International Journal of Applied Earth Observation and Geoinformation 18 (2012) 82–90

Contents lists available at SciVerse ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Modifying geographically weighted regression for estimating aboveground biomass in tropical rainforests by multispectral remote sensing data

Pavel Propastin[a,b,*]

[a] Institute of Geography, Georg-August University Göttingen, Goldschmidtstr. 5, 37077 Göttingen, Germany
[b] Department of Bioclimatology, Büsgen-Institute, Georg-August University Göttingen, Büsgenweg 2, 37077 Göttingen, Germany

ARTICLE INFO

Article history:
Received 24 May 2011
Accepted 17 December 2011

Keywords:
Geographically and altitudinal weighted regression
Spatial non-stationarity
Aboveground biomass
Tropical rainforest
Vegetation index
Sulawesi

ABSTRACT

The present study uses a local regression approach for estimation of aboveground biomass (AGB) in a tropical rainforest area with highly diverse terrain conditions from remote sensing-based multi-spectral vegetation indices (VI). By incorporating altitudinal effects into the spatial weighting matrices of the common geographically weighted regression (GWR), an extended GWR model, geographically and altitudinal weighted regression (GAWR), has been developed to deal with both spatial (horizontal) and altitudinal (vertical) non-stationarity in the data set. Unlike the common GWR model, the presented GAWR approach captures both horizontal and altitudinal drifts in the relationships between aboveground biomass and remote sensing data. In order to test its improved performance, the GAWR method was compared with the traditional GWR technique and global ordinary least squares regression (OLS) in a region of mountainous tropical rainforest in Sulawesi, Indonesia. The relationships between AGB and VIs were found to be significantly spatially variable. The results showed that there were substantial benefits in capturing both horizontal and vertical non-stationarity simultaneously. The GAWR method significantly improved AGB prediction in all simulations relative to both the traditional GWR and OLS methods, as indicated by accuracy and precision statistics. From the results of empirical tests, it seems proper to say that for this data set, the GAWR model is better than the traditional GWR model.

1. Introduction

Several studies have used different kinds of remotely sensed data for forest inventories and the spatial modelling of biophysical vegetation parameters in various regions around the world (among them: Muukkonen and Heiskanen, 2005; Gonzalez-Alonso et al., 2006; Rosenqvist et al., 2003). These and other related studies have proven the generally good suitability of the remote sensing data for estimation of different forest attributes at national and over-national level (Dong et al., 2003; Tomppo et al., 2002; Sivanpillai et al., 2007). Nonetheless, the remote sensing-based mapping of biomass in tropical regions is a challenging task with many problems (Clark et al., 2011; Steiniger, 2000; Foody et al., 2001; Lu, 2006; Boyd and Danson, 2005).

The common approach to model biophysical variables from satellite data is the statistical regression procedure, where the quantitative relationship between spectral reflectance in satellite bands and ground-based data is calibrated by interrelating known coincident observations of remotely sensed and ground data. Empirical models, which relate multi-spectral satellite data to field observations of aboveground biomass, have proven their effectiveness at different scales all over the world (Steiniger, 2000; Lu et al., 2004; Zheng et al., 2004). Models based on regression analysis are considered to be relatively easy to implement and can provide accurate results by application at all spatial scales. They have already demonstrated their effectiveness in numerous studies throughout the world (Muukkonen and Heiskanen, 2005; Tomppo et al., 2002; Gonzalez-Alonso et al., 2006). Nonetheless, the common regression approach has been known to mislead the prediction of the variable under study at values beyond a saturation point of the canopy reflectance (Puhr and Donoghue, 2000; Zheng et al., 2004; Boyd and Danson, 2005).

The k-nearest-neighbour (kNN) estimation method has been established as a non-parametric alternative to the use of regression approaches for modelling forest biomass (Keller et al., 1985; Tomppo, 1991; Tokola et al., 1996; Fazakas et al., 1999). Artificial neural networks (ANN) are general purpose computing tools that can solve complex non-linear problems (Fischer, 1998). ANN has already been employed in remote sensing of forest resources and yield stronger correlations with forest biomass in both tropical and non-tropical environments than statistical regression methods (Muukkonen and Heiskanen, 2005; Foody et al., 2001).

* Correspondence address: Institute of Geography, Georg-August University Göttingen, Goldschmidtstr. 5, 37077 Göttingen, Germany. Tel.: +49 551 39 79 79; fax: +49 551 39 80 20.
E-mail address: ppropas@uni-goettingen.de

0303-2434/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jag.2011.12.013

Both the kNN and ANN are based on the assumption that the pixel values depend only on the forest conditions, and not on the geographic location. If a spatial dependence exists, field plots from outside the estimation area cannot be used without taking into account the bias (Tokola et al., 1996). This is the reason why both techniques as well as conventional regression approaches are considered to be non-spatial methods. However, most of geographical phenomena are characterized by spatial dependence, meaning that points situated in geographical space near to each other are likely to show more similar characteristics than points situated far away from each other (Griffith, 2003). Moreover, recent studies have shown that the relationships between spatially distributed environmental variables can vary across the geographical space (Brunsdon et al., 1999; Fotheringham et al., 2002). With respect to the remote sensing applications in forestry, spatial drift has been found in the relationships between satellite data and forest structure characteristics such as net primary production (Wang et al., 2005) and leaf area index (Propastin, 2009). Other related non-remote sensing studies have reported spatial instability of the relationship between tree diameter and tree height (Zhang et al., 2004), aboveground biomass and elevation (Propastin, 2011), basal area and elevation (Kupfer and Farris, 2007). These studies have shown that accounting for spatial drift in a studied relationship significantly improves explanatory power of the model.

In the field of spatial modelling, a number of statistical techniques have been developed to model spatial variations in relationships between variables over space. One of the most powerful tools in exploring spatial heterogeneity of relationships is geographically weighted regression (GWR) (Brunsdon et al., 1996; Fotheringham et al., 2002). The GWR method estimates parameters for all sample points in a data set, taking into account non-stationarity of relationships. GWR allows greater insights into the nature and accuracy of the data under study providing a detailed understanding of the relationships and their spatial variation (Foody, 2003, 2004). The ability to not only model spatial variation but also to map the regression parameters and goodness-of-fit statistics makes GWR a very attractive tool for remotely sensed biomass modelling. Since a GWR-derived map represents each coefficient rather than a single number this method produces a great amount of output information that may be used for better understanding the response of forest biomass to spectral reflectance of satellite bands. However, in the field of remote sensing applications in forestry, GWR-based modelling is a relatively unexplored area.

The present study uses fine resolution satellite imagery (Landsat ETM+) to estimate the aboveground biomass of tropical rainforest at the regional scale. In doing so, this study is not unique: examples of this type of research abound in the literature (e.g., Lu et al., 2004; Zheng et al., 2004; Muukkonen and Heiskanen, 2005; Mäkelä and Perkkarinen, 2004; Puhr and Donoghue, 2000). Instead using the well-known common approaches for the remote sensing based estimation of biomass, the present study strives to make a contribution to current literature through introducing and testing a novel modelling technique. The introduced approach (1) emphasizes the possible spatial variation in the response of forest biomass to the remote sensing-measured reflectance by developing a local model using geographically weighted regression; and (2) extends the traditional GWR framework with altitudinal information adapting GWR for application in a mountainous forest area. The presented geographically and altitudinal weighted regression (GAWR) addresses the problem of spatial non-stationarity in relationships involving both horizontal and altitudinal variations. The study bases on an extensive field survey carried specially out in a tropical rainforest region with diverse terrain conditions where both spatial and altidudinal factors determine distribution of aboveground biomass.

This article is structured as follows. After introduction of the study area and data used in the study in Sections 2 and 3, Section 4 presents a basic framework for GWR and then extend it to include altitudinal data. Section 4 also offers key technical implementation details of the GAWR model, including optimal selection of model parameters, and model comparison criterion. In Section 5, the results of GWR and GAWR application to a case study of biomass mapping in a tropical rainforest in Sulawesi (Indonesia) are reported and compared with respect to models accuracy. Finally, in Section 6, we draw conclusions.

2. Study area

The analysis area is located in Central Sulawesi, Indonesia (Latitude 0°55′–01°54′ South, Longitude 119°40′–120°29′ East) and comprises the region of the Lore-Lindu National Park together with the bordering areas (Fig. 1a). The area has a complicated relief with elevations from zero in the north to about 2500 m above sea level in the middle part and is cut by four river valleys: the Palolo to the north, Napu to the east, Bada to the south and Kulawi to the west. The highest peaks are Mt. Nokilalaki (2355 m) and Mt. Rorekatimbu (2610 m).

Large parts of the study area, especially major parts of the mountain ranges, are covered by undisturbed tropical forest, whose natural vegetation is generally classified into two major vegetation types based on altitudinal distribution with lowland rainforest below 1000 m and mountain rainforest above 1000 m. These forests are dominated by wood species such as *Aglaia argentea*, *Pimelodendron amboinicum*, *Bischofia javanica*, *Cananga odorata*, *Meliosma sumatrana*. The natural forest has been subject to remarkable forest conversion activities along the park boundaries. Near the national park boundaries, many previously forested areas have been transformed into perennial agro-forestry areas with cocoa and/or coffee cultivation. Some of these areas have been abandoned after short-term cultivation and reverted to secondary forest. The secondary forests are dominated by *Acalypha caturus*, *Grewia glabra*, *Homalanthus populneus*, *Macaranga hispida*, *Mallotus mollissimus*, *Pipturus argenteus*. Most areas of the river valleys are completely deforested and used for production of paddy rice.

3. Data

3.1. Ground data

Extensive forest inventory was carried out during a number of field campaigns in 2003–2008, whereas the most part of the measurements was done in 2006–2008. Measurements of several forest structure parameters (diameter at breast height, tree height, and crown diameter) were done in 116 sampling plots located across the Lore-Lindu National Park. A number of sampling plots were also located in forested areas outside of the park borders (Fig. 1b). The location of the plots in the study area was established using a stratified random sampling scheme. The natural vegetation stratification (lowland rainforest, pre-montane rainforest, and montane rainforest, according to Whitten et al., 2002) in the study area served as a base for establishing the sampling plots location. A satellite-based land cover classification (Erasmi et al., 2007) and a digital terrain model were used to determine areas covered by the three rainforest strata.

The employed sampling strategy was organized by plot and subplots (Fig. 1c). All the 85 plots had a size of 40 m × 60 m (~0.25 ha). Geographic coordinates of the corners as well as the elevation above the sea level of each plot were measured by GPS (±5 to 8 m horizontal accuracy). For most of the plots, the accuracy of the GPS elevation measurements was comparable with the horizontal accuracy (±6