REPLACING RADIATIVE TRANSFER MODELS BY SURROGATE APPROXIMATIONS THROUGH MACHINE LEARNING

J. Verrelst*, J.P. Rivera, J. Gómez-Dans, G. Camps-Valls and J. Moreno
Image Processing Laboratory (IPL), University of Valencia, Spain

Laboratorio de Procesado de Imágenes

*Jochem Verrelst
Image Processing Laboratory
University of Valencia, Spain
jochem.verrelst@uv.es

1. Introduction

Physically-based radiative transfer models (RTMs) help in understanding the processes occurring on the Earth's surface and their interactions with vegetation and atmosphere. However, advanced RTMs can take a long computational time, which makes them unfeasible in many real applications. To overcome this problem, it has been proposed to substitute RTMs through so-called ***emulators***. **Emulators are statistical models that approximate the functioning of RTMs.** They are advantageous in real practice because of the computational efficiency and excellent accuracy and flexibility for extrapolation. We here present an '**Emulator toolbox**' that enables analyzing three **multi-output machine learning regression algorithms (MO-MLRAs)** on their ability to approximate an RTM. As a proof of concept, a case study on emulating sun-induced fluorescence is presented. The toolbox is foreseen to open new opportunities in the use of advanced RTMs, in which both consistent physical assumptions and data-driven machine learning algorithms live together.

Objective: to present a Emulator toolbox that enables building surrogate models that approximate radiative transfer models through MO-MLRAs.
Thereby related are the following goals:

1. to evaluate multiple MO-MLRAs on their performance to function as an emulator and as a proof of concept;
2. to apply the best performing MO-MLRA as the emulator to approximate SCOPE.

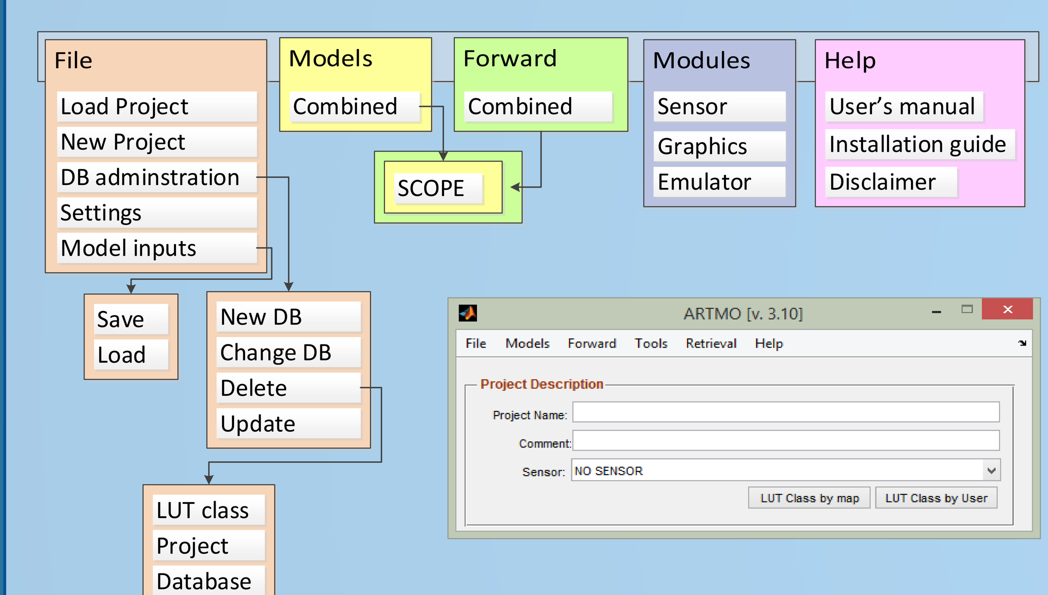
6. Conclusions

To facilitate the use of emulators, **ARTMO's new Emulator toolbox** enables analyzing three multi-output machine learning regression algorithms (MO-MLRAs), both linear (PLSR and nonlinear (KRR, NN). The toolbox enables the user to train the MO-MLRA models with data coming from RTMs that are available within ARTMO. Options are provided to optimize the training phase, such a PCA pre-processing step, ranging training/validation distributions or through cross-validation sub-sampling procedures. Performance and processing speed of the MO-MLRAs are then calculated. A successfully validated MO-MLRA can function as emulator.
We analyzed the ability of the implemented MO-MLRAs to substitute the SVAT model SCOPE in the generation of sun-induced fluorescence (SIF) outputs. NN and KRR emulated SIF profiles with great precision (relative errors below 0.5% when trained with 500 or more samples), and this **with a gain in processing speed** of about 50 (NN) **up to about 800 (KRR) times faster than SCOPE v1.60**.

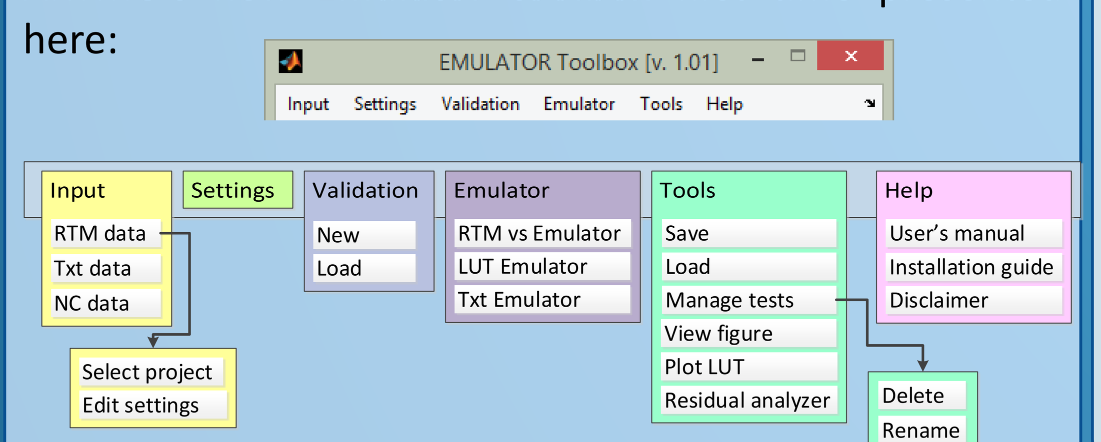
The emulator toolbox opens up a diverse range of new applications using advanced RTMs, such as improved inversion strategies, and fast rendering of simulated scenes in preparation for new satellite missions.

2. ARTMO

ARTMO brings multiple leaf and canopy radiative transfer models (RTMs) together along with essential tools required for semi-automatic retrieval of biophysical parameters in a modular toolbox. The software package is freely downloadable at http://ipl.uv.es/artmo.



ARTMO's new **Emulator toolbox** first time presented here:



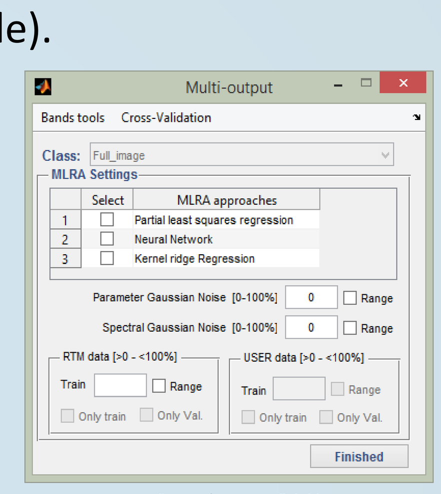
3. Emulator toolbox:

Input → **Settings** → **Validation** → **Emulator**

Input: An ARTMO-generated LUT **or** external LUT (.txt file).

Settings:
The following MO-MLRAs have been implemented:

- **Partial least squares regression (PLSR)**
- **Neural networks (NN)**
- **Kernel ridge regression (KRR)**

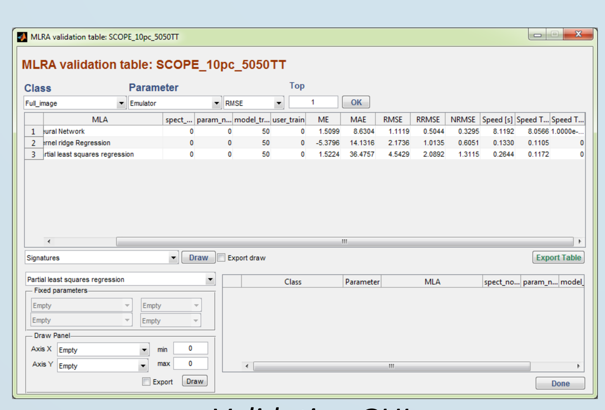

Settings GUI

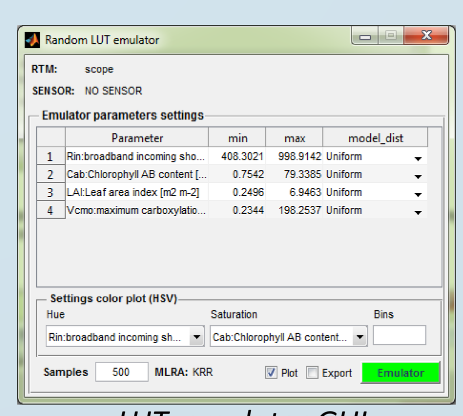
The following options are provided:

- **PCA** (greatly speeds up the training phase)
- **Cross-validation sub-sampling**. Enables a more robust assessment of the accuracy.

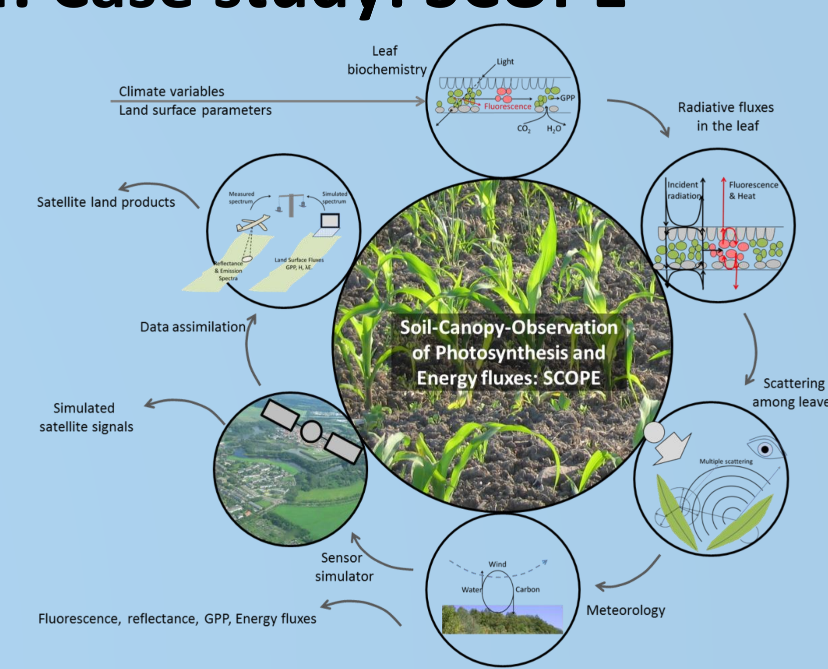
Multiple MO-MRLAs can be validated at once.

Validation: An overview table with RMSE accuracies is provided. The most accurate MO-MLRA can then function as **emulator**.


Validation GUI


LUT emulator GUI

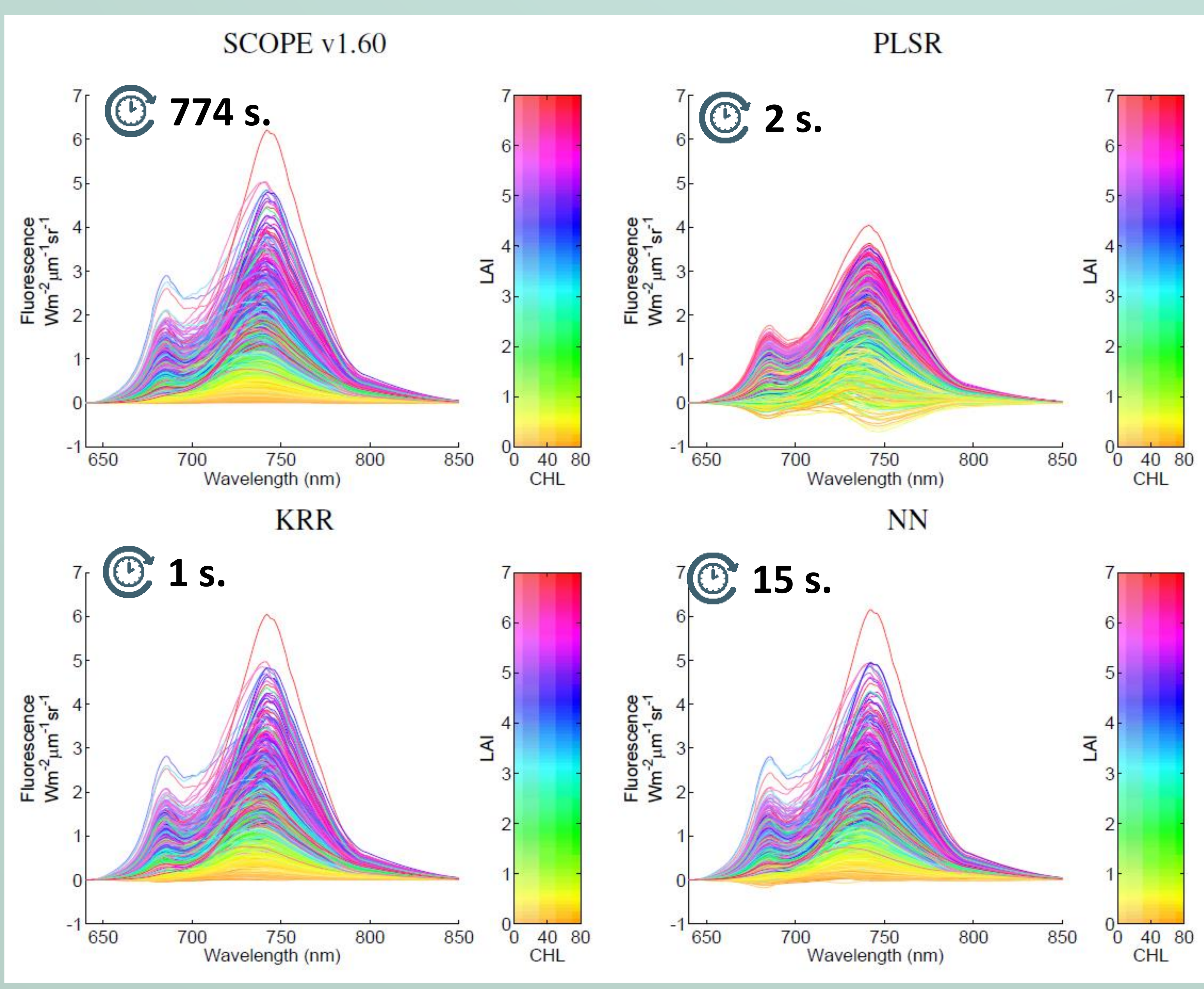
4. Case study: SCOPE



SCOPE is a vertical (1-D) integrated radiative transfer and energy balance SVAT model, with sun-induced chlorophyll fluorescence (SIF) as one of their outputs. A LUT of 1000# entries were generated based on the most important input variables.

	Variable Names	Units	Range
Leaf biochemistry			
Vcmo	Maximum carboxylation capacity	μmol m^{-1} s^{-1}	0.1 to 100
Leaf variables			
CHL	Leaf chlorophyll content	μg/cm^2	0 to 80
C_m	Leaf dry matter content	g/cm^2	0.001 to 0.05
Canopy variables			
LAI	Leaf area index	m^2/m^2	0.01 to 7
rwc	Within-canopy-layer resistance	m^2/m^2	0 to 20
SZA	Solar zenith angle	$\circ$	0 to 60
Micrometeorology variables			
Ca	CO_2 concentration in the air	ppm	350 to 450
P	Air pressure	hPa	1000 to 1090
ea	Atmospheric vapour pressure	hPa	10 to 50
Ta	Air temperature	°C	5 to 25
Rin	Incoming shortwave radiation	W m^{-2}	400 to 1000

5. Emulation results

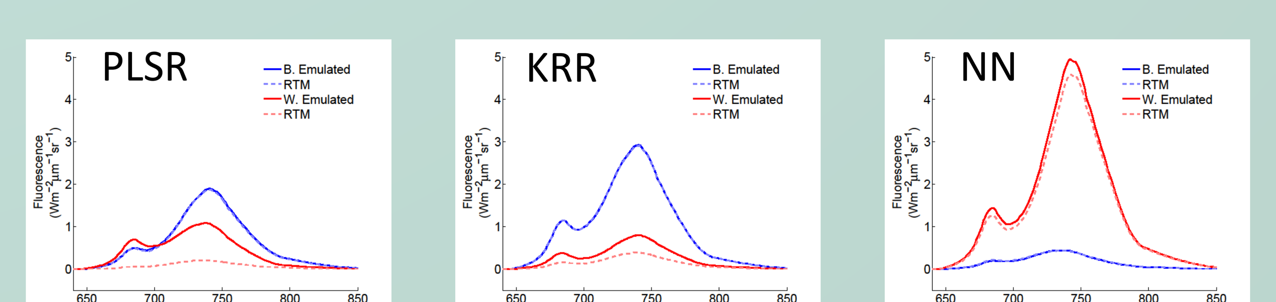


Validation SCOPE fluorescence emulation

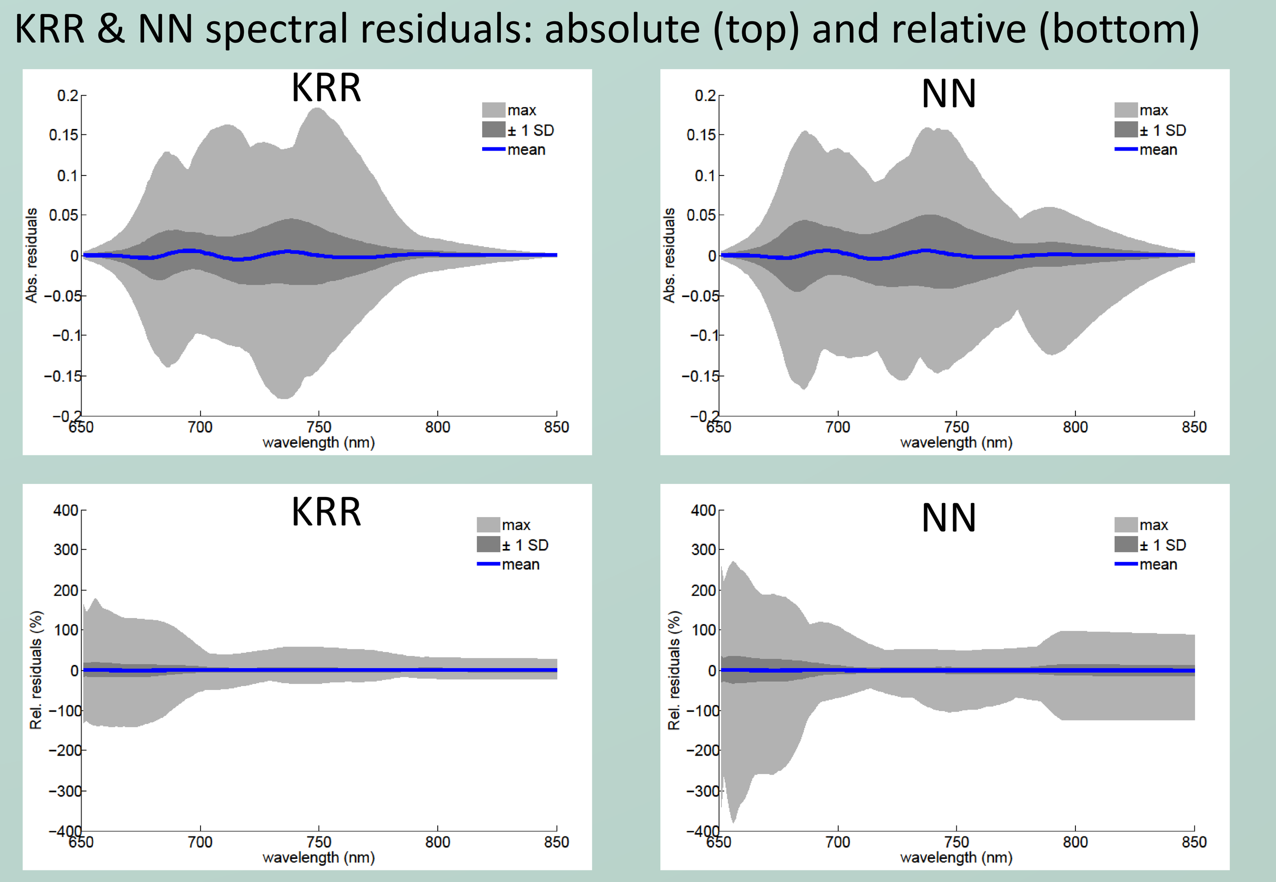
MO-MLRA	$RMSE_{CV}$	$NRMSE_{CV}$ (%)	Speed Training (s)	Speed Validation (s)	Gain in Speed (x)
# 1000					
PLSR	2.99	1.03	0.58	0.05	423
KRR	0.85	0.29	7.88	0.05	790
NN	0.64	0.22	65.56	0.06	51

KRR and NN deliver accurate SIF emulations. PLSR not recommended as emulator.

Inspecting best (B) and worst (W) emulated fluorescence profiles



KRR & NN spectral residuals: absolute (top) and relative (bottom)



Rivera, J.P., Verrelst, J., Gómez-Dans, J., Muñoz-Marí, J., Moreno, J., Camps-Valls, J. 2015. An emulator toolbox to approximate radiative transfer models with statistical learning. Remote Sensing. *In press.*