Information Fusion 16 (2014) 59–67

Contents lists available at SciVerse ScienceDirect

Information Fusion

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

Data fusion for high spatial resolution LAI estimation

Carmen Hernández [a,b], Leónia Nunes [a,c], Domingos Lopes [a], Manuel Graña [b,*]

[a] Centro de Investigação e de Tecnologias Agro-Ambientais e Tecnológicas (CITAB), Departamento das Ciências Florestais e Arquitectura Paisagista (CIFAP), Universidade de Trás-os-Montes e Alto Douro (UTAD), Quinta de Prados, Apartado 1013, 5001-801 Vila Real, Portugal
[b] Computational Intelligence Group, University of the Basque Country, UPV/EHU, Spain
[c] Agrarian School of Viseu (ESAV)/Polytechnic Institute of Viseu (ISPV), Estrada de Nelas, Quinta da Alagoa, Ranhados, 3500-606 Viseu, Portugal

ARTICLE INFO

Article history:
Received 1 February 2011
Received in revised form 8 April 2012
Accepted 10 April 2012
Available online 26 April 2012

Keywords:
Kriging
Leaf Area Index (LAI)
Remote sensing
Vegetation indices
Hybrid validation

ABSTRACT

Leaf Area Index (LAI) is a critical variable for forest management. It is difficult to obtain accurate LAI estimations of high spatial resolution over large areas. Local estimations can be obtained from *in situ* field measurements. Extrapolation of local measurements is prone to error. Remote sensing LAI estimation products, such as the one provided by MODIS are of very low resolution and subject to criticism in recent validation works. Forest management requires increasingly high resolution estimations of LAI. We propose a data fusion process for high spatial resolution estimation of the LAI over a large area, combining several heterogeneous information sources: field sampled data, elevation data and remote sensing data. The process makes use of spatial interpolation techniques. We follow a hybrid validation approach that combines the conventional prediction error measures with a spatial validation based on image segmentation. We obtain encouraging results of this information fusion process on data from a forest area in the north of Portugal.

1. Introduction

Forest management is guided by the accurate estimation of the current biomass and the prediction of its evolution. There are several dynamical models of the biomass, based on the modeling of the interactions between the soil, the atmosphere and the vegetation [1,2]. The Leaf Area Index (LAI) is a critical input variable for these models. The LAI is the ratio of total upper leaf surface of vegetation divided by the surface area of the land on which the vegetation grows. LAI is a dimensionless value, typically ranging from 0 for bare mineral ground in deserts up to 6 for a dense forest. The accurate LAI estimation over extended forested areas allows the spatial estimation of the biomass evolution, and, therefore, the careful management of the forest. It can be critical from several economical points of view, including the negotiation of CO_2 quotas and the optimal balance between ecological preservation and exploitation of the forest.

In situ field measurements are the most trustworthy sources of information for LAI estimation. Unfortunately, they are very expensive and local. Some remote sensing products, such as the ones provided by Terra Moderate Resolution Imaging Spectroradiometer (MODIS) [3,4], give very low spatial resolution estimations of the LAI. The resolution of MODIS data is very coarse, thus the pixel size is very large allowing for a great heterogeneity of vegetation cove in each pixel. Besides there are growing concerns about its validity over specific terrains [5]. Previous works [6] on the estimation of the Net Primary Production (NPP) have shown the simplification of reality incurred by MODIS and its bias underestimating extreme values of NPP.

High spatial resolution remote sensing, i.e. Landsat, do not provide LAI estimation, but it is possible to compute some variables correlated with LAI. In this paper we propose an innovative information fusion process based on spatial interpolation methods to provide high resolution accurate estimations of LAI. The information sources are the *in situ* field measurements, the remote sensing images and the altitude data obtained from digital elevation maps. The process fits a random field to the data provided by the *in situ* measurements using the remote sensing and altitude data to improve the spatial interpolation of the LAI values, specially in regions outside the *in situ* sampled areas. We propose also a hybrid validation approach combining conventional prediction error measures and the spatial validation based on the mutual information between the estimated LAI values and the results of remote sensing image segmentation evaluated by the authors LN and DL which are experts on the specific characteristics of this study area where they have done field work for many years. The experts were confronted with a Likert scale of LAI range values into five categories (lowest, low, medium, high, highest) to assign them to each image region obtained by the segmentation software, so that they are not required to provide detailed quantifications.

Previous works on information fusion involving remote sensing data were concerned with the registration of images from different

* Corresponding author.
E-mail address: manuel.grana@ehu.es (M. Graña).

1566-2535/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.inffus.2012.04.001

sensors [7–9] or the combination of classifiers for improved thematic map generation [10]. There are few instances [11] combining remote imaging data and other kinds of information. Specifically, *in situ* measurements have been used mostly for validation of classification or regression results. Recent validation works [5] show that the MODIS LAI product underestimated LAI in some croplands. Therefore, there is a need for innovative accurate LAI estimation procedures such as the one proposed in this paper. The preliminary results provided in this paper encourage further exhaustive computational experiments with greater data support which would allow detailed statistical analysis and validation.

The contents of the paper is the following: Section 2 contains some background information on LAI estimation. Section 3 gives the formal definitions of the spatial information methods used to perform information fusion. Section 4 describes the data used in the experiment and some validation methodological issues. Section 6 reports our experimental results. Finally, Section 7 gives our conclusions and directions for future work.

2. Background on LAI estimation

Leaf Area Index (LAI) is a fundamental characteristic of terrestrial ecosystems because it is directly connected with evapotranspiration, photosynthetic rate, forest production and site water balance [12,13]. The LAI or the leaf area per unit ground area is especially important in many global terrestrial processes for computing the exchanges of energy, water and other gases [14]. LAI constitutes a key variable for ecological studies because is strongly related to physiological processes. Borak and Jasinski [15] consider the LAI, an important quantity in many global climate and ecological models.

Estimation of LAI from remote sensing data has been the subject of active research. There are algorithms to solve the inverse problem of finding LAI from MODIS data [16–18] based on the bidirectional reflection model and patterns of reflectance observed in the nature which have become standard products. Using Synthetic Aperture Radar (SAR) the regions with specific LAI values can be segmented from the image [19]. For Landsat data, such as the one used in our work, Bayesian networks have been trained to predict the LAI value [20]. Recent validations of the MODIS LAI product using *in situ* measurements and Landsat derived vegetation indices [5] demonstrated that MODIS LAI underestimates LAI on some croplands. Therefore, innovative and more accurate LAI estimation methods are needed.

Althought direct LAI estimates can not be obtained from Landat image, it is possible to compute [21] physiologically-based vegetation indices strongly correlated with LAI, such as the Normalized Difference Vegetation Index (NDVI) defined as follows:

$$NDVI = \frac{NIR - R}{NIR + R} \tag{1}$$

where R and NIR stand for the spectral reflectance measurements acquired in the red and near-infrared regions, respectively. Vegetation indices such as NDVI are correlated with reductions in photosynthetically active radiation (PAR). The dependence of NDVI on LAI is based on the correspondence between the amount of leaves and the absorption of PAR (APAR) in the satellite-observed spectrum of solar reflection [22]. The relationship between APAR and NDVI is linear [23] and the relationship between the fraction of APAR (fAPAR) and LAI is exponential:

$$fAPAR = 1 - exp(-kLAI), \tag{2}$$

where k is a coefficient related to the spatial distribution and structure of foliage and leaves. Fensholt [24] reported that numerous studies have analysed this relation and there is general agreement that a stronger relation exists between fAPAR and NDVI than between LAI and NDVI. Based on satellite data, the relationship between NDVI and fAPAR has been found to be linear or approximately linear for green vegetation. The equations relating LAI and NDVI are very local and difficult to estimate. Their parameters depend on vegetation types, seasonal and annual variations. We do not attempt to fit them from the scarce available data. Instead, we use the NVDI as ancilliary information that can be useful to perform accurate extrapolations of *in situ* LAI measurements performed through spatial interpolation methods.

3. Spatial interpolation models

Geostatistical methods use statistical properties of the given observations of an spatial process (i.e. spatial autocorrelation). The theory of regionalized variables [25] and geostatistical methods based on it are an effective tool for studying spatial distributions of spatial variables, and to perform interpolation of unevenly distributed samples over a regular sampling grid. Geostatistical studies have been used in many applications, such as hydrological data modeling [26,27], mining [28,29], studies of air quality [30,31], biology [32,33], and economics [34].

Kriging is a statistical interpolation method. Given a set of spatial samples, it fits a random field to them to produce predictions of the values of the latent spatial process in positions where no observation is given. Therefore, the goal of kriging is to estimate the value of a random function, $Z(s)$ in non-sampled positions s of a region D, given a set of observation $\{Z(s_1), \ldots, Z(s_n)\}$ obtained in positions, $s_1, \ldots, s_n$. The kriging predictor, $\widehat{Z}(s_0)$, is a linear combination of the observed values:

$$\widehat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i \cdot Z(s_i), \tag{3}$$

where the linear combination parameters λ_i are solutions of a system of equations obtained assuming that the data come from a sample surface of a random process $\mathcal{Z}(s)$ and that we want to minimize the error of prediction:

$$\varepsilon(s) = \mathcal{Z}(s) - \sum_{i=1}^{n} \lambda_i \cdot Z(s_i). \tag{4}$$

Kriging computes the best linear unbiased estimator $\widehat{Z}(s_0)$ based on a stochastic model of the data spatial dependence, which is quantified either by the variogram or by the mean and covariance function of the random field. The Simple Kriging (SK) assumes that the process mean and variance are constant and known, while the Ordinary Kriging (OK) [35] assumes that the process mean and covariance are unknown and must be estimated. The estimation of the kriging parameters is given by the following equation:

$$\lambda = \Gamma^{-1} \gamma(s), \tag{5}$$

where $\lambda = [\lambda_1, \ldots, \lambda_n]$ corresponds to the vector of kriging parameters, $\Gamma = [\gamma(s_i, s_j);\ i, j = 1, \ldots, n]$ is the variogram matrix, $\gamma(s_i, s_j)$ is the value of the variogram between two sampling points, and $\gamma(s)$ denotes the vector of variogram values between the sampling points and the point of actual estimation.

3.1. Universal krigging

The Universal Kriging (UK) assumes the existence of a trend in the data, so that the local mean of the process depends on the actual position $E[\mathcal{Z}(s)] = m(s)$. Therefore, the residual process obtained after removing the trend

$$R(s) = Z(s) - m(s) \tag{6}$$

is assumed to be a (zero mean) stationary process which can be interpolated using simple kriging. If there is some deterministic