Available online at www.sciencedirect.com
ScienceDirect
Remote Sensing of Environment 112 (2008) 613–622
Remote Sensing of Environment
www.elsevier.com/locate/rse

A Bayesian network algorithm for retrieving the characterization of land surface vegetation

Yonghua Qu [a,b,*], Jindi Wang [a,b], Huawei Wan [d], Xiaowen Li [a,b], Guoqing Zhou [c]

[a] School of Geography, Beijing Normal University, Beijing, China
[b] State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and the Institute of Remote Sensing Applications of Chinese Academy of Sciences, Beijing, China
[c] Department of Civil Engineering and Technology, Old Dominion University, Norfolk, VA 23529, United States
[d] Environmental Satellite Center Preparing Office, State Environmental Protection Administration, Beijing, China

Received 15 March 2006; received in revised form 23 March 2007; accepted 31 March 2007

Abstract

A hybrid inversion technique based on Bayesian network is proposed for estimating the biochemical and biophysical parameters of land surface vegetation from remotely sensed data. A Bayesian network is a unified knowledge-inferring process that can incorporate information derived from multiple sources including remote sensing and information derived from a priori knowledge. Using this inversion approach, content of chlorophyll *a* and chlorophyll *b* (Cab) and leaf area index (LAI) of winter wheat were estimated from data derived from simulations as well as field measurements. Estimations from the simulated data proved accurate, with root mean square errors (RMSEs) of 0.54 m^2/m^2 in LAI and 4.5 $\mu g/cm^2$ in Cab. In validating the estimates against field measurements, it was found that prior knowledge of target parameters improved the accuracy of estimates, in terms of RMSEs from 0.73 to 0.22 m^2/m^2 in LAI and 9.6 to 4.0 $\mu g/cm^2$ in Cab. Bayesian inference in this hybrid inversion scheme produces a posterior probability distribution, which can reveal such properties of the inferred results as updated information contained in the inversion result. Using entropy, the revision of posterior information about the parameters of interest was calculated. Including more data may allow more information to be retrieved about parameters in general. Exceptions were also observed where data from some viewing angles slightly reduced the information on the parameters of interest. It was also found that data from these viewing angles were less sensitive to the parameters. The method proposed here was also validated using LandSat ETM+ imagery provided by the BigFoot project. When used for mapping LAI with ETM+ imagery, the proposed method with an RMSE of 0.70 and a correlation of 0.67 produced a slightly better result than that from empirical regression.

Keywords: Bayesian network; Leaf area index; Chlorophyll concentration; Hybrid inversion; Information entropy

1. Introduction

Estimating such biophysical and biochemical parameters of land surface vegetation as leaf area index (LAI) and leaf chlorophyll content (Cab) is an important application of remote sensing (Koetz et al., 2005; Myneni et al., 1995; Verstraete et al., 1996). LAI, defined as half the total leaf surface area per unit area of horizontal surface, is an important structural variable of vegetation and a key variable in understanding several ecophysiological processes within the vegetation canopy (Gong et al., 2003; Tian et al., 2003). Cab, the sum of the contents of chlorophyll *a* and chlorophyll *b* per unit leaf area, is intimately associated with physiological functions of leaves (Gitelson & Merzlyak, 1997; Sims & Gamon, 2002). Both LAI and Cab are affected when vegetation is exposed to natural and anthropogenic stresses, and non-destructive determination of these parameters from a distance is a good method of studying leaf function, and plant physiological state and stress (Koetz et al., 2005).

LAI and Cab can be estimated either by empirical methods or by inverting a radiative transfer model (Baret & Guyot, 1991;

* Corresponding author. School of Geography, Beijing Normal University, Beijing, China. Tel.: +86 10 58809966; fax: +86 10 58805274.
E-mail address: qyh@bnu.edu.cn (Y. Qu).

0034-4257/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2007.03.031

Verstraete et al., 1996). The empirical method involves constructing empirical formulae that link spectral features (e.g. vegetation index) and parameters of the earth's surface using experimental data (Gong et al., 1992; Walthall et al., 2004). The empirical method is simple and efficient in estimating the parameters but has a few inherent disadvantages. Since the relationship is derived using data at a specific time and place, the empirical formulae are limited temporally and spatially. An alternative is to invert a physical model, a method that has a clear physical basis and the retrieved results can be explained using physical models (Myneni et al., 1995).

However, physical models are often complicated and non-linear; as a result, the inversion process is often ill-posed (Combal et al., 2003). Many mathematical techniques have been developed to handle such problems, such as the regularization method (Doicu et al., 2003, 2004; Fymat, 1979) and artificial neural network techniques (Fang & Liang, 2003). The problem of inversion can also be solved by the knowledge reasoning method rather than by mathematical optimization (Kimes et al., 1991). In applying knowledge reasoning, the ill-posed problem can be regarded as a result of insufficient information during the process of knowledge reasoning. Therefore, additional knowledge is used in the inversion process (Combal et al., 2003; Li et al., 2001).

Recent research has shown that combining the empirical method and physical model inversion into a new hybrid inversion scheme is a promising approach to estimating surface parameters (Fang & Liang, 2003; Liang, 2004). The scheme uses simulated data sets to fit an empirical formula and the fitted equation is then used for estimating land surface parameters. Although non-linear fitting methods such as artificial neural network (ANN) and projection pursuit regression (PPR) have been used in earlier hybrid inversion models, and have shown how inversion efficiency and accuracy can be improved, there is still scope for improvement by incorporating more information in the inversion process. In the earlier hybrid inversion schemes, parameters of the physical model alone were included; other parameters that may influence the parameters of interest were not incorporated into the process of estimation. For example, although it is known that LAI is affected by crop growth stages, current hybrid inversion methods lack a way to introduce such temporal information into the estimations.

This paper proposes an alternative hybrid inversion method, which uses a Bayesian network to map the simulated reflectance to its corresponding biophysical parameters. As a hierarchical probability model, a Bayesian network can be used not only as a non-parametric regression model but also to deduce information from multi-layer parameters (Marcot et al., 2001). Kalacska et al. (2005) used it for estimating LAI of a tropical dry forest from ETM+ data and obtained better results than those obtained from spectral vegetation indices or ANN. However, for the initial network estimates the authors used data from ground surveys combined with known forest structure, LAI, and satellite reflectance. The method, therefore, was not free of the disadvantages inherent in the empirical method. In our proposed Bayesian network approach, the initial network estimates use the data obtained by simulating a physical model. Therefore, the approach can be used for estimating the biophysical and biochemical parameters of a standing crop over a wider temporal and spatial range than is possible with a limited amount of ground measurements. In our approach, we also focus on incorporating ancillary information extracted from a spectral library, namely the Spectral Library on Typical Land Surface Objects in China (SLTLSOC) (Qu et al., 2003), to support the inversion, which differs from other non-parametric regression methods such as ANN and PPR.

Our study sought to develop a new hybrid inversion scheme supported by the SLTLSOC and was tested against data sets obtained from both simulations and field measurements. The second purpose was to study changes in the values of the parameters of interest after sequentially adding multi-angle observation data to the inversion process.

2. Methods

2.1. Radiative transfer model

A coupled radiative transfer model, PROSPECT+SAIL (PROSAIL), was used to simulate the reflectance of vegetation canopies. PROSPECT can simulate the reflectance and transmittance of leaves using their biochemical and biophysical parameters (Baret & Fourty, 1997; Jacquemoud & Baret, 1990). These parameters include Cab, leaf water content (Cw), dry matter content (Cm), and the leaf mesophyll structural parameter (N). SAIL (Scattering by Arbitrarily Inclined Leaves) is a physics-based radiative transfer model used for simulating the hemispheric reflectance spectra of canopies under different viewing directions (Verhoef, 1984). The version of SAIL used in this study was developed by Kuusk, which included the hot-spot effect in the original SAIL model (Kuusk, 1991). The SAIL model needs seven input parameters: LAI, average leaf angle (ALA), ratio of leaf length to canopy height (SL), leaf hemispheric reflectance (LR), leaf transmittance (LT), soil reflectance, and atmospheric visibility (VIS). LR and LT can be simulated by PROSPECT. The coupled PROSAIL model computes multi-spectral reflectance under different incident and observation directions. We used the simulated BRDF (Bidirectional Reflectance Distribution Function) data from wavelengths of 400–900 nm at intervals of 5 nm, with the illumination-viewing angle of 55°. Because Landsat ETM+ data and field measurements were to be used, the response functions of ETM+ on green (525–605 nm), red (630–690 nm), and near-infrared (775–900 nm) bands were used to resample the simulated reflectance from simulated narrow bands into broad ETM+ bands 2, 3, and 4 respectively.

To reduce the number of model parameters to be retrieved, some parameters can be fixed when simulating the canopy reflectance. PROSPECT computes leaf reflectance and transmittance based on the specific absorption coefficient (SAC), the main factor influencing leaf reflectance, of each component in every band (Jacquemoud & Baret, 1990). The effect of Cw and Cm is mainly in wavelengths longer than 1300 nm and of Cab, in approximately 400–800 nm. Since reflectance in wavelengths less than 900 nm is used to estimate chlorophyll content,