Predicting Ecosystem Functional Properties at ICOS sites with hyperspectral PRISMA data using machine learning: a comparison between random forest and extreme gradient boosting

Consiglio Nazionale delle Ricerche

CNR IRET Istituto di Ricerca sugli Ecosistemi Terrestri

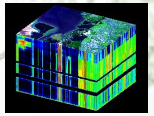
L. Nardella [1], A. Sebastiani [2], A. Montaghi [3], D. Papale [2], G. Vaglio Laurin [2]

[1] Research Institute on Terrestrial Ecosystems, National Research Council, Porano, Italy, [2] Research Institute on Terrestrial Ecosystems, National Research Council, Montelibretti Research Area, Italy, [3] Research Institute on Terrestrial Ecosystems, National Research Council, Sesto Fiorentino, Italy

Background

Ecosystem Functional Properties (EFPs) characterize key ecosystem processes (e.g. photosynthesis, respiration, nutrient or water cycles), and help monitor ecosystem response to biotic and abiotic factors, including climate change. EFPs are derived from Eddy Covariance (EC) fluxes of carbon, water and energy, collected in EU by the **ICOS network** at footprint (local) scale.

Innovative hyperspectral satellite remote sensing data (**PRISMA**) and derived **Vegetation Indices** (VIs), collect vegetation spectral response/health status in hundreds of fine bands, and can support the upscaling of EFPs over large regions.



Objectives

1. Test the capacity of PRISMA VIs to predict EFPs in different ecosystems/plant functional types (PFT);
2. Compare the results obtained from two different Machine Learning modelling approaches: **Random Forest** (RF) and **eXtreme Gradient Boosting** (XGB).

Methods

ICOS: 15 sites in 5 EU countries, 5 PFTs

EFPs elaborated by ICOS:

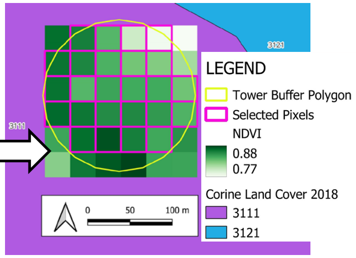
- **GPP** = Gross Primary Productivity
- **NEE** = Net Ecosystem Exchange

EFPs computed here:

- **LUE** = Light Use Efficiency: **GPP / SW_{IN}** (ShW. in. rad.)
- **WUE** = Water Use Efficiency: **GPP / LE** (latent heat)
- **BW** = Bowen Ratio: **H** (sensible heat)**/ LE**

Extraction of 29 PRISMA VIs over homogeneous areas

- Min **70%** for pixel inclusion
- **Area-based statistics** (NDVI-based homogeneity)



Modelling

1. **Default parameters:**
   - > **RF** (*randomForest* R)
   - > **XGB** (*XGB* Python)
2. **Hyperparameter tuning:**
   - > **RF** (*caret* R)
   - > **XGB** (*optuna* Python)
3. **Feature Selection (VSURF R)**
   - Hyperp. tuning and Cross Validation **RF** (*caret* R)
   - Hyperp. tuning and Cross Validation **XGB** (*optuna + DART* Python)

Results

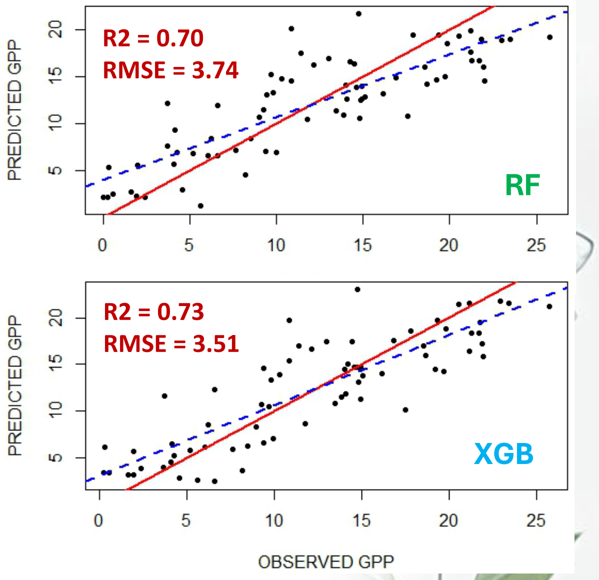
TUNING (post Feature Selection)

Model	HyperPar.	GPP	NEE	LUE	WUE	BW
RF	ntree	500	500	500	500	500
	nodesize	8	10	10	3	4
	mtry	2	3	2	1	2
XGB	eta	0.767	0.998	0.591	0.998	0.357
	lambda	1.87E-05	4.91E-08	0.334	6.37E-04	0.153
	alpha	0.0232	7.28E-07	2.13E-08	4.46E-05	2.34E-05
	gamma	2.11E-06	2.50E-03	1.57E-05	1.84E-08	3.20E-03
	max_depth	7	7	11	7	8
	max_leaves	320	60	870	990	210
	max_bin	448	448	576	512	704
	grow_policy	depthwise	depthwise	depthwise	lossguide	depthwise
	min_child_weight	14	13	7	15	8
	max_delta_step	6	6	7	6	5
	subsample	1	0.9	0.5	0.5	1
	colsample_bylevel	1	0.7	0.7	0.7	0
	colsample_bytree	0.9	0.5	0.9	0.8	0.9
	rate_drop	0.070	0.357	0.191	0.909	0.023
	n. estimators	1040	1420	1340	900	970

EFP	Selected VIs
GPP	VOG, RENDVI, IRECI, OSAVI, Simple_Ratio, NIRv, CAI **(7)**
NEE	VOG, IRECI, OSAVI, CAI **(4)**
LUE	VOG, IRECI, CAI **(3)**
WUE	VIgreen_Index, VARI **(2)**
BW	EVI, NIRv, NDLI, MCARI, SATVI, CRI, ARVI **(7)**

EFP	Metric	RF	XGB
GPP	R2	0.70	**0.73**
	RMSE	3.74	**3.51**
NEE	R2	0.58	**0.58**
	RMSE	3.54	**3.49**
LUE	R2	0.58	**0.61**
	RMSE	0.01	**0.01**
WUE	R2	-0.33	**-0.02**
	RMSE	0.06	**0.05**
BW	R2	**0.34**	0.32
	RMSE	**1.66**	1.69

MODEL PERFORMANCE: GPP



Conclusion

Results show that PRISMA VIs can predict with good accuracy GPP, NEE and LUE in EU independently on the natural ecosystems considered (wetlands, grasslands, or forests). Further studies exploiting other VIs are ongoing, to assess the lower accuracy obtained by WUE. VOG, IRECI, NIRv resulted frequently selected, highlighting which spectral regions mostly contributed to accurate models. Hyperparameter tuning improved performances for both RF and XGB models in all cases. Extreme gradient boosting provides a more sofisticated tuning framework which improves model perfomances in most cases.

LN acknowledges the project PRIN 2020 "MULTIFOR - Multi-scale observations to predict Forest response to pollution and climate change_PRIN_2020_LS9". GVL and DP acknowledge the PNRR, Missione 4, Componente 2, Avviso 3264/2021, IR0000032—ITINERIS - Italian Integrated Environmental Research Infrastructures System CUP B53C22002150006.