Estimating biophysical vegetation parameters in a *dehesa* ecosystem using non-parametric statistical analysis: a multi-scale approach

Vilar, L.[1], Pacheco-Labrador[1], J., Pérez-Cabello[2], F., Martín, M.P[1]

[1] Institute of Economics, Geography and Demography, Centre for Human and Social Sciences. Spanish National Research Council (CSIC), Albasanz 26-28. 28037 Madrid, Spain
[2] GEOFOREST Group, IUCA, Department of Geography and Spatial Management, University of Zaragoza, Pedro Cerbuna 12, E-50009, Zaragoza, Spain

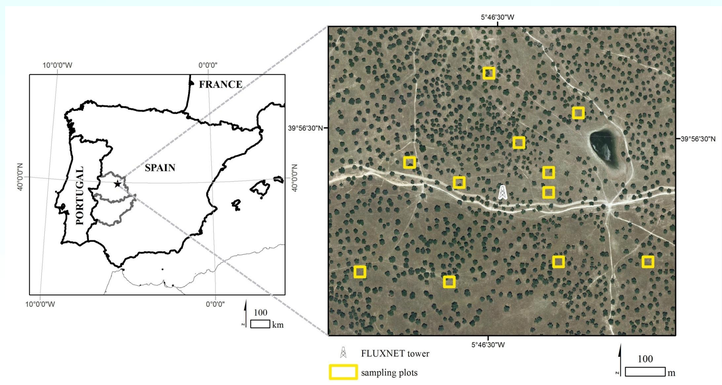
ABSTRACT

Remote sensing (RS) allows monitoring vegetation biophysical properties (VBP) at synoptic temporal and spatial scales. Some of these variables are being systematically monitored at coarse spatial resolution by global remote sensing programs. However, these products, which offer information at pixel sizes ranging between 250 to 3000 m, have demonstrated lack of accuracy and consistency in complex ecosystems such as the savannas. In these tree-grass ecosystems (TGE), structural heterogeneity causes within-pixel spectral mixture and presence of shadows and roughness that can cause considerable errors in the estimation of vegetation status at ecosystem level. Proximal sensing can be used to quantify the impact of these effects on the estimation of biophysical vegetation parameters and enhance our remote sensing and modeling abilities in these complex ecosystems. In this study we used hyperspectral field spectroscopy measurements to calibrate predictive models of **Live Fuel Moisture Content** (**LFMC**), **Canopy Water Content** (**CWC**), **Plant Area Index** (**PAI**) and **Specific Leaf Area** (**SLA**) in the grass layer of a *dehesa*, a Mediterranean tree-grass ecosystem. Models based on Landsat 5 TM spectral vegetation indices (SVIs) assessed the performance of two non-parametric statistical techniques: partial least squares regression (PLRS) and random forest regression (RFR) using field spectroscopy measurements. These models were up-scaled to Landsat 5TM images to assess the impact of ecosystem and atmospheric correction uncertainties on their predictions. Results showed that RFR slightly outperformed PLSR. CWC was the VBP best predicted by both methods (R^2~0.7 and RRMSE~42%). As a result of models up-scaling to Landsat 5 TM imagery, prediction and accuracy of the estimations decreased in about 20% except for LFMC (R^2~0.81 and RRMSE~47.63%).

Keywords: Biophysical variables, field spectroscopy, Landsat, partial least squares regression, random forest regression, tree-grass ecosystem

STUDY AREA

Majadas del Tiétar, Cáceres (Spain)



- Mediterranean TGE *(dehesa)*
- Scattered oak trees (~20% fraction cover) + grazed pasture
- Mediterranean climate type: average annual temperature 16.7ºC with 30 daily average and ~40ºC maximum values in summer months (July-August). Precipitation annual average ~572 mm while in summer is below 70 mm
- Eddy covariance flux towers in site belonging to CARBORED and FLUXNET to track the water and carbon fluxes

DATA

Biophysical variables

- 2009-2015 field campaigns = 54 sample dates
- 22 25x25cm plots; destructive sampling

$$LFMC\ (\%) = \frac{W_{Fresh} - W_{Dry}}{W_{Dry}} \times 100$$

$$CWC\ (g \cdot cm^{-2}) = \frac{W_{Fresh} - W_{Dry}}{A}$$

$$PAI\ (cm^2 \cdot cm^2) = \frac{\frac{W_{Dry}}{W_{Sub.Dry}} \times A_{Sub.Leaf}}{A}$$

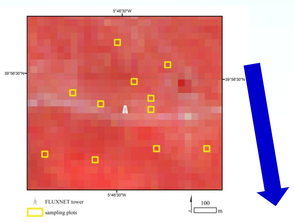
$$SLA\ (cm^2 \cdot g^{-1}) = \frac{A_{Leaf}}{W_{Dry}}$$

Hand-held hyperspectral data

- In 25*25m plots Hemispherical-Conical Reflectance Factors sampled at nadir along two diagonal transects NE-SW and NW-SE
- Hyperspectral data resampled to the Landsat 5 and Landsat 8 bands

Landsat images

- 14 Landsat 5 images
- (202/32) (http://glovis.usgs.gov)
- Pixel value extracted from
- 25x25 m plots



Spectral vegetation indices calculated

NDVI, EVI, NDII, MSI, NDII2, SAVI, GEMI, VARI

METHODS

Statistical analysis

Non-parametric linear: Partial Least Squares Regression (PLSR)

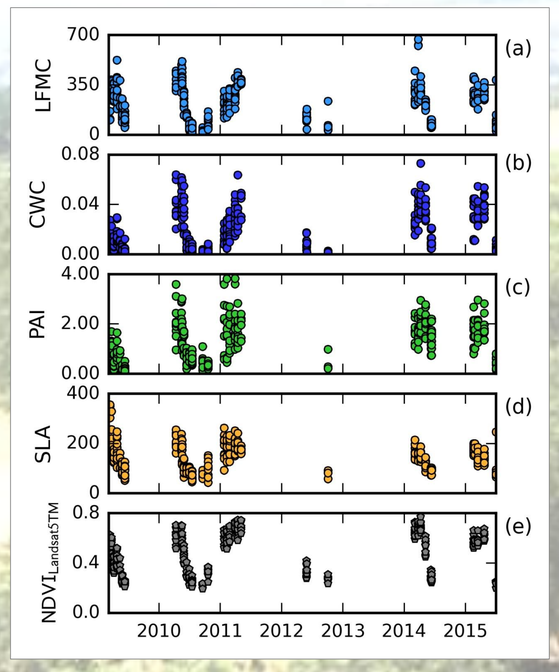
Non-parametric non-linear: Random Forest Regression (RFR)

65% sample to calibrate
35% sample to validate

PLSR and RFR Models from proximal sensing → Landsat 5TM → **Up-scaling** → R^2, RRMSE

RESULTS AND DISCUSSION

Time data of VBP and NDVI



PLSR performance

Biophysical variable	R_{cal}^2	$RRMSE_{cal}$ (%)	R_{val}^2	$RRMSE_{val}$ (%)
LFMC	0.59	35.53	0.63	37.48
CWC	0.71	42.71	0.77	37.68
PAI	0.58	39.62	0.53	45.16
SLA	0.47	30.07	0.58	36.27

RFR performance

Biophysical variable	R_{cal}^2	$RRMSE_{cal}$ (%)	R_{val}^2	$RRMSE_{val}$ (%)
LFMC	0.62	34.07	0.58	39.33
CWC	0.69	44.08	0.68	43.49
PAI	0.65	36.20	0.50	45.51
SLA	0.48	29.57	0.53	25.60

cal: calibration; val: validation

Up-scaling performance

	PLSR		PFR	
Biophysical variable	R_{cal}^2	$RRMSE_{cal}$ (%)	R_{val}^2	$RRMSE_{val}$ (%)
LFMC	0.81	47.63	0.53	43.93
CWC	0.46	112.34	0.49	58.01
PAI	0.25	54.03	0.43	48.31
SLA	0.58	27.33	0.61	25.54

- **Calibration:** RFR R^2_{cal} where slightly larger than PLSR R^2_{cal} except for CWC; $RRMSE_{cal}$ behaved consequently. R^2_{val} were always larger for PLSR, but $RRMSE_{val}$ was ~ 10% larger than RFR for SLA predictions. CWC was the variable best predicted.
- **Models up-scale:** Models performance varied when up-scaled to Landsat 5 TM, prediction errors increased respect to $RRMSE_{val}$, especially CWC predicted by PLSR.

CONCLUSIONS

Biophysical variables of the grass canopy of a TGE were retrieved from SVI derived from hyperspectral field data and field sampling campaign. Two statistical methods were tested, PLSR and RFR. CWC was the variable better predicted and RFR slightly outperformed PLSR. Those models were applied to Landsat 5 TM images in an up-scale exercise. Modest performance of both methods in the estimation of the VBP can be related with the ecosystem characteristics and the field sampling protocol as the areas sampled for the estimation of the VBP and the calculation of VIs do not completely match. This is specially important in grazed ecosystem, as the one under investigation, as this causes a large heterogeneity at the sub-plot scale (25x25m) that can be spectrally characterized using transects but not by destructive sampling of a limited number of sub-plots (25x25cm). Further investigation is required to explore this issue in order to define field protocols better adapted to grazed TGEs.

REFERENCES

Breiman, L., 2001. Random Forests. Machine Learning 45, 5-32.
Hanan, N.P., Hill, M.J., 2010. Savannas in a Changing Earth System. The NASA Terrestrial Ecology Tree-Grass Project. Outline for a Coordinated NASA Field Campaign for Earth Observation and Modeling in Mixed Tree-Grass Ecosystems. In.
Wold, H., 1985. Partial least squares. In: Kotz, S.J., Norman L. (Ed.), Encyclopedia of statistical sciences 6, New York: Wiley, pp. 581–591.

AKNOWLEDGEMENTS

This research has been funded by the FLUXPEC project "Monitoring changes in water and carbon fluxes from remote and proximal sensing in a Mediterranean *dehesa* ecosystem" (CGL2012-34383, Ministry of Economy and Competitiveness).We would like to acknowledge the personal from CCHS-CSIC, INIA, CEAM and University of Zaragoza that contributed to the field works.