REGRESSION TECHNIQUES AND NEURAL NETWORKS FOR ESTIMATING LAI FROM REMOTE SENSING DATA – AN APPLICATION CASE USING FOREST-BGC

Leónia NUNES[a], Carmen HERNANDEZ[b], Miguel VEGANZONES[b], Margarida L. R. LIBERATO[cd], Manuel GRAÑA[b] and Domingos LOPES[e]

[a] Agrarian School of Viseu (ESAV) / Polytechnic Institute of Viseu (ISPV), Estrada de Nelas, Quinta da Alagoa - Ranhados, 3550-606 Viseu, Portugal. Phone: 00 351 232 446 600, Fax: 00 351 232 426 536, e-mail: lnunes@esav.ipv.pt.

[b] Grupo de Inteligencia Computacional, Universidad del País Vasco (UPV/EHU), Pº Manuel Lardizabal 1, 20018 San Sebastián, Spain. Phone: 00 34 943 015 245, Fax: 00 34 943 015 590, e-mail: mamen.hernandez@ehu.es, miguelangel.veganzones@ehu.es, cpgrrom@si.ehu.es.

[c] University of Trás-os-Montes e Alto Douro, School of Sciences and Technology, Physics Dept, Apartado 1013, 5001-801 Vila Real, Portugal. Phone: 00 351 259 350 000, Fax: 00 351 259 350 480, e-mail: mlr@utad.pt.

[d] CGUL, IDL, University of Lisbon, 1749-016 Lisbon, Portugal.

[e] University of Trás-os-Montes and Alto Douro (UTAD), Forestry and Landscape Architecture Department, Apartado 1013, 5000-911 Vila Real, Portugal. Phone: 00 351 259 350 000, Fax: 00 351 259 350 480, e-mail: dlopes@utad.pt.

ABSTRACT

In order to build a model that could routinely estimate carbon sequestration for Portuguese forest ecosystems, we are designing a system based on FOREST-BGC, a well known ecophysiological model. The most critical model's input variable is the Leaf Area Index (LAI). In this work we propose and validate LAI estimation from LANDSAT TM image data.

Field data showed that the Portuguese forest ecosystems are heterogeneous and difficult for modelling, although previous results proved that it is possible to estimate LAI from vegetation indices computed from remote sensing data. A LANDSAT TM image from 2006 was geometric and radiometric corrected and maps of vegetation indices images were created.

Regression models over the vegetation indices and a Multilayer Perceptron (MLP) taking as input the six LANDSAT TM bands were compared for LAI estimation. The resulting final NPP maps obtained applying FOREST-BGC to the estimated LAI values can be a powerful tool for decision makers, allowing to evaluate the evolution of forested areas and the quantification of carbon stocks and, more important, the carbon dynamics.

Keywords: Net Primary Production; FOREST-BGC; Leaf Area Index; Vegetation Index

METHODOLOGY

Figure 1 describes a global view of the followed methodology. At this stage, the work was mainly focused on the remote sensing investigation line, but it is important to underline that fieldwork measurements were made, in order to acquire data for the model validation.

A sample of 30 plots of *Eucalyptus globulus* was marked in the study area. Sampling plots were 500m^2 circles and included dendrometric measurements, physiological and soil.

A GPS file was collected in the centre of each sampling plot for its accurate localisation and was used afterwards to locate each sampling plots on maps, aerial and satellite images. Figure 2 presents false colour aerial photo with the sampling plots in the study area.

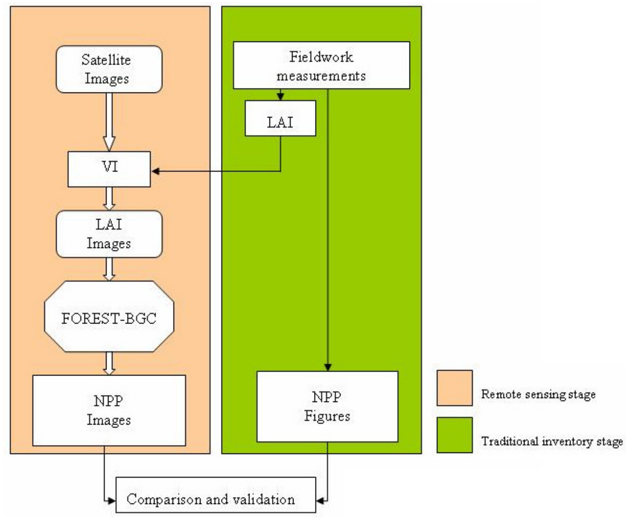

Figure 1. Methodology followed in this research.

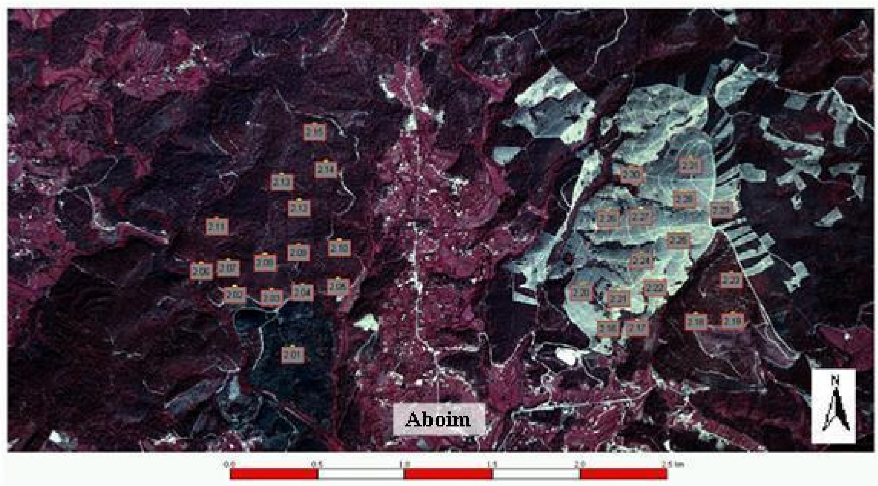

Figure 2. Aerial photo showing the sampling plots for the *Eucalyptus globulus.*

In terms of remote sensing, the available LANDSAT TM image from September 2006 were radiometrically and geometrically corrected, following the same methodology as described by Pons and Sole-Sugrañes, (1994) and Pons (2002). This correction was made using the software MIRAMON, developed by a team of researchers in remote sensing of the Autonomous University of Barcelona (UAB), in Spain (Pons and Sole-Sugrañes, 1994).

The images were classified in order to identify the *Eucalyptus* area. A Boolean final image was defined where 1 represented *Eucalyptus* stands and 0 represented non-*Eucalyptus* areas.

Later, some vegetation indices with different degrees of complexity were tested. The vegetation indices analysed, described in Table 1, were calculated overlapping the GPS coordinates of each sampling plot for LANDSAT TM image from 2006.

The following stages were related with the LAI estimation based on the vegetation indices. The scale used to estimate the LAI was LANDSAT pixel. In GIS software the average of pixels surrounding the centre of the sample was determined automatically.

Different approaches were followed ranging from a linear regression adjustment till a neural network using as starting point the reflectance units from each band. In this last case, the Multi-layer Perceptron (MLP) procedure produces a predictive model for one or more dependent (target) variables based on the values of the predictor variables. The neural network was adjusted based on the matrix of data to fit the models, using the SPSS software.

Table 1. Vegetation indices tested.

Designation	Mathematical expression	Designation	Mathematical expression
NDI(MIR1)	$\frac{(NIR - MIR1)}{(NIR + MIR1)}$	TVI1	$\sqrt{\frac{NIR}{R}}$
NDI (MIR2)	$\frac{(NIR - MIR2)}{(NIR + MIR2)}$	TVI2	$\frac{(NIR + R)}{(NIR - R)}$
NDTI	$\frac{(MIR1 - MIR2)}{(MIR1 + MIR2)}$	TVI3	$\frac{G}{R}$
NDVI	$\frac{(NIR - R)}{(NIR + R)}$	TVI4	$\sqrt{\frac{G}{R}}$
MVI1	$\frac{MIR1}{MIR2}$	TVI7	$\frac{(G - R)}{(G + R)}$
MVI2	$\frac{NIR}{MIR2}$	TVI8	$\frac{(G + R)}{(G - R)}$
MVI3	$\frac{NIR}{(MIR1 + MIR2)}$	TVI9	$\sqrt{\frac{(G - R)}{(G + R)} + 0,5}$
RVI	$\frac{NIR}{R}$		

R is reflectance in the red wavelength; **G** is reflectance in the green wavelength; **NIR** is reflectance in the near infrared wavelength; **MIR1** is reflectance in the mid infrared band 1; **MIR 2** is reflectance in the mid infrared band 2.

RESULTS AND DISCUSSION

Lai estimation from remotely sensed data -empirical equation linking LAI to VI

In a first stage and in order to identify the best independent variable for the LAI estimation, the correlation matrix between LAI and the vegetation indices was analysed. NDI(MIR2) is the better index (LAI X NDI(MIR2) = 0.718), followed by NDVI (LAI x NDVI = 0.697).

As reported in the literature, the merging of remote sensing data (in this case to estimate LAI) with ecophysiological models (in this case the FOREST-BGC, which uses LAI as the main input variable) will allow the development of a practical methodology for monitoring NPP and thus the sequestered carbon over large areas.

A linear equation was adjusted (equation 1), for future comparison with more sophisticated approaches ($R^2 = 0.539$; $R^{2adj.} = 0.523$; RMSE = 0.706). The model is statistically significant (sig < 0.05) which means the model fits to the data and the residual analysis showed required data normality.

$$LAI = -1.139 + 5.876\, NDI(MIR2), \qquad (1)$$

LAI estimation from remote sense data –neural network method

After this, the feasibility to estimate LAI using a neural network was also tested. The results were obtained with a 10 units neural network (Figure 3) whose input are the reflectance values from all the available bands and has a hyperbolic activation function, where reflectance of Band 7 (DN_MIR2) was the strongest unit. The neural network has 1 hidden layer composed of 3 units and uses as input the reflectance values from all the LANDSAT TM bands. The results obtained from this study are in agreement with the ones obtained from Kavzoglu (2009). Lucas (1995) tried to estimate LAI by adopting vegetation indices as predicting variables. He concluded that it is possible to evaluate a forest stand parameter from these images and that NDVI was the best vegetation index (RMSE = 2.21). Using LANDSAT

TM imagery Fassnacht *et al.* (1997) obtained correlation coefficient values of 0.96 and a RMSE of 0.315 when using red, near and mid infrared wavelengths simultaneously, in predictive models developed for conifers. The results obtained on the regression phase of this research are in good agreement with these published results. Despite the regression approach is more practical and easily applicable, the best results of this research were achieved with the neural networks. Although it is more complex than the linear regression, anyone with a basic background of programming can replicate the results.

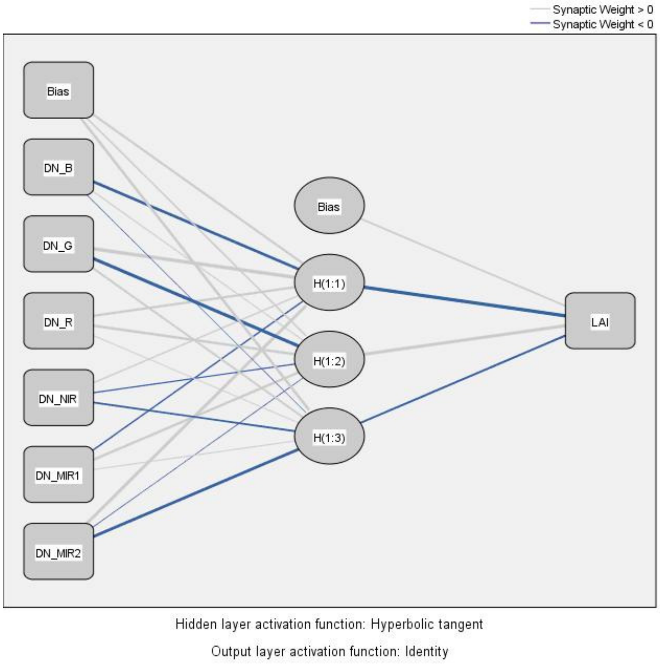

Figure 3. Neural network adjusted using as initial units the reflectance from the LANDSAT TM bands.

Comparison of estimated LAI values

Later LAI estimated by the regression approach was compared with the ones from the neural network (Figure 4). From the comparison of only graphical analysis can be concluded that neural networks describes better LAI mainly for LAI values under 2.5 $m^2.m^{-2}$.

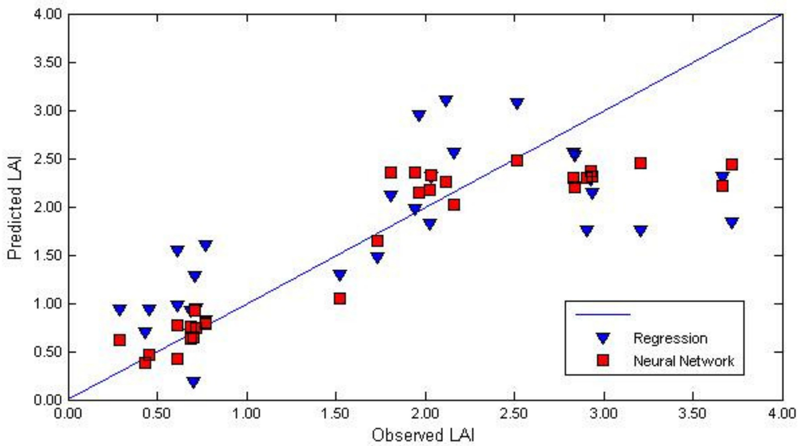

Figure 4. Comparison between LAI from the regression and from the neural network.

The worst results are occurring only for LAI values upper than 2.5$m^2.m^{-2}$ (Table 2) which represents only 9 sampling plots. This means further studies should be implemented for higher LAI stands. In some cases, the spectral vegetation indices saturate at LAI values of 2-3 due to multiple scattering effects (Chen, 2002; Heiskanen, 2006). That is, the canopies with a high LAI value have similar NDVI values.

Table 2. Quantification of the average deviation on LAI estimation from the regression and the neural network approaches.

		RMSE	BIAS
Global Data	**Regression**	0,725932928	-0,02828858
	Neural Network	0,489663862	-0,16573262
LAI < 2.5	**Regression**	0,5135593	0,30519345
	Neural Network	0,2306704	0,07145425
LAI > 2.5	**Regression**	1,068267198	-0,80641332
	Neural Network	0,821633543	-0,71916865

REFERENCES

Borak, J.S. and Jasinski, M.F., 2008. Effective interpolation of incomplete satellite-derived leaf-area index time series for the continental United States. *Agricultural and Forest Meteorology*, 149: 320-332.

Chen, J.M., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S.G., White, H.P., Hall, R.J., Peddle, D.R., King, D.J., Trofymow, J.A., Swift, E., Van Der Sanden J. and Pellikka, P.K.E., 2002, Derivation and validation of Canada-wide coarse resolution leaf area index maps using high-resolution satellite imagery and ground measurements. *Remote sensing of environment*, 80: 165-184.

Fassnacht, K.S., Gower, S.T., MacKenzie, M.D., Nordheim, E.V., Lillesand, T.M., 1997. Estimating the leaf area index of north central Wisconsin forests using the Landsat Thematic Mapper. *Remote Sensing of Environment*, 61: 229-245.

Goetz, S.J and Prince, S.D., 1996. Remote sensing of net primary production in boreal forest stands. *Agricultural and Forest Meteorology*, 78. 149-179.

Goetz, S.J., 1997. Multi-sensor analysis of NDVI, surface temperature and biophysical variables at a mixed grassland site. *International Journal of Remote Sensing*, 18 (1): 71-94.

Govaerts, Y.M., Verstraete, M.M., Pinty, B. and Gobron, N., 1999. Designing optimal spectral indices: a feasibility and proof of concept study. *International Journal of Remote Sensing*, 20 (9): 1853-1873.

Gower, S.T., Kucharik, C.J. and Norman, J.M., 1999. Direct and indirect estimation of leaf area index, fAPAR, and net primary production of terrestrial ecosystems. *Remote Sensing of Environment*, 70: 29-51.

Heiskanen, J., 2006, Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data. *International Journal of Remote Sensing*. 27: 1135-1158.

Kavzoglu, T., 2009, Increasing the accuracy of neural network classification using refined training data. *Environmental Modelling and Software*, 24: 850-858.

Pons, X. and Solé-Sugrañes, L., 1994. A simple radiometric correction model to improve automatic mapping of vegetation from multispectral satellite data. *Remote Sensing of Environment*, 48: 191-204.

Pons, X., 2002, *MiraMon. Geographic information system and remote sensing software* Centre de Recerca Ecològica i Aplicacions Forestals, CREAF. Bellaterra. ISBN: 84-931323-5-7.

Running, S. W., Coughlan, J., 1988. A general model of forest ecosystem processes for regional applications – I. Hydrologic balance, canopy gas exchange and primary production processes. *Ecological Modelling*, 42: 125-154.