radar and hyperspectral observations. Geophysical Research Letters 30 (9), 1472-1479.

Underwood E., Ustin S., DiPietro D., 2003. Mapping nonnative plants using hyperspectral imagery. Remote Sensing of Environment 86(2), 150–161.

Vaglio Laurin G., Liesenberg V., Chen Q., Guerriero L., Del Frate F., Bartolini A., Coomes D., Wilebore B., Lindsell J., and Valentini R., 2013. Optical and SAR sensor synergies for forest and land cover mapping in a tropical site in West Africa. International Journal of Applied Earth Observation and Geoinformation 21, 7-16.

Vogelmann, J.E., B.N. Rock, Moss D.M., 1993. Red Edge Spectral Measurements from Sugar Maple Leaves. International Journal of Remote Sensing 14, 1563-1575.

Williams A.P., Hunt E.R., 2002. Estimation of leafy spurge cover from hyperspectral imagery using mixture tuned matched filtering. Remote Sensing of Environment 82(2–3), 446–456.

Wold, S., Sjostrom, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58 (2), 109–130.

Zhang, H., Hu, H., Yao, X. , & Zheng, K., 2009. Estimation of above-ground biomass using HJ-1 hyperspectral images in Hangzhou Bay, China. International Conference on Information Engineering and Computer Science, 2009. ICIECS 2009.

Zolkos S.G., Goetz S.J., Dubayah R., 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. Remote Sensing of Environment 128, 289-298.

"List of Figure Captions"

Figure 1. The study area located along the border between Sierra Leone and Liberia, and included in the GRNP. The flight lines, realized during an airborne survey, cover part of the permanent field plots established in the Park.

Figure 2. In (A) and (B) false-color composite of hyperspectral data at 807.5(R), 597.3 (G) and 467.3 (B) nm. In (A): strip of data where large crowns are visible. In (B): example of edge-effect for a specific plot. In (C): the same edge effect in (B) is visualized in the aerial photograph. Plot edges are represented as black circles. In (D): the same plot in (B) and (C) is visualized as a lidar point cloud.

Figure 3. Scatterplots of predicted vs. field observed AGB for the following inputs: (a) lidar metrics, (b) hyperspectral bands, (c) lidar metrics and VIs, (d) lidar metrics and hyperspectral bands.

Figure 4. VIP scores for individual lidar (a) and hyperspectral (b) datasets.

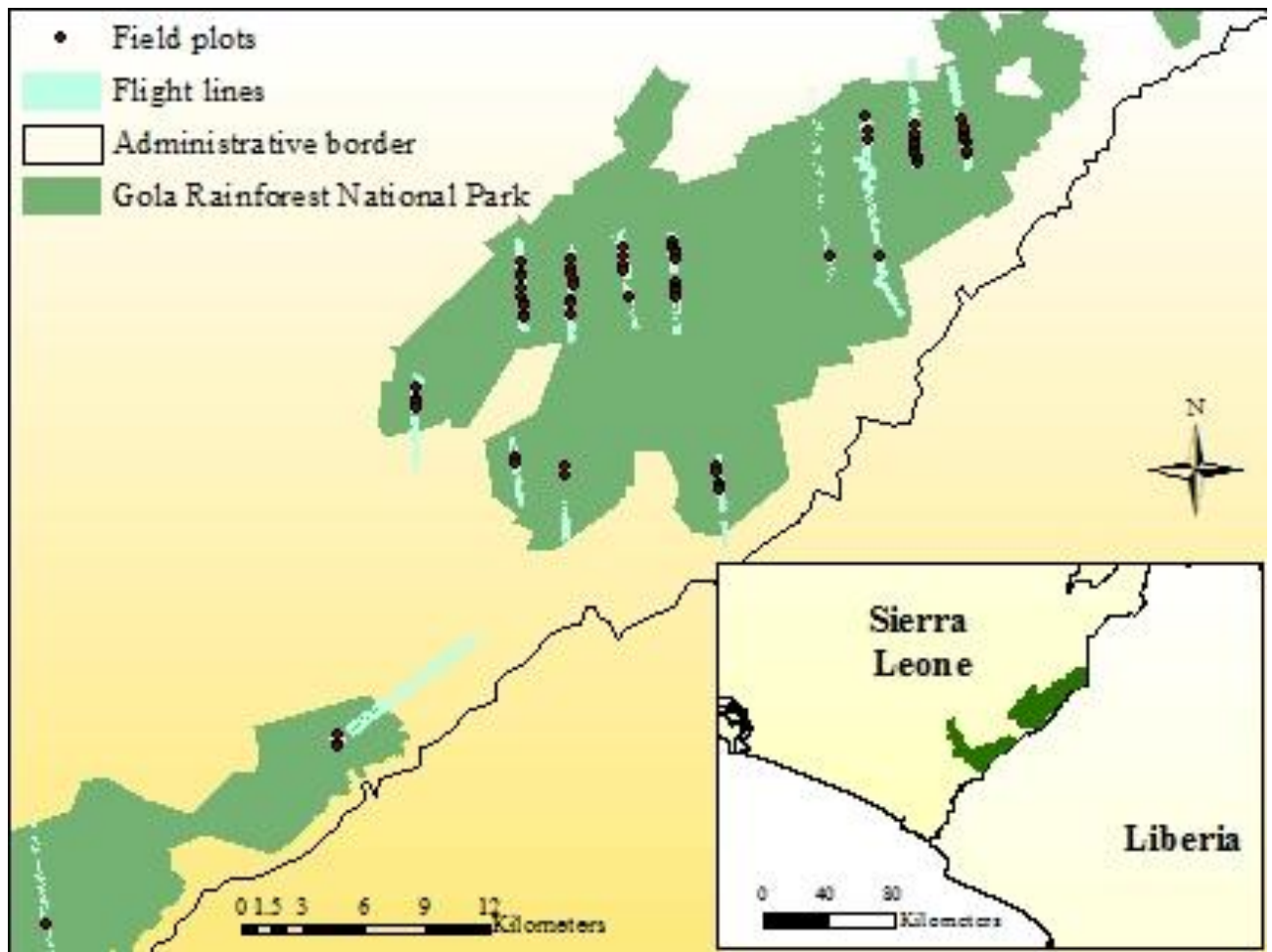*Figure 5. AGB and number of trees in the 70 plots (total area = 87500 m^2) according to different ranges of field-observed height.*
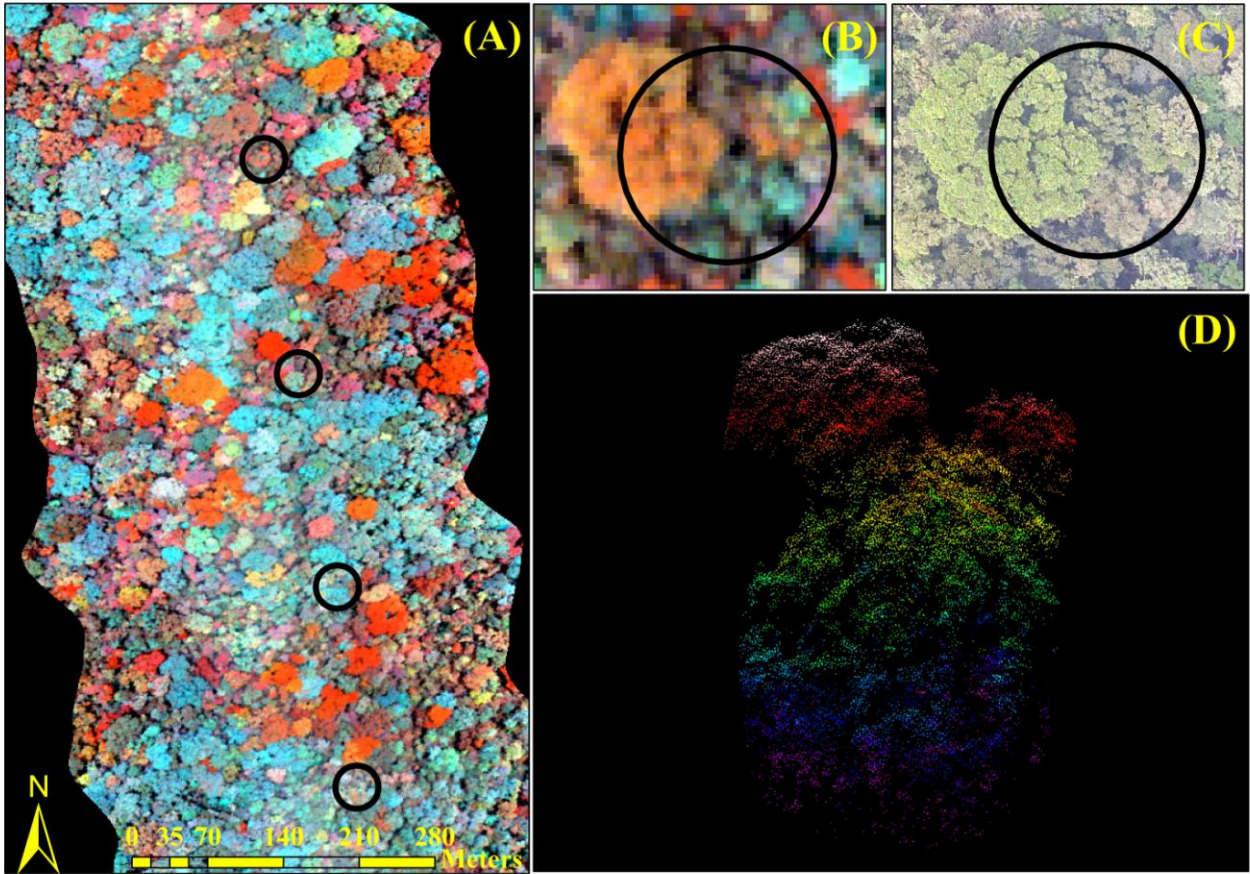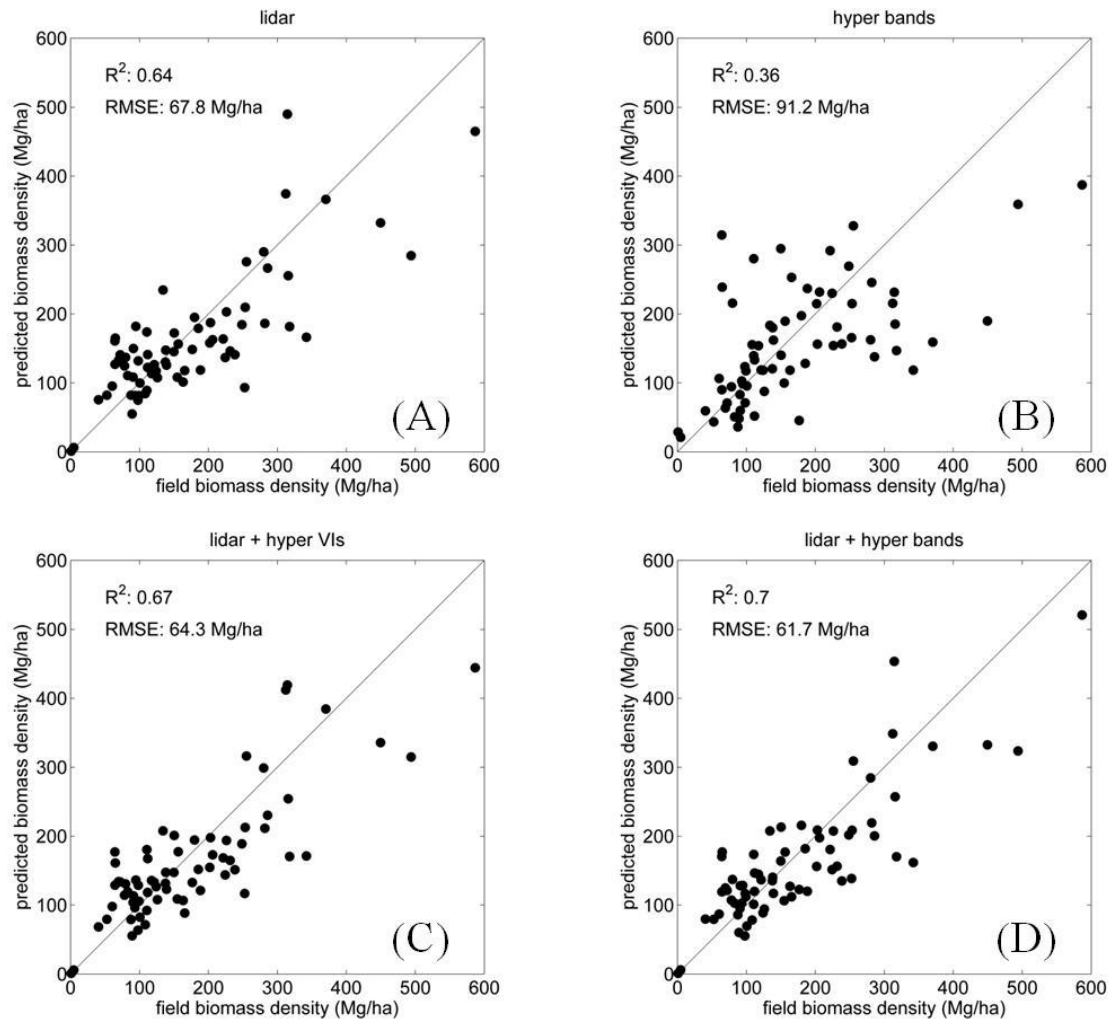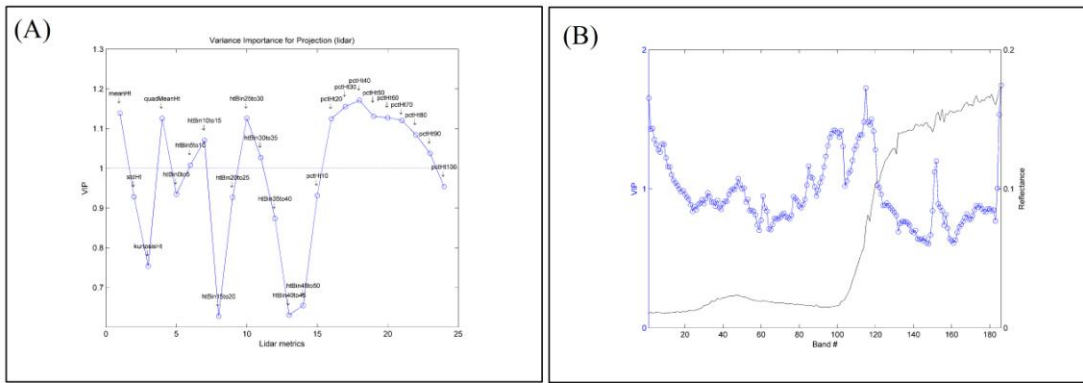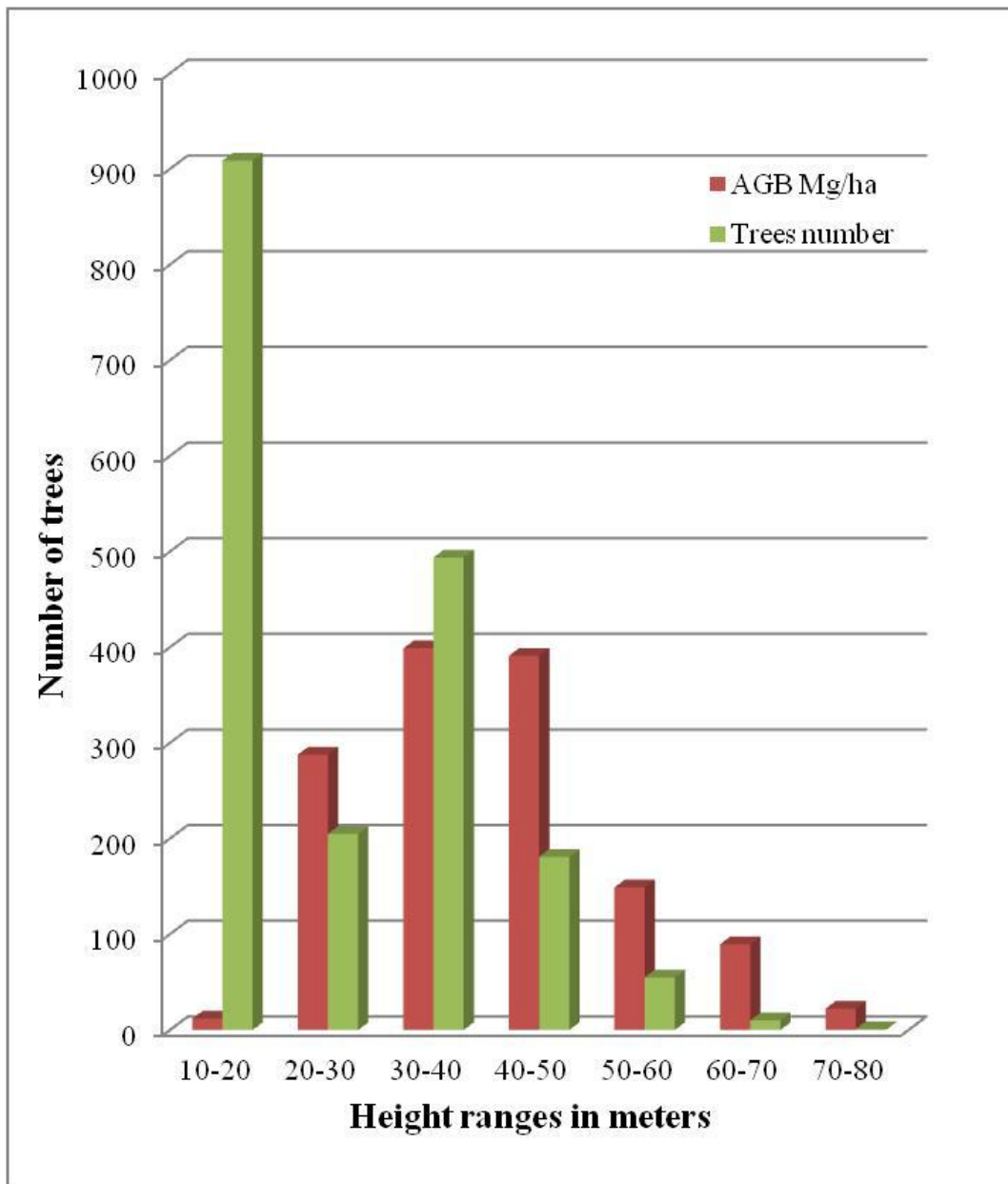
Fig. 1

Fig. 2

Fig.3

Fig.4

Fig. 5

Chapter 7

Biodiversity mapping in a tropical West African forest with airborne hyperspectral data

Research paper under review at PLOS One.

Biodiversity mapping in a tropical West African forest with airborne hyperspectral data

Gaia Vaglio Laurin [a,g], Jonathan Cheung-Wai Chan [b], Qi Chen [c], Jeremy Lindsell [d], David Coomes [e], Leila Guerriero [f], Fabio Del Frate [f], Franco Miglietta [b], Riccardo Valentini [a,g]

[a]CMCC - Centro Mediterraneo per i Cambiamenti Climatici, via Augusto Imperatore (Euro-Mediterranean Center for Climate Change), IAFENT Division, via Pacinotti 5, Viterbo 01100, Italy

[b]Fondazione Edmund Mach di San Michele all'Adige. Via E. Mach, 1 38010 S. Michele all'Adige (TN) – ITALY

[c]Department of Geography, University of Hawai`i at Mānoa, 422 Saunders Hall, 2424 Maile Way, Honolulu, HI, 96822, USA

[d]The Royal Society for the Protection of Birds, The Lodge, Sandy, Beds. SG19 2DL, UK.

[e]Forest Ecology and Conservation Group, Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK.

[f]Tor Vergata University, Department of Civil Engineering and Computer Sciences Engineering, Via del Politecnico 1, 00133 Rome, Italy

[g]Department of Forest Resources and Environment, University of Tuscia, Viterbo I-01100 Italy

Short title: Tropical forest biodiversity mapping

Corresponding author: Gaia Vaglio Laurin, CMCC/Iafent Division, via Pacinotti 4, 01100 Viterbo, Italy,

e-mail: gaia.vagliolaurin@cmcc.it, laurin@disp.uniroma2.it

Keywords: biodiversity, tropical, forest, Africa, hyperspectral

Word count: 5946

Abstract

Tropical forests are a major repository of biodiversity. Spatial information about tropical forest biodiversity is scarce but it is fundamental for conservation decision making, especially at fine scales. Remote sensing is increasingly contributing to biodiversity mapping and monitoring. Airborne hyperspectral data have been successfully used for tree species classification and retrieval of species richness, but studies linking airborne hyperspectral data with the Shannon-Wiener biodiversity index are not available, and no previous investigation has been done with hyperspectral sensors in African forests. We retrieved the Shannon-Wiener biodiversity index in a moist tropical forest in Sierra Leone with Random Forests (Pseudo R^2 = 84.91% and Out-of-bag RMSE = 0.30), and using as inputs reflectance values from 186 bands in the VIS-NIR spectral range, collected by an AISA Eagle airborne sensor. Lower accuracy was obtained when using 1st derivative of reflectance (R^2 = 71.42%, OOB RMSE= 0.35), while the use of vegetation indices derived from reflectance was unsuccessful in predicting the index. The inputs ranking procedure embedded in Random Forests allowed evaluating the contribution of different regions of the spectra to models. The present research demonstrates for the first time in Africa and, to our knowledge in tropical forests, the ability of airborne hyperspectral sensor to predict the canopy Shannon-Wiener index. These results, together with those obtained by similar using a spaceborne hyperspectral sensor, support the overall use of hyperspectral data for biodiversity mapping and the integrated use of platforms, especially in view of forthcoming hyperspectral satellite missions. Areas with high biodiversity or vulnerability to change could be monitored with airborne sensors, while regional monitoring can be instead done by means of satellite-borne hyperspectral systems, further allowing multi-temporal studies.

1. Introduction

Mapping biological diversity is a major conservation priority (Gaston 2000), due to increasing threat from anthropogenic pressure, habitat loss and fragmentation, and climate change effects (Thomas et al. 2004). Mapping biodiversity is also a main activity in international agreements promoting biodiversity conservation, i.e. the Aichi Biodiversity Targets and the Millennium Development Goals.

Tropical forests are a major repository of biodiversity (Chapin et al. 2000) hosting over half of the world's plant species (Foody 2003; Thomas et al. 2004). Spatial information about tropical forest biodiversity is scarce but it is also fundamental for conservation decision making (Balmford and Whitten 2003), and in view of long-term carbon storage (Diaz et al. 2009) and co-benefits of the Reduction of Emission from Deforestation and Degradation (REDD+) program (Diaz et al. 2009; Paoli et al. 2010).

Remote sensing is critical for biodiversity mapping and monitoring, allowing the extrapolation of local field measures, especially difficult to collect in inaccessible tropical regions, to larger areas (Tuner et al. 2003; Oldeland et al. 2010). Palmer et al. (2002) proposed the 'spectral variation hypothesis' to explain why electromagnetic measures are related to biological diversity: the spectral variation of reflectance values is correlated with spatial variation in the environment by means of landscape structure and complexity. Habitat heterogeneity is further linked to niche complexity which is known to enhance species diversity (Simonson et al. 2012). Considering tree diversity in forests, the remote sensing measures are in relationship with the chemical and structural properties of the vascular species.

Multispectral imagery has been already employed to map biodiversity: relationships have been found between Normalized Difference Vegetation Index (NDVI) derived from satellite imagery and diversity of different ecosystems (Gould 2000; Kerr and Ostrovsky; Leyequien et al. 2007; Nagendra 2001). Foody and Cutler (2003, 2006) used neural networks to correlate tropical tree species richness and evenness in Borneo with Landsat Thematic Mapper (TM) reflectance, obtaining a correlation coefficient of 0.54 in the first study and of 0.69 in the second one. At a slightly higher spatial resolution, Feilhauer and Schmidtlein (2009) predicted species richness and diversity of a German walnut-fruit forest using the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) reflectance, obtaining respectively a coefficient of determination of 0.51 and 0.61. Rocchini et al. (2004) used very high spatial resolution Quickbird reflectance to predict species richness of an Italian wetland, finding that at the plot scale (1100 m^2) the measure of spectral heterogeneity was able to predict only 20% of the variance in species richness while using aggregated data (1 ha), the coefficient of determination reached 0.48. In tropical forests, arboreal diversity can be considered a proxy measure of overall biodiversity (Gentry 1988) but, as emerged from different studies, the variability in vegetation biodiversity can only be partially captured with medium spatial resolutions and multispectral data (Carlson et al. 2007; Foody and Cutler 2003; Gould 2000; Rocchini et al.

2004, 2007). Furthermore, indirect mapping methods, such those coupling land cover maps with remote sensing data to evaluate diversity of species, cannot be used where land cover information is lacking or where the cover is homogeneous, such as in tropical forests (Nagendra 2001).

Fine scale biodiversity maps (< 0.5 km resolution) are needed by land managers and scientists, because they can provide an understanding of species distributions on a scale commensurate with conservation, management and policy development activities (Carlson et al. 2007). The increased availability of very high spectral resolution sensors has provided the opportunity to conduct detailed ecological studies on terrestrial ecosystems characteristics (Kumar et al. 2001; Thenkabail et al. 2004a), but few previous researches related hyperspectral data to biodiversity. Hyperspectral data may provide information on how chemical and structural properties of vascular plants vary within and across ecosystems (Martin and Aber 1997; Ustin et al. 2004) and have mainly been applied in ecological research to temperate areas. The tropics, hosting a greater variety of landscapes, habitats, and species, and with more complex canopy structures, have received so far less attention with few published studies and none for Africa (Levin et al. 2007; Nagendra and Rocchini 2008; Rocchini 2007; Townsend et al. 2008). In tropical forests, spaceborne hyperspectral data has been proved useful to estimate plant species richness (Kalacksa et al. 2007), while airborne hyperspectral data have been successfully used for tree species classification (Clark et al. 2005; Zhang et al. 2006) and retrieval of species richness (Carlson et al 2007).

The objective of the present study is to assess the usefulness of airborne hyperspectral data to predict the arboreal biodiversity of a West African moist forest. As a biodiversity measure we used the Shannon-Wiener Index (Shannon 1948) which is possibly the most used index allowing comparison with similar studies, while as modeling tool we selected Random Forests (Breiman 2001), a machine learning algorithm able to deal with high number of inputs, solving nonlinear problems. According to our knowledge, this is the first research illustrating that airborne very high resolution data are useful in predicting the Shannon-Wiener index of a tropical forest.

2. Material and methods

2.1 Study area and field data

The study area is located at the westernmost end of the West African Upper Guinean Forest Belt, in Sierra Leone, covering the central portion of the Gola Rainforest National Park (GRNP) and for a smaller extent the southern portion (Fig. 1). The region is characterized by lowland moist evergreen forests, with some drier types in place, dominated by *Fabaceae, Euphorbiaceae* and *Sterculiaceae* families (Cole 1993). The GRNP area has been protected through conservation programs since 1989 but commercial logging, most intensively in the southern block, was carried out in 1963–1965 and 1975–1989. Recent land cover mapping highlighted the importance of the GRNP in conserving this forest from anthropogenic pressure in the surrounding areas (Vaglio Laurin et al. 2013). The climate is moist tropical, with annual rainfall around 2500–3000 mm, a dry season from November to April coincident with leaf-off condition of some semi-deciduous tree species, and an altitude of 70–410 m. Floristic information has been derived from a field survey carried out in 2006-2007 (Lindsell and Klop 2013). During that survey all trees with Diameter at Breast Height (DBH) > 30 cm were recorded in circular plots sized 0.125 ha. We selected the plots surveyed by an hyperspectral airborne campaign, excluding those located less than 1 km from the park boundary and those affected by cloud shadow in the hyperspectral data, retaining a total of 64 ground truth plots.

The biodiversity of a particular group of organisms in a location can be quantified in terms of richness and evenness (Magurran, 2004). An abundance-based measure of plant diversity, like the Shannon-Wiener Index, should reflect the structural variability of a landscape much better than species richness, because it captures differences in composition and dominance structure of a given plant community (Foody and Cutler 2003). We calculated the Shannon-Wiener index for each plot, according to the formula:

$$H' = -\sum_{i=1}^{R} p_i \ln p_i$$

where, p_i is the proportion of individuals belonging to the *i*th species in the plot data.

2.2 Remote sensing data

In March 2012 an airborne survey collected hyperspectral data over parts of the Gola GRNP, using an AISA Eagle sensor with FOV equal to 39.7°, set to record 244 bands with 2.3 nm spectral resolution in

the 400-1000 nm range and spatial resolution of 1 m after radiometric correction and orthorectification (Fig. 2). Atmospheric correction of the hyperspectral image strips was performed using the Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm (Felde et al. 2003). Due to high noise levels, all the bands out of the 450-900 nm range and four bands in the 759-766 nm range were removed, reducing the total number of bands to 186. Minimum Noise Fraction (MNF) transformation (Green et al. 1988) was used to further reduce noise in the dataset. For each image strip, 9 to 15 MNF components were selected by visual screening and used to compute the inverse MNF to transform back the bands in the original data space. For each plot, we then calculated different statistical metrics for the 186 hyperspectral bands after MNF inversion, including minimum, maximum, mean, and standard deviation.

Derivatives can be useful for data analysis as small variations of spectral curve can be enhanced and background noise suppressed (Tsai and Philpot 1998, Gong et al. 1997). First order derivatives were generated dividing the difference between successive reflectivity values by the wavelength interval separating them. Then a seven-point moving filter was applied for smoothing (Han and Rundquist, 1997, Demetriades-Shah et al., 1990) and statistical metrics where calculated. We also calculated Photochemical Reflectance Index (Gamon et al. 1992), Red Edge Normalized Difference Vegetation Index (Sims and Gamon 2002), Atmospherically Resistant Vegetation Index (Kaufman and Tanre 1996), Vogelmann Red Edge Index (Vogelmann et al. 1993), Red Green Ratio (Gamon and Surfus 1999), Simple Ratio (Sellers 1985), Anthocyanin Reflectance Index (Gitelson et al. 2001).

2.3 Retrieval method

Random Forests (RF) is an ensemble learning method for regression and classification, which creates multiple decision trees and provides in output the regression model that is the mode of the regression output by individual trees. The method combines bagging (Breiman 1996), which is a bootstrap aggregating method, and the random selection of features in order to build a collection of decision trees with controlled variation. Bagging improves the stability and accuracy of machine learning algorithms, reducing variance and avoiding overfitting. The RF algorithm estimates the importance of a variable by looking at how much prediction error increases when out-of-bag data for that variable is permuted while all others are left

unchanged; the necessary calculations are carried out tree by tree as the RF is constructed (Liaw and Wiener 2002). Out-of-bag samples can be used to calculate an unbiased error rate and variable importance, eliminating the need for a test set or cross-validation; because a large number of trees are grown, there is limited generalization error (that is, the true error of the population as opposed to the training error only).

In previous studies using hyperspectral data, RF has been used by Clark and Roberts (2012) to discriminate tropical tree species and by Leutner et al. (2012) to analyze the species richness of a temperate montane forest in Germany. Using coarse resolution satellite-derived and climate data, Parmentier et al. (2011) predicted the species richness in African rain forests.

In our study, we created three different datasets, one including all the metrics (minimum, maximum, mean, standard deviation) derived from hyperspectral bands reflectance (n = 744; 186 bands by four metrics each), the other including the same metrics derived from first derivatives of reflectance (n = 716; 179 bands by four metrics each) and the third from vegetation indices derived by reflectance (n = 7; seven vegetation indices calculated). We used the three data sets as input to predict the Shannon diversity index. Two assessment criteria are provided by RF using Out-of-bag (OOB) strategy, an unbiased internal estimate of RF: Pseudo R^2 and OOB-MSE. Pseudo R^2 is equal to 1- (MSE/variability explained). MSE is the OOB mean squared error of residuals. We have used Pseudo R^2 and OOB-RMSE, corresponding to the squared root of OOB-MSE, to assess the performance of the model. The OOB strategy also enables two methods of feature ranking. The first is the increase in MSE if a particular feature is being removed. The second is the increase of purity among the splitting groups in the process of building a decision tree if a particular feature is used. We have chosen the first strategy 'increase in MSE' to understand the importance of spectral regions.

3. Results

3.1 Field data

The field data analysis showed that the 64 plots contained a total of 133 species. In the cumulated sampled area (8.125 ha) the total number of recorded trees was 676. The most common species and families, considered as those having a minimum of 10 individuals per species, composed more than 50% of the samples with *Caesalpinioideae* as the most represented family, and are illustrated in Table 1, which shows

that only 15 species are represented by 10 or more individuals. The species-area curve (Connon and McCoy 1979) shows that the sampled area was enough to capture most of the diversity of the site (Fig. 3). The Shannon-Wiener index ranged between 0 and 2.63, with a mean value of 1.68 and a standard deviation of 0.48.

3.2 Regression results

The results obtained from the RF OOB estimation indicates that Shannon-Wiener index can be accurately predicted using the plot-level statistics derived by hyperspectral bands, which were used as inputs in the best model, resulting in a Pseudo R^2 = 84.91% and a OOB-RMSE = 0.30. The results obtained with statistics from first derivatives were lower, with a Pseudo R^2 = 71.42% and OOB-RMSE = 0.35. Figure 4 shows the plot between RF predictions and Shannon Index using metrics and 1st order derivatives. The model based on vegetation indices was unsuccessful, with a negative Pseudo R^2 that means lack of ability to explain more variability than those expressed by the average of Shannon values for the plots. All results are presented in Table 2.

Fig. 5 and 6 were generated to illustrate the differences in inputs ranking for the two models: Fig. 5 for the ranking of hyperspectral metrics (maximum, minimum, mean, standard deviation of band reflectance) and Fig. 6 for the same metrics derived from the 1st order derivative of reflectance. The y-axis represents the 'percentage of increase in MSE' and the x-axis is the band region. When derivatives are used, standard deviations from the near infrared region provides by far the highest ranking inputs, possibly due to the ability of the derivatives to suppress background signals making this region more useful. When hyperspectral band metrics are used, the most important inputs come from the standard deviations from the green region, but contributions come from all the available spectra and other metrics too. In both models the most ranked statistical metric has been the standard deviation, which indicates that is the spectral variation which provides most information on diversity variation. However, unreported tests run with standard deviation metrics only produced slightly lower results in accuracy.

4. Discussion

Our results indicate that is possible to obtain important biodiversity information at very fine spatial resolution over tropical forests. To our knowledge, this is the first study which demonstrates the usefulness of airborne hyperspectral data to predict the Shannon-Wiener index in a tropical forest.

Few previous researches related hyperspectral data to biodiversity. At very high spatial resolution, Carlson et al. (2007) mapped the vascular plants species richness in lowland forest in Hawaii using the NASA Airborne Visible and Infrared Imaging Spectrometer (AVIRIS), covering the Visible (VIS) to Short Wave Infrared (SWIR) range and with a pixel size of 3.6 m. They found that the derivative reflectance in those wavelengths regions associated with upper-canopy pigments, water and nitrogen content, was well correlated with species richness across field sites. A linear regression based on the most useful wavelengths, centered at 530, 720, 1201 and 1523 nm, was used to predict species richness, obtaining a coefficient of determination equal to 0.85. A much weaker correlation was instead obtained using the Shannon-Wiener index. In a temperate montane forest, Leutner et al. (2012) used HyMap VIS-SWIR hyperspectral and lidar full waveform airborne data to model species richness and community composition using data at 7 m spatial resolution, finding with the addition of lidar data a coefficient of determination equal to 0.29, and obtaining lower result for the Shannon index ($R^2 < 0.17$). The same sensor, with 5 m pixel size, was also used by Oldeland et al. (2010) in a savanna ecosystem, who reported a moderate relationship, with a R^2 of 0.41 for Shannon-Wiener index. To our knowledge, the only other study conducted in a tropical forest with hyperspectral data was realized by Kalacska et al. (2007) using the satellite Hyperion sensor in Costa Rica with 30m pixel size. Using wavelet decomposition followed by a stepwise regression they found that the Shannon index could be predicted with a R^2 of 0.84; vegetation indices were not as good predictors as wavelet features. The selected bands were those from the shortwave infrared region and one from the visible region of the spectra (621 nm). Among the studies conducted without hyperspectral data, one successful result has been obtained by Feilhauer and Schmidtlein (2009) who retrieved the Shannon index with a coefficient of determination of 0.61 in a temperate forest using ASTER data.

Our results are very similar to those obtained by Kalacska et al. (2007) with respect to the ecosystem under analysis, the ability to retrieve the Shannon-Wiener index, and the poor results obtained with vegetation indices. Our results are also in accordance to the Grime's (1998) mass ratio hypothesis, which states that immediate control over ecosystem processes, such as water balance and nutrient cycles, depends

primarily on the functional characteristics of the most abundant species, which are generally better suited to the abiotic environment of the study site. By contrast, rare species are a relatively incidental set of species that are more variable in their functional characteristics compared with common species. Hence, the ecological implications of the most abundant species make the Shannon index a powerful tool for relating spectral and species diversity at a local scale, taking species abundance into account instead of relying solely on local richness (*from* Kalacska et al 2007). A key open question in biodiversity studies is whether information on canopy biodiversity can be a surrogate for sub-canopy biodiversity; Luetner et al. (2012) suggested that hyperspectral data reflects environmental conditions acting upon plants, such as soil pH, water availability, nitrogen availability and others, which are known to influence species distributions and community composition; Asner et al. (2008) estimated the diversity of foliar chemicals within the canopy as a whole using hyperspectral data, relating this to faunal and floral distributions.

Similarly to what has been found by Zhang et al. (2006) in a tropical tree classification study, our results suggest that derivative analysis it might not be optimal for our aims and data, resulting in a lower R^2 (Fig. 4; Table 2). A possible explanation can be attributed to the fact that derivative is very sensitive to noise in the original spectrum. The residual noise is emphasized in the derivative spectra and this may vary according to the pixel location on the tree crown. In addition, environmental or stress factors such as moisture content and leaf age introduce subtle variations in crown reflectance that are enhanced by differentiation. Consequently, spectral variation within crowns can be greatly enlarged in the derivative domain, interfering with identification of differences among crowns (Zhang et al. 2006).

The selection as most important variables of a majority of standard deviation metrics, for both models, indicates that is the amount of spectral variability among plots the valuable input for predicting biodiversity. Spectral variation can be determined by variation within individual tree crowns, between tree crowns of the same species, and among spectra of different species, and the overall picture is complicated by reflectance, absorption, and transmission properties of the leaves and wood, viewing geometry and other environmental variables such as microclimates, soil characteristics, precipitation, topography and soil moisture, foliage age, position in the canopy, chlorophyll content, forest vigor and the presence of liana (Zhang et al. 2006). Within canopy variation, even within the same tree, is commonplace as leaves may be of varying age and suffer from different levels of necrosis, herbivory and epiphyll cover (Lucas et al. 2008). At

the crown level, contribution from non-foliar surfaces and variations in shading and anisotropic multiple scattering relative to illumination and view geometry may led to further confusion (Clark et al., 2005). The selected scale of analysis can also have an influence on accuracy of results: when pixel dimensions shrink below the size of the object studied (for instance individual tree crowns), then a sudden increase in the variability of reflectance values from pixels that cover the same individual tree may happen, due to some pixels covering leaves in sunshine and others located over dark gaps between leaves, or on the tree bark (Nagendra and Rocchini 2008; Rocchini and Vannini 2008). In this research the obtained results are similar to those recorded for another tropical area with a hyperspectral satellite sensor having larger spatial resolution (Kalacska et al. 2007), indicating that our processing successfully removed noise, and that airborne hyperspectral data are an optimal tool to predict forest biodiversity.

There are specific limitations affecting the present study, including: the time lag between field data collection and airborne survey; the quality of field plots geo-location, obtained without the differential Global Positioning System correction and thus prone to errors; the limited spectra collected by the sensor, which does not include the SWIR region. Considering that these most of these limitations can be easily overcome in other studies, we evaluate the presented results as very encouraging for future biodiversity studies at local scale.

5. Conclusions

The present research demonstrates for the first time in Africa and, to our knowledge in tropical forests, the ability of an airborne hyperspectral sensor to predict the canopy Shannon-Wiener index.

The use of airborne hyperspectral sensors can specifically target areas with high biodiversity or vulnerability to change (e.g., occurring on deforestation fronts) and/or tree species that are of particular importance (Lucas et al. 2008). Regional monitoring can be instead done by means of satellite-borne hyperspectral systems, further allowing multi-temporal studies.

With different forthcoming satellite hyperspectral missions, there is a clear need to increase research on hyperspectral applications, as well as to increase the collection and dissemination of quality biodiversity field information, and to incorporate these remote sensing data into mapping and monitoring activities, as a way to improve our understanding of the distribution of life on earth.

The use of hyperspectral imagery can be preferred in inaccessible landscapes and, with a proper sampling strategy, airborne data can provide transects useful as ground truth surrogates to estimate the biodiversity of larger areas.

Studies using hyperspectral and lidar data for species distribution mapping mostly report increased accuracy when using both sensors with respect to their single use (Dalponte et al. 2008; Feret and Asner 2012; Jones et al. 2010), but this integration has yet not been proven to be equally effective for estimation of biodiversity; the inclusion of lidar derived geomorphological predictors, which have been shown to influence species distribution patterns at various scales (Hofer et al. 2008), can be a way to successfully integrate these two sensors. It is also important to explore the extent of generalization of the hyperspectral-biodiversity relationship, at least at regional biome level, as this could further reduce the need of site-specific ground truth. All these topics will be the addressed by further research of our group.

Literature cited

Asner GP, Martin RE, Ford AJ, Metcalfe DJ, and Liddell MJ (2009). Leaf chemical and spectral diversity in Australian tropical forests. Ecological Applications 19:236–253

Asner GP (2008). Hyperspectral Remote Sensing of Canopy Chemistry, Physiology, and Biodiversity in Tropical Rainforests. In: Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests Edited by Margaret Kalacska and G . Arturo Sanchez-Azofeifa. CRC Press.

Balmford A & Whitten T (2003). Who should pay for tropical conservation, and how could the costs be met?. Oryx, 37(02), 238-250.

Breiman L (2001). Random Forests. Machine Learning 45 (1): 5–32.

Breiman L (1996). Bagging predictors. Machine Learning 24 (2): 123–140.

Carlson KM, Asner GP, Hughes RF, Ostertag R, Martin RE (2007). Hyperspectral remote sensing of canopy biodiversity in Hawaiian lowland rainforests, NY, Print. Ecosystems 10:536–549

Chapin FS, Zavaleta ES, Eviner VT, Naylor RL, Vitousek PM, Reynolds H.L, Hooper DU, Lavorel S, Sala OE, Hobbie SE, Mack MC & Diaz S (2000). Consequences of changing biodiversity. Nature, 405, 234–242.

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

Clark ML & Roberts DA (2012). Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. Remote Sensing, 4(6), 1820-1855.

Cole NHA (1993). Floristic association in the Gola rain forests: a proposed biosphere reserve. Journal of Pure and Applied Science, 2, 35-50.

Connor EF and McCoy ED (1979). The statistics and biology of the species-area relationship. American Naturalist 113:791-833.

Dalponte M, Bruzzone L, Gianelle D (2008). Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. IEEE Trans. Geosci. Remote Sens. 46, 1416–1427.

Demetriades-Shah TH, Steven MD and Clark J (1990). High resolution derivatives spectra in remote sensing. Remote Sensing of Environment, 33:55-64.

Díaz S, Hector A, Wardle DA (2009). Biodiversity in forest carbon sequestration initiatives: not just a side benefit. Curr. Opin. Environ. Sustain., 1, pp. 55–60

Feilhauer H and Schmidtlein S (2009). Mapping continuous fields of forest alpha and beta diversity. Applied Vegetation Science 12: 429-439.

Felde GW, Anderson GP, Cooley TW, Matthew MW, Adler-Golden SM, Berk A, Lee J (2003). Analysis of Hyperion Data with the FLAASH Atmospheric Correction Algorithm. Proceedings of the International Geoscience and Remote Sensing Symposium, Toulouse, France, (IGARSS'03), 90-92.

Feret JB, Asner GP (2012). G.P. Semi-supervised methods to identify individual crowns of lowland tropical canopy species using imaging spectroscopy and LiDAR. Remote Sens. 4, 2457–2476.

Foody GM (2003). Remote sensing of tropical forest environments: Towards the monitoring of environmental resources for sustainable development. International Journal of Remote Sensing, 24, 4035, 2003.

Foody GM, Cutler MEJ (2003). Tree biodiversity in protected and logged Bornean tropical rain forests and its measurement by satellite remote sensing. J. Biogeogr. 30, 1053–1066.

Foody GM, Cutler MEJ (2006). Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks. Ecol. Model. 195, 37–42.

Gamon JA, Penuelas J, and Field CB (1992). A Narrow-Waveband Spectral Index That Tracks Diurnal Changes in Photosynthetic Efficiency. Remote Sensing of Environment 41:35-44.

Gamon JA and Surfus JS (1999). Assessing Leaf Pigment Content and Activity With a Reflectometer. New Phytologist 143:105-117.

Gaston KJ (2000). Global patterns in biodiversity. Nature 405:220–227.

Gentry AH (1988). Changes in plant community diversity and floristic composition on environmental and geographic gradients. Annals of the Missouri Botanical Garden. 75 1-34

Gitelson AA, Merzlyak MN, and Chivkunova OB (2001). Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. Photochemistry and Photobiology 71:38-45.

Gong P, Ru R, and Yu B (1997). Conifer species recognition: An exploratory analysis of in situ hyperspectral data, Remote Sensing of Environment, 62:189-200.

Gould W (2000). Remote Sensing of Vegetation, Plant Species Richness, and Regional Biodiversity Hotspots. Ecological Applications Vol. 10, No. 6, pp. 1861-1870.

Green AA, Berman M, Switzer P, and Craig MD (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal: IEEE Transactions on Geoscience and

Remote Sensing, v. 26, no. 1, p. 65-74.

Grime JP (1998). Benefits of plant diversity to ecosystems: immediate, filter and founder effects. J. Ecol. 86, 902–910.

Han J and Rundquist DC (1997). Comparison of NIR/RED ratio and first derivative of estimating algal-chlorophyll concentration: A case study in a turbid reservoir. Remote Sensing of Environment, 62:253-261.

Hofer G, Wagner HH, Herzog F, Edwards PJ (2008). Effects of topographic variability on the scaling of plant species richness in gradient dominated landscapes. Ecography, 31, 131–139.

Jones TG, Coops NC, Sharma T (2010). Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada. Remote Sens. Environ. 114, 2841–2852.

Kalacska M, Sanchez-Azofeifa GA, Rivard B, Caelli T, Peter White H, Calvo-Alvarado JC (2007). Ecological fingerprinting of ecosystem succession: Estimating secondary tropical dry forest structure and diversity using imaging spectroscopy, Remote Sensing of Environment, Volume 108, Issue 1, Pages 82-96.

Kaufman YJ and D Tanre (1996). Strategy for Direct and Indirect Methods for Correcting the Aerosol Effect on Remote Sensing: from AVHRR to EOS-MODIS. Remote Sensing of Environment 55:65-79.

Kerr JT and Ostrovsky M (2003). From space to species: Ecol. Appl. for remote sensing. Trends Ecol. Evol. 18, 299–305.

Kumar L, Schmidt K, Dury S & Skidmore A (2001). Imaging spectrometry and vegetation science. In Imaging Spectrometry: Basic Principles and Prospective Applications (eds.). F. D. Van der Meer & S. M. De Jong, pp.111-155. Dordrecht: Kluwer Academic Publishers.

Leutner BF, Reineking B, Müller J, Bachmann M, Beierkuhnlein C, Dech S, Wegmann M (2012). Modelling forest α-diversity and floristic composition—On the added value of LiDAR plus hyperspectral remote sensing. Remote Sens., *4*, 2818–2845.

Levin N, Shmida A, Levanoni O, Tamari H, Kark S (2007). Predicting mountain plant richness and rarity from space using satellite-derived vegetation indices. Divers Distrib 13:692–703

Leyequien E, Verrelst J, Slot M, Schaepmanstrub G, Heitkonig I, Skidmore A (2007). Capturing the fugitive: Applying remote sensing to terrestrial animal distribution and diversity. Int. J. Appl. Earth Obs. Geoinform. 9, 1–20.

Liaw A & Wiener M (2002). Classification and Regression by Random Forest. R news, 2(3), 18-22.

Lindsell JA and Klop E (2013). Spatial and temporal variation of carbon stocks in a lowland tropical forest in West Africa. Forest Ecology and Management, Volume 289, 1 February 2013, pp. 10-17.

Lucas R, Mitchell A, and Bunting P (2008). Hyperspectral Data for Assessing Carbon Dynamics and Biodiversity of Forests. In: Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests. Edited by Margaret Kalacska and G . Arturo Sanchez-Azofeifa. CRC Press

Magurran AE (2004). Measuring Biological Diversity. Blackwell Publishers, Oxford, 260.

Martin ME and Aber JD (1997). High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. Ecol. Appl. **7**: 431–443.

Metzge JP (2000). Tree functional group richness and landscape structure in a Brazilian tropical fragmented landscape. Ecological Applications 10, 1147–1161.

Nagendra H (2001). Using remote sensing to assess biodiversity. International Journal of Remote Sensing 22(12): 2377-2400.

Nagendra H and Rocchini D (2008). High resolution satellite imagery for tropical biodiversity studies: the devil is in the detail. Biodiversity Conservation, 17:3431-3442.

Oldeland J, Wesuls D, Rocchini D, Schmidt M, Jurgens N (2010). Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity? Ecological Indicators, 10, 390-396.

Palmer MW, Earls PG, Hoagland BW, White PS, Wohlgemuth T (2002). Quantitative tools for perfecting species lists. Environmetrics 13, 121–137.

Parmentier I, Harrigan RJ, Buermann W, Mitchard ET, Saatchi S, Malhi Y, ... & Hardy O J (2011). Predicting alpha diversity of African rain forests: models based on climate and satellite-derived data do not perform better than a purely spatial model. Journal of biogeography, 38(6), 1164-1176.

Paoli GD, Wells PL, Meijaard E, Struebig MJ, Marshall AJ, Obidzinski K, Tan A, Rafiastanto A, Yaap B, Ferry Slik J, Morel A, Perumal B, Wielaard N, Husson S, D'Arcy L (2010). Biodiversity Conservation in the REDD. Carbon Balance and Management, 23;5:7.

Rivard B, Sanchez-Azofeifa AG, Foley S, Calvo-Alvarado JC (2008). Species classification of tropical tree leaf reflectance and dependence on selection of spectral bands. In Hyperspectral remote sensing of tropical and subtropical forests, Eds. Margaret Kalacksa and G. Arturo Sanchez-Azofeifa. CRC Press, Boca Raton, FL.

Rocchini D, Chiarucci A, Loiselle SA (2004). Testing the spectral variation hypothesis by using satellite multispectral images. Acta Oecologica 26, 117–120.

Rocchini D, Ricotta C, Chiarucci A (2007). Using remote sensing to assess plant species richness: the role of multispectral systems. Applied Vegetation Science, 10: 325-332.

Rocchini D and Vannini A (2010). What's up? Testing spectral heterogeneity vs. NDVI relationship by quantile regression. International Journal of Remote Sensing, vol. 31 issue 10.

Sellers P J (1985). Canopy reflectance, photosynthesis and transpiration. International Journal of Remote Sensing, 6(8), 1335-1372.

Shannon CE (1948). A mathematical theory of communication. The Bell System Technical Journal, 27, 379-423 and 623-656.

Simonson WD, Allen HD & Coomes DA (2012). Use of an airborne lidar system to model plant species composition and diversity of Mediterranean oak forests. Conservation Biology, 26(5), 840-850.

Sims DA and Gamon JA (2002). Relationships Between Leaf Pigment Content and Spectral Reflectance Across a Wide Range of Species, Leaf Structures and Developmental Stages. Remote Sensing of Environment 81:337-354.

Thenkabail PS, Enclona EA, Ashton MS, Legg C and De Dieu MJ (2004a). Hyperion, IKONOS, ALI and ETM plus sensors in the study of African rainforests. Remote Sensing of Environment, 90, pp. 23-43.

Thenkabail PS, Enclona EA, Ashton MS, & Van Der Meer B (2004b). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. Remote sensing of environment, 91(3), 354-376.

Thomas CD, Cameron A, Green RE, et al. (2004). Extinction risk from climate change, Nature, 427, 145.

Townsend AR, Asner GP, Cleveland CC (2008). The biogeochemical heterogeneity of tropical forests. Trends Ecol Environ 43(8):424–431.

Tsai F and Philpot W (1998). Derivative analysis of hyperspectral data. Remote Sensing of Environment, 66:41-51.

Ustin SL, Roberts DA, Gamon JA, Asner GP, & Green RO (2004). Using imaging spectroscopy to study ecosystem processes and properties. BioScience, 54, 523– 534.

Vaglio Laurin G, Liesenberg V, Chen Q, Guerriero L, Del Frate F, Bartolini A, Coomes D, Wilebore B, Lindsell J, and Valentini R (2013). Optical and SAR sensor synergies for forest and land cover mapping in a tropical site in West Africa. International Journal of Applied Earth Observation and Geoinformation, 21, 7-16.

Vogelmann JE, Rock BN, Moss DM (1993). Red Edge Spectral Measurements from Sugar Maple Leaves. International Journal of Remote Sensing 14:1563-1575.

Whitmore TC (1990). An introduction of tropical rain forests. Clarendon, Oxford, England.

Zhang J, Rivard B, Sanchez-Azofeifa A, Castroesau K (2006). Intra- and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: implications for species identification using HYDICE imagery. Remote Sensing of Environment 105 (2), 129–141.

Tables:

Table 1: list of most common trees (by individuals' number and % over all trees) found in the study area

Species name	# of trees	% of trees	Family name
Heritiera utilis	61	9.0	*Malvaceae*
Protomegabaria stapfiana	57	8.4	*Phyllanthaceae*
Cynometra leonensis	50	7.4	*Caesalpinioideae*
Brachystegia leonensis	28	4.1	*Caesalpinioideae*
Gilbertiodendron bilineatum	28	4.1	*Caesalpinioideae*
Stachyothyrsus stapfiana	24	3.6	*Caesalpinioideae*
Phyllocosmus africanus	20	3.0	*Ixonanthaceae*
Xylopia quintasii	18	2.7	*Annonaceae*
Parinari excelsa	18	2.7	*Chrysobalanaceae*
Calpocalyx brevibracteatus	16	2.4	*Mimosoideae*
Sacoglottis gabonensis	14	2.1	*Humiriaceae*
Octoknema borealis	13	1.9	*Olacaceae*
Uapaca guineensis	13	1.9	*Euphorbiaceae*
Berlinia confusa	10	1.5	*Caesalpinioideae*
Bussea occidentalis	10	1.5	*Caesalpinioideae*
Total	**380**	**56.2**	

Table 2: Random Forests models results using the three input sets.

Random Forests Out-of-bag estimates	
	Shannon index
Hyperspectral band reflectance metrics	Pseudo R^2 = 84.91%, OOB-RMSE = 0.30
First derivatives reflectance metrics	Pseudo R^2 = 71.42%, OOB-RMSE = 0.35
Vegetation indices	Pseudo R^2: -15.97%, OOB-RMSE = 0.51

Figure legends:

Figure 1: the Gola Rainforest National Park, the study area in Sierra Leone. Strips of hyperspectral data collected over the area are shown, as well as location of field plots.

Figure 2: example of one of the strips of hyperspectral data (in false-color composite at 807.5 (R), 597.3 (G) and 467.3 (B) nm) with overlapped field plots in yellow.

Figure 3: species-area curve.

Figure 4: scatterplots of predicted versus expected values obtained by model based on hyperspectral reflectance band metrics (left side) and first derivatives reflectance metrics (right side).

Fig. 5: ranking of hyperspectral metrics, with maximum, minimum, mean, standard deviation of band reflectance in the four different frames of the figure. The y-axis represents the percentage in increase in MSE and the x-axis is the band region (in nm).

Fig. 6: ranking of derivative metrics, with maximum, minimum, mean, standard deviation of first derivatives of band reflectance in the four different frames of the figure. The y-axis represents the percentage in increase in MSE and the x-axis is the band region (in nm).
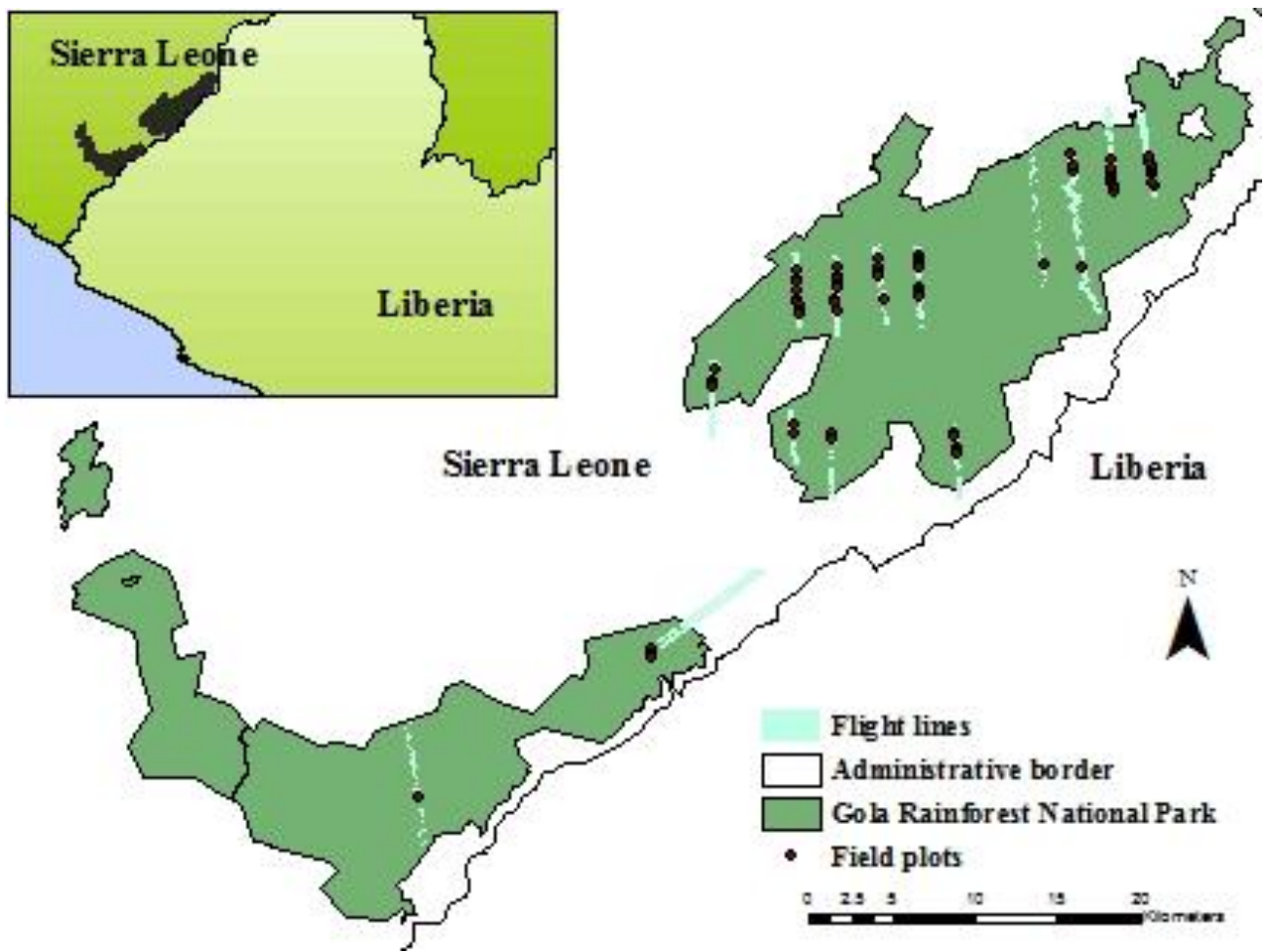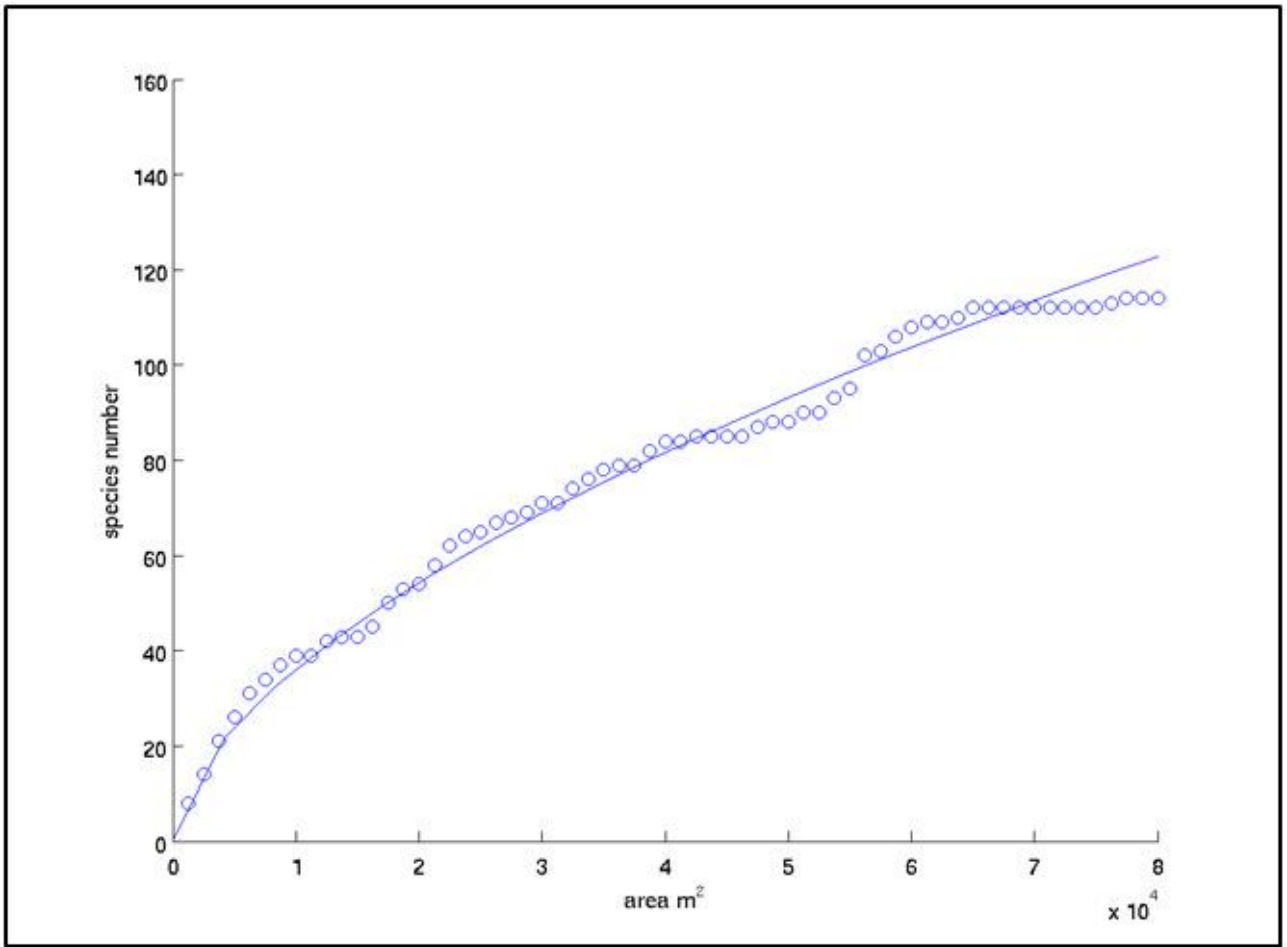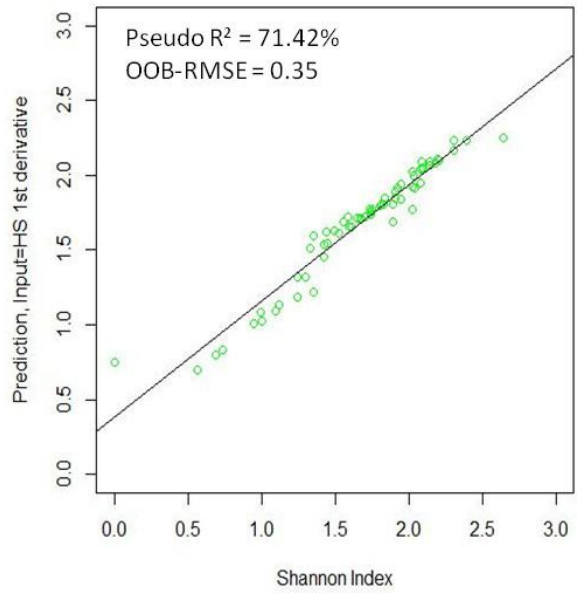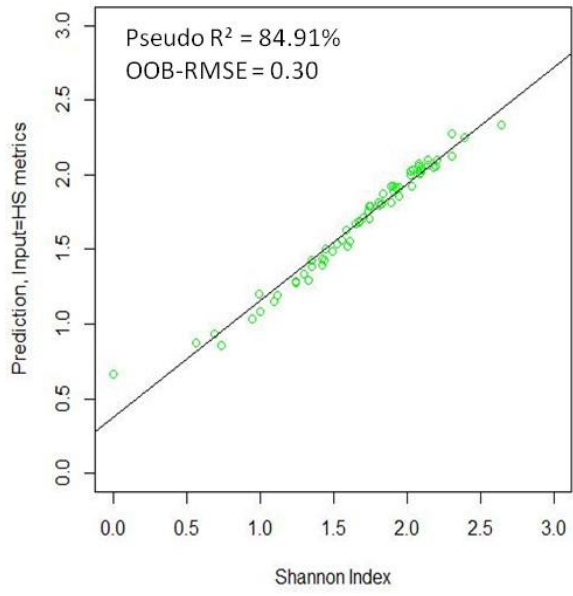
Fig.1

Fig. 2

Fig. 3

Fig. 4

Chapter 8

Research summary

This thesis focuses on two main topics in forest research: classification and modeling of forest parameters applying remote sensing data. Information provided by reliable land cover map is fundamental for sustainable management of forest resources. The statistical modeling of forest parameters (e.g. biomass and biodiversity), on the other hand, is required to assess the forest resources more efficiently, and coupled with remote sensing data it allows to gather information over large areas.

The main contributions of the present work include:

1. The application of innovative techniques and data to improve land cover mapping, above ground biomass estimation, and biodiversity estimation.
2. The adoption of advanced statistical modeling tools and classification algorithms to deal with complex retrieval and classification problems.

In the first case, the integration of data from active and passive sensors has proven to increase the accuracies in the different case studies, in which multispectral, hyperspectral, SAR and LiDAR datasets were used. This integration confirmed that data fusion is probably the best approach to capture the variability of complex forest environments. In the second case, the use of advanced algorithms not only improved the tests results, but allowed the joint use of different sensors, from which a large amount of information was extracted; it also allowed ranking the features extracted from remote sensing data with respect to their usefulness in solving classification and estimation problems.

8.1 Challenges addressed

According to the results obtained from this study the research challenges listed in Chapter 2 are addressed as follows:

1. Ancillary data usefulness in AGB LiDAR-based estimations

The case study illustrated in Chapter 5 investigates the estimation of AGB in a temperate forest in Sierra Nevada, California, in a biome in which LiDAR already proved to be a very accurate instrument for this task (Lefsky et al. 2002). In such studied areas, many different vegetation information layers are available thanks to government efforts devoted to improve the conservation and management of these forests. The challenge in this specific study was therefore to examine innovative strategies to further improve the ability to estimate biomass with remote sensing tools. An accurate literature review showed that the relationship between LiDAR height metrics and AGB is influenced by vegetation type (Drake et al. 2003; Naesset and Gobakken 2008) and this information, at the fine scale needed considering the extent of the area under examination, is often available in protected or managed areas of developed countries. Specifically, in the area of interest aerial photographs were collected and used to generate a State detailed vegetation map. Orthophotos are often available in the European areas too, as well as in other non tropical countries, and detailed land cover maps can also be found over specific areas such as reserves or where key natural resources are found. However, this potentially useful information has never been used as ancillary data for generating biomass estimations. The following step was then identify the best way to integrate categorical information, in the case study a vegetation map, into a retrieval algorithms, avoiding the development of distinct biomass models for each vegetation type. In fact, in forestry research the field validation dataset is always limited due to the huge resources needed for its collection. Splitting the field data into subsets belonging to each vegetation type is to be avoided, as a reduced number of samples negatively affect the robustness of the statistics behind the models. A solution to this issue was found by means of MEM, which attribute random effects to vegetation types while developing unique biomass estimation for the area of

interest. The case study demonstrated that the incorporation of ancillary information in MEM, compared to the use of LiDAR data alone in multiplicative models, resulted in an increase of the R^2 of the biomass estimation from 0.77 to 0.83 with RMSE (root mean square error) reduced by 10% (from 80.8 to 72.2 Mg/ha).

2. Ancillary data (LiDAR-derived) usefulness in discriminating vegetation types

Ancillary data are known to be useful in land cover mapping at medium resolution scale, as exemplified by the use of terrain elevation data in the CORINE program of the EU (Buttner et al. 2002). Vegetation type mapping can be a challenging task due to spectral similarities among vegetation, and has to be conducted at a spatial resolution compatible with the size of observed objects. For the peculiar case of land cover mapping of Chapter 4, involving the detection of fine differences in vegetation, the ancillary information generated from traditional remote sensing tools has been rarely used. The case study illustrated in Chapter 3 introduces an innovative way to exploit LiDAR information together with SAR data, therefore two unconventional tools for vegetation resources mapping. With respect to more traditional multispectral sensors, both instruments present advantages. LiDAR by-products, in this case the Canopy Height Model, provide height data which can be important in the detection of specific vegetation types. LiDAR-derived products are often available from government data collection efforts, as exemplified in Italy, in which all the coastal strip and some inland areas have been surveyed with LiDAR, with datasets and by-products made available through the Italian Ministry of Environment Geoportal website (http://www.pcn.minambiente.it/GN/index.php?lan=en, accessed on November 11, 2013). SAR data offer the advantage of being insensitive to cloud cover and can provide information on volumetric scattering which is in relationship to the above ground biomass content of vegetation (Imhoff 1995). In this case study, the integration of multiple data not only provides a mapping accuracy superior to the one obtained by means of high resolution multispectral data (SPOT 5), but allows to better distinguish vegetation types which are less accurately identified by the optical dataset. In fact, the overall accuracy obtained by joint datasets reached 97.7%, against 93.0% obtained with

SPOT 5 data; separability analysis indicated a better performance of joint datasets in the distinction of conifer forests from dwarf pines. Neural networks, an advanced machine learning tool, allow the successful integration of the two different datasets providing classification accuracy significantly superior to the widespread Maximum Likelihood classification algorithm. Therefore, this study identifies an innovative way to solve a complex vegetation classification problem, making full usage of the most advanced remote sensing dataset and data exploitation techniques.

3. Data fusion: evaluating the benefits of optical and RADAR sensors integration for tropical land cover classification

Tropical forested landscapes present different challenges with respect to their classification, including the presence of numerous vegetation classes characterized by spectral similarities, the similarities between natural vegetation and tree plantations, the high fragmentation of land cover types creating a mosaic of classes, the remoteness of the areas and related constraints for field data collection, and the atmospheric adverse conditions which negatively affect the acquisition of quality optical remote sensing data. On the other hand, detailed land cover maps of tropical regions are needed to support local management, international efforts such as REDD+, and to better understand the role of tropical forests for climate change science (Valentini et al. 2013). While optical data have been for decades the primary data source for land cover mapping activities, in recent years SAR sensors have emerged as important tools in vegetation studies (Rahman and Sumantyo 2010). Not only SAR data are insensitive to persistent cloud cover affecting tropical regions (Lu et al. 2007), but they also provide volumetric scattering information which is complementary to the canopy reflectance data provided by optical sensors. For instance two distinct vegetation type can have similar reflectance response but very different volumes, as shown in the case of Alpine vegetation in previous paragraph. Therefore, the joint use of these two sensors is highly recommended in challenging land cover tasks (Lefsky and Cohen, 2003). The tropical land cover mapping activity illustrated in Chapter 5 is an example of successful integration of two different optical data (Landsat TM and ALOS AVNIR), characterized

by different spatial resolutions, with the ALOS PALSAR L-band data. In this area cloud cover is so persistent that, over approximately a 10-year period, only two scenes were found in archive, and both presented regions of isolated clouds that prevent full coverage of the area of interest. The SAR-optical data fusion significantly improved the classification accuracy of the area with respect to single data use, reaching the highest overall accuracy (97.5% with AVNIR as optical data and 92.7% with Landsat TM as input) allowing the detection of different successional forest stages and the distinction of semi-deciduous from moist forest type. The use of texture analysis, a well-known technique in image processing (Berberoglu et al. 2007; Chica-Olmo and Abarca-Hernandez 2000), confirmed this as an important technique to improve classification accuracy, especially with respect to the identification of tree plantations having a regular geometry. A semi-automatic method for the selection of the numerous features derived from optical and SAR data processing was also identified. The use of Neural Networks algorithm optimized the integration of the two sensors and brought a significant increase in classification accuracy with respect to the traditional Maximum Likelihood classification approach. Finally, the SAR data single use, even if it did not allow resolving ambiguities among closely related forest types, was also successful in indentifying main land cover classes. The obtained overall accuracy equal to 78.1% confirms the importance of this data type in case of optical data unavailability, also in view of the biomass-related information that SAR can provide.

4. Data fusion: evaluating the integration of LiDAR and hyperspectral sensors for AGB estimation

While LiDAR sensor proved to be very useful in estimating AGB in certain regions and ecosystems, its application in African environments has been very limited and mainly confined to the satellite GLAS instrument (Baccini et al. 2008; Mitchard et al. 2012). A recent meta-analysis on LiDAR-based biomass estimation revealed a broad range of accuracy values for tropical forests studies (Zolkos et al. 2013). Therefore, assessing LiDAR small footprint ability to retrieve biomass in African tropical forests is a priority, also in view of carbon mitigation strategies on-going in the continent. While

LiDAR provides structural forest information, hyperspectral data record canopy spectral information that is potentially related to AGB, by means of functional plant differences which in turn affect species parameters, such as wood density and biomass content (Baker et al. 2004). Hyperspectral data are also a powerful tool to stratify vegetation types prior to the estimation of its biophysical parameters. New satellite missions, which are planned for the very next few years, will provide innovative data to foster ecological monitoring, and additional efforts to fully understand the role hyperspectral data can have in environmental applications are therefore requested. The case study illustrated in Chapter 6 is an example of innovative data fusion, involving small footprint LiDAR and very high spatial resolution hyperspectral imagery from an airborne survey conducted over a West African moist tropical forest. The limited improvement in estimation accuracy obtained with hyperspectral data addition to LiDAR metrics (0.7 R^2 versus 0.64 obtained with LiDAR only) is discussed in the context of the scarce available research regarding the joint use of these two datasets for AGB estimation. Sources of uncertainty are also identified, and innovative approaches to preprocess and extract data from the vast amount of available information are also used. Specifically, the use of Partial Least Squares Regression allowed the simultaneous use of approximately 766 collinear features as regression inputs, generated from statistics calculated out of the 186 hyperspectral bands and LiDAR point cloud. PLSR results outperformed those obtained by a traditional multiplicative power model (0.7 R^2 versus 0.57 obtained with power model) and demonstrated its usefulness in integrating very complex datasets. The scores calculated by means of the Variable Importance in the Projection, a tool available for PLSR results and applied in Chapter 6, identified the LiDAR and hyperspectral most important features for the biomass model. This helped in better understanding the biomass-height relationship for this African forest and to relate the hyperspectral information to foliar biochemistry.

5. Evaluating the impact of field data geolocation in LiDAR-based AGB estimates

One of the main problems and uncertainty causes encountered in the modeling of forest parameters by means of remote sensing is related to the availability and quality of field data, which are used to train, validate and test the models. In the different case studies

here presented the importance of quality field data is stressed, especially with respect to geolocation. The adoption of a fine scale of analysis, made possible by airborne data, further sharpens the geolocation accuracy problem. The delay of GPS signal due to tree cover is a problem which can partially be mitigated by the use of differential GPS technique. But sometimes, as in the case of our African area, the network of GPS bases needed for differential correction is not available, and surrogate base stations have to be set up using two GPS units, causing additional operational problems. The dimension of field plots is also an important factor in view of increasing geolocation accuracy, as larger plots better accommodate positional errors, besides reducing field biomass variability. The establishment of plots near features recognizable from aerial ortho-photographs is another option, which can help in the co-registration of remote sensing and field data. Use of professional GPS units is also recommended.

This research discusses the geolocation accuracy problem in the context of three different cases of estimation of forest parameters: for the AGB and biodiversity modeling in Gola Rainforest National Park (Chapters 6 and 7), and for biomass estimation in Sierra Nevada (Chapter 5). In the first two cases, a precise quantification of the error introduced in the estimates by geolocation poor accuracy was not possible, but evidences such as those visible in aerial photographs indicated this as a primary source of uncertainty. In Sierra Nevada the simultaneous presence of two GPS measurements, one taken with a recreational unit and the other obtained by means of differential correction technique, allowed the exact quantification of the error introduced by poor field plot geolocation, which caused a decrease of R^2 by 0.1-0.13 and an increase in RMSE of 21-31%. Finally, as evidenced in the hyperspectral and LiDAR fusion case study in Chapter 6, geolocation accuracy problems can interest not only field data, but also remote sensing data itself: a mismatch in the order of 1-4m was observed between LiDAR and hyperspectral, thus being another source of error and an area of possible improvement in data collection planning.

8.2 Conclusions

Overall, spanning over different ecosystems and research problems, this study offered to me the opportunity of appreciating the difficulties, challenges and future directions in forest and vegetation monitoring and research based on remote sensing data. A number of recommendations emerged from single case studies, which could be potentially useful in other forest researches:

1. Maximize the use of ancillary information, which is sometimes available through other research or monitoring activities, also as a way to foster collaboration and data sharing.

2. Use advanced algorithms, which are demonstrated to improve results and which may better deal with new generation remote sensing data, characterized by increased spatial and spectral resolutions or polarizations.

3. Combine multi-source remote sensing data by means of data fusion, as different sensors can contribute to bring valuable and complementary information over the target.

4. Apply advanced image processing techniques to remote sensing data, thus improving preprocessing steps such as noise reduction in SAR and hyperspectral data, or feature extraction, such as in the case of texture analysis in classification studies.

5. Improve field data collection in quality and quantity, to better represent the vegetation variability and improve modeling, taking into account the spatial resolution of the remote sensing data being used.

6. Increase the use of SAR satellite data, which are an invaluable tool in tropical areas affected by persistent cloud cover.

While the collection of proper field data remains a difficult point to address, especially in remote regions, my future research will keep on focusing on data fusion and the application of innovative techniques and algorithms to deal with complex and multiple datasets. Datasets from forthcoming missions (i.e. ALOS2, Sentinel-1, Sentinel-2) will soon provide new data to exploit and evaluate, hopefully having improved capability to provide information for forestry research.

8.3 References

Baccini A., Laporte N., Goetz S. J., Sun M., & Dong H., 2008. A first map of tropical Africa's above-ground biomass derived from satellite imagery. Environmental Research Letters, 3(4), 045011.

Baker T.R., Phillips O.L., Malhi Y., Almeida S., Arroyo L., Di Fiore A., Erwin T., Higuchi N., Killeen T.J., Laurance S.G., Laurance W.F., Lewis S.L., Lloyd J., Monteagudo A., Neill D.A., Patino S., Pitman N.C.A., Silva J.N.M., Vasquez Martinez R., 2004. Variation in wood density determines spatial patterns in Amazonian forest biomass. Global Change Biology 10:545–562.

Berberoglu S., Lloyd C.D., Atkinson P.M., Curran P.J., 2007. Texture classification of Mediterranean land cover. International Journal of Applied Earth Observation and Geoinformation 9 (3), 322–334.

Büttner G., Feranec J., Jaffrain G., Steenmans C., Gheorghe A., & Lima V., 2002. Corine land cover update 2000. Technical guidelines. Copenhagen, Denmark: European Environment Agency.

Chica-Olmo M., Abarca-Hernandez F., 2000. Computing geostatistical image texture for remotely sensed data classification. Computers and Geosciences 26 (4), 373–383.

Drake J.B., Knox R.G., Dubayah R.O., Clark D.B., Condit R., Blair J.B., & Hofton M., 2003. Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: Factors affecting the generality of relationships. Global Ecology & Biogeography, 12, 147–159.

Imhoff M. L., 1995. A theoretical analysis of the effect of forest structure on synthetic aperture radar backscatter and the remote sensing of biomass. Geoscience and Remote Sensing, IEEE Transactions on, 33(2), 341-352.

Lefsky M.A., Cohen W.B., Harding D.J., Parker G.G., Acker S.A., & Gower S.T., 2002. Lidar remote sensing of above ground biomass in three biomes. Global ecology and biogeography, 11(5), 393-399.

Lu D., Batistella M., Moran E., 2007. Land-cover classification in the Brazilian Amazon with the integration of Landsat ETM+ and RADARSAT data. International Journal of Remote Sensing 28 (24), 5447–5459.

Mitchard E.T.A., Saatchi S.S., White L.J.T., Abernethy K.A., Jeffery K.J., Lewis S.L., ... & Meir P., 2012. Mapping tropical forest biomass with radar and spaceborne LiDAR in Lopé National Park, Gabon: overcoming problems of high biomass and persistent cloud. Biogeosciences, 9, 179-191.

Næsset E., & Gobakken T., 2008. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. Remote Sensing of Environment, 112, 3079–3090.

Rahman M.M. and Sumantyo J.T.S., 2010. Mapping tropical forest cover and deforestation using synthetic aperture radar (SAR) images. Applied Geomatics 2 (4),113–121.

Valentini R., Arneth A., Bombelli A., Castaldi S., Cazzolla Gatti R., Chevallier F., Ciais P., Grieco E., Hartmann J., Henry M., Houghton R.A., Jung M., Kutsch W. L., Malhi Y., Mayorga E., Merbold L., Murray-Tortarolo G., Papale D., Peylin P., Poulter B., Raymond P.A., Santini M., Sitch S., Vaglio Laurin G., van der Werf G.R., Williams, C.A., and Scholes R.J., 2013. The full greenhouse gases budget of Africa: synthesis, uncertainties and vulnerabilities, Biogeosciences Discuss., 10, 8343-8413.

Zolkos S.G., Goetz S.J., & Dubayah R., 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. Remote Sensing of Environment, 128, 289-298.

Appendix I

Curriculum Vitae and publications list

Curriculum Vitae

Name	**Gaia Vaglio Laurin**
Address	Via Marin Sanudo 4, Roma 00176, Italy
Mobile	+328 4117518
e-mail	gaia.vl@tiscali.it
Date and Place of Birth	Rome, 26th May 1967
Nationality	Italian
Marital status	Unmarried

Education

Jan. 2014	Geoinformation PhD, University of Tor Vergata, Computer Science and Civil Engineering Department, Rome, Italy
June 2004	Istituto Agronomico per l'Oltremare (Italian Ministry of Foreign Affairs): Master Course in **Remote Sensing/GIS** and Natural Resources Evaluation.
Dec. 2004	University of London: Master's Degree in **Biodiversity Conservation**, Imperial College at Wye, UK.
June 2002	Post-graduate course in **Communication** and Marketing **for the Environment**, Ateneo Impresa, Rome, Italy.
Dec. 1996	Master in **Demo-Ethno-Anthropological Sciences**, University of Rome 'La Sapienza', Italy.
Nov. 1997	Professional qualification for Biologists (National Exam).
May 1995	*'Laurea'*(Master equivalent) in **Biological Sciences**, University of Rome 'La Sapienza', Italy. Note: 110/110. Ecology major.

Languages

Italian (mother tongue)
English (proficiency)
Spanish (proficiency)
Portuguese (good)
French (basic)

Computer skills

MS Office; GIS and Image processing (ArcView, ArcGIS, MapInfo, ENVI, ERDAS, eCognition); Matlab.

Other

Excellent communication skills, very good in partnering and team work, result-oriented, scientific, management and fund raising capability.

References

Riccardo Valentini, La Tuscia University, Italy
Qi Chen, University of Hawaii, USA
David Coomes, University of Cambridge, UK
Luca Ongaro, IAO, Ministry of Foreign Affairs, Firenze, Italy

Professional experience

December 2012 – present
CMCC centro euromediterraneo sui cambiamenti climatici: Earth Observation scientist at IAFENT division. Main duty: FP7EU ERC grant AFRICA GHG for remote sensing of vegetation component.

February 2010 – present
Cambridge Conservation Initative: consultant in the two CCI grants in favour of Gola Rainforest National Park, Sierra Leone. Tasks: Remote sensing of vegetation.

December 2010 – May 2011
CMCC centro euromediterraneo sui cambiamenti climatici: consultant in the EU ERC grant AFRICA GHG. Remote sensing of vegetation.

Dec. 2010 –April 2011 and Dec. 2012 – March 2013
University of Hawaii at Manoa, Honolulu, Hawaii, USA: visiting scholar at the Geography Department, Qi Chen laboratory, lidar analysis for forest research.

Nov. 2009 – Oct. 2012
University of Tor Vergata, Rome, Italy: PhD in Geoinformation, with scholarship, Computer Science and Civil Engineering Department, Tor Vergata University, Rome, Italy

Feb 2007 – March 2009
Planetek srl: consultant technical and management officer for GMES - Land projects (land cover classification in EU), ESA and Regione Veneto funding. *Duties*: user relationships, project management, review of object oriented algorithms applied to land cover classification, image classification (SPOT5), training, reporting, GIS mapping.

July 2005 – June 2008
Advanced Computer Systems spa: consultant technical and management duties for 'Desertwatch' project (desertification monitoring - http://dup.esrin.esa.it/desertwatch/), technical leader for WADE-Tiger project (radar remote sensing for water monitoring in Sahelian Africa - http://www.tiger.esa.int/) both funded by ESA. *Duties*: end users contacts, project management, review of algorithms, image classification, training, reporting, GIS mapping.

November 2004 – June 2005:
UN FAO Consultant for the Global Land Cover Network, SRDN: Technical Support Unit at the Land Cover Thematic Centre based in Florence at IAO. Duties: image interpretation, land cover mapping methodological review, communication contents, e-learning tools, outreaching, GIS.

January 2002 – June 2004:
Consultant for "Strategic Environmental Assessment (VAS) of the Chieti Province" and for "Research and experimentation on inventory, cataloguing, restoration and re-evaluation of the transhumant ways and lands", **Italian Ministry of Environment/University of Chieti**, Italy. Duties: GIS, remote sensing, environmental planning.

March 1998 – May 2002:
Coordinator of the " Sustainable Development and Environment" sector/projects, Italian **NGO Terra Nuova**. In charge of co-operation programs in developing countries (East and West Africa, Central and Latin America). Duties: project planning, design, management, evaluation and follow up. Management of funds and relations with donors and different institutions. Training of personnel. Fields of activities: development, desertification fight, poverty reduction, environmental management, eco-tourism, natural resources and biodiversity conservation, environmental policies and strategies, environmental education. Based in Terra Nuova international offices (Rome, Nairobi, Bamako, Managua, Salvador da Bahia) and in the field.

September 1999 – November 2001:
Project Leader of "Esplorambiente", **environmental education** projects for italian high schools, co-funded by the Italian Ministry of Foreign Affairs (MAE).

July – August 2001:
Expert in ecology, Nicaragua. European Commission (EC) – co-funded project "Biodiversidad y Humedales en àreas indìgenas" B7-6200/00-09/DEV/ENV.

November 1997 – February 1998
Expert in protected areas, Brazil. Italian Ministry of Foreign Affairs (MAE) - co-funded project 1775/COCIS/BRA.

Design and management of many environmental/development projects in the agriculture, rural development, conservation domain, funded by the EC and development governmental agencies, including:
❖ Ecoscambio, Brazil. Conservation and sustainable development in 5 protected areas in Brazil.
❖ Los Pantanos de Villa, Lima. Management and conservation of Peru Wetlands.
❖ Biodiversidad y humedales en àreas indìgenas de Centro America (Panama, Costa Rica, Nicaragua, Honduras).
❖ Conservation and sustainable development in a Ramsar site in Mali (Lac Debo et Walado Debo), Mali.
❖ Community-based conservation in Tanzania (Rungwa Game Reserves) and Kenya (Garissa District).
❖ Pastoral development and natural resources management in Turkana District, Kenya.
❖ Sustainable use of natural resources along the Mata Atlantica (Brasil).
❖ Trans-boundary Environmental Project: conservation of natural resources and sustainable development in pastoral semi-arid regions of Eastern Africa.
❖ Supporting *in-situ* and *ex-situ* phytogenetic resources conservation in Cuba: the Germoplasm Bank.
❖ Tiger Innovator: Water monitoring in West Africa by remote sensing.

Main Publications

Valentini, R., Arneth, A., Bombelli, A., Castaldi, S., Cazzolla Gatti, R., Chevallier, F., Ciais, P., Grieco, E., Hartmann, J., Henry, M., Houghton, R. A., Jung, M., Kutsch, W. L., Malhi, Y., Mayorga, E., Merbold, L., Murray-Tortarolo, G., Papale, D., Peylin, P., Poulter, B., Raymond, P. A., Santini, M., Sitch, S., Vaglio Laurin, G., van der Werf, G. R., Williams, C. A., and Scholes, R. J., **2013**. The full greenhouse gases budget of Africa: synthesis, uncertainties and vulnerabilities, Biogeosciences Discuss., 10, 8343-8413.

Gaia Vaglio Laurin, Fabio Del Frate, Luca Pasolli, Claudia Notarnicola, Leila Guerriero, Riccardo Valentini. **2013**. Discrimination of vegetation types in alpine sites with alos palsar, radarsat2 and lidar derived information. International Journal of Remote Sensing 34(19), 6898-6913.

Vaglio Laurin, G., Liesenberg V., Chen Q., Guerriero L., Del Frate F., Bartolini A., Coomes D., Wilebore B., Lindsell J., and Valentini R., **2013**. Optical and SAR sensor synergies for forest and land cover mapping in a tropical site in West Africa , International Journal of Applied Earth Observation and Geoinformation, 21, 7-16.

Chen, Q., G. Vaglio Laurin, J. Battles, and D. Saah, **2012**. Integration of airborne lidar and vegetation types derived from aerial photography for mapping aboveground live biomass. Remote Sensing of Environment, 121, 108-117.

Lu, D., Q. Chen, G. Wang, E. Moran, M. Batistella, M. Zhang, G. Vaglio Laurin, and D. Saah, **2012**. Aboveground forest biomass estimation with Landsat and LiDAR data and uncertainty analysis of the estimates. International Journal of Forestry Research, doi:10.1155/2012/436537.

L. Pasolli, C. Notarnicola, L. Bruzzone, G. Bertoldi, S. Della Chiesa, V. Hell, G. Niedrist, U. Tappeiner, M. Zebisch, F. Del Frate, and G. Vaglio Laurin. Estimation of Soil Moisture in an Alpine Catchment with RADARSAT2 Images. Applied and Environmental Soil Science, Volume **2011**, Article ID 175473, 12 pages

Gaetano Pace, Gaia Vaglio Laurin, Lucio Pires Do Rosario, Maurizio Sciortino. Space for the UNCCD and the Desertwatch project. 2008 In: Desertification and Risk Analysis Using High and Medium Resolution Satellite Data Training Workshop on Mapping Desertification. Book Series NATO Science for Peace and Security Series C: Environmental Security. Springer Netherland.

A. Arledler, P. Castracane, A. Marin, S.Mica, G.Pace, M.Quartulli, G. Vaglio Laurin, I.Alfari, H. Trebossen. 2007. Detecting water bodies and water related features in the Niger basin area by SAR data: the ESA TIGER WADE project. UNESCO's technical documents on Hydrology.

G. Vaglio Laurin and L. Ongaro, 2006. Mapping the suitability of potential conservation sites: a case study in Northern Tunisia. Journal of Agriculture and Environment for International Development, 2006, 100 (1/2): 3-28.

Fabietti W., G. Vaglio Laurin, 2004. 'Censimento delle risorse naturali'. In: Antiche infrastrutture per uno sviluppo compatibile. Pp. 118-124. Università di Chieti-Pescara, Università del Molise, Ministero dell'Ambiente. Di Erre edizioni.

Salvatori V., G. Vaglio Laurin, , P.L. Meserve, L. Boitani & A. Campanella. 1999. Spatial organization, activity, and social interactions of culpeo foxes (*Pseudalopex culpaeus*) in north-central Chile. Journal of Mammalogy,80 (3):980-985.

Conferences

Gaia Vaglio Laurin, Qi Chen, Jeremy Lindsell, David Coomes, Roberto Cazzolla-Gatti, Elisa Grieco, Riccardo Valentini, 2013. Above ground biomass estimation from lidar and hyperspectral airbone data inWest African moist forests. Geophysical Research Abstracts Vol. 15, EGU2013-6227-1, 2013. EGU General Assembly 2013

G Vaglio Laurin, F Del Frate, L Pasolli, Notarnicola C. 2012 Forest/vegetation types discrimination in an alpine area using RADARSAT2 and ALOS PALSAR polarimetric data and Neural Networks. Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International.

Luca Pasolli, Claudia Notarnicola, Lorenzo Bruzzone, Giacomo Bertoldi, Georg Niedrist, Ulrike Tappeiner, Marc Zebisch, Fabio Del Frate, Gaia Vaglio Laurin, 2011. Spatial and temporal mapping of soil moisture content with polarimetric RADARSAT 2 SAR imagery in the Alpine area. IGARSS 2011: 1067-1070

F. Del Frate, G. Vaglio Laurin, C. Notarnicola, "SOFIA, SOil and Forest Information Retrieval by using RADARSAT2 images: first results of the project", 2010, in Proceedings ESA Living Planet Symposium. Bergen, Norway, 28 Jun.– 2 Jul 2010.

G. Vaglio Laurin, Q. Chen, D. Coomes, F. Del Frate, G. A. Licciardi, J. Lindsel, 2010. "Mapping the biomass of Gola forest reserve in Sierra Leone with remote sensing data and neural networks", Santiago de Compostela, Spain, FORESTSAT Sept. 2010.

L. Pasolli, C. Notarnicola, L. Bruzzone, G. Bertoldi, M. Zebisch, F. Del Frate, G. Vaglio Laurin, 2010 "Analysis of polarimetric RADARSAT2 images for soil moisture retrieval in an alpine catchment", in Proceedings of SPIE Conference on SAR Image Analysis, Modeling, and Techniques. Toulouse, France.

A. Di Gregorio, P. Sarfatti, G. Vaglio Laurin, 2005. The Global Land Cover Network: innovation and experiences in land cover mapping for ARD and land use planning. In: Abstracts of parallel sessions, poster presentations, EFARD 2005, 27-29 of April, Swiss Federal Institute of Technology, Zurich.