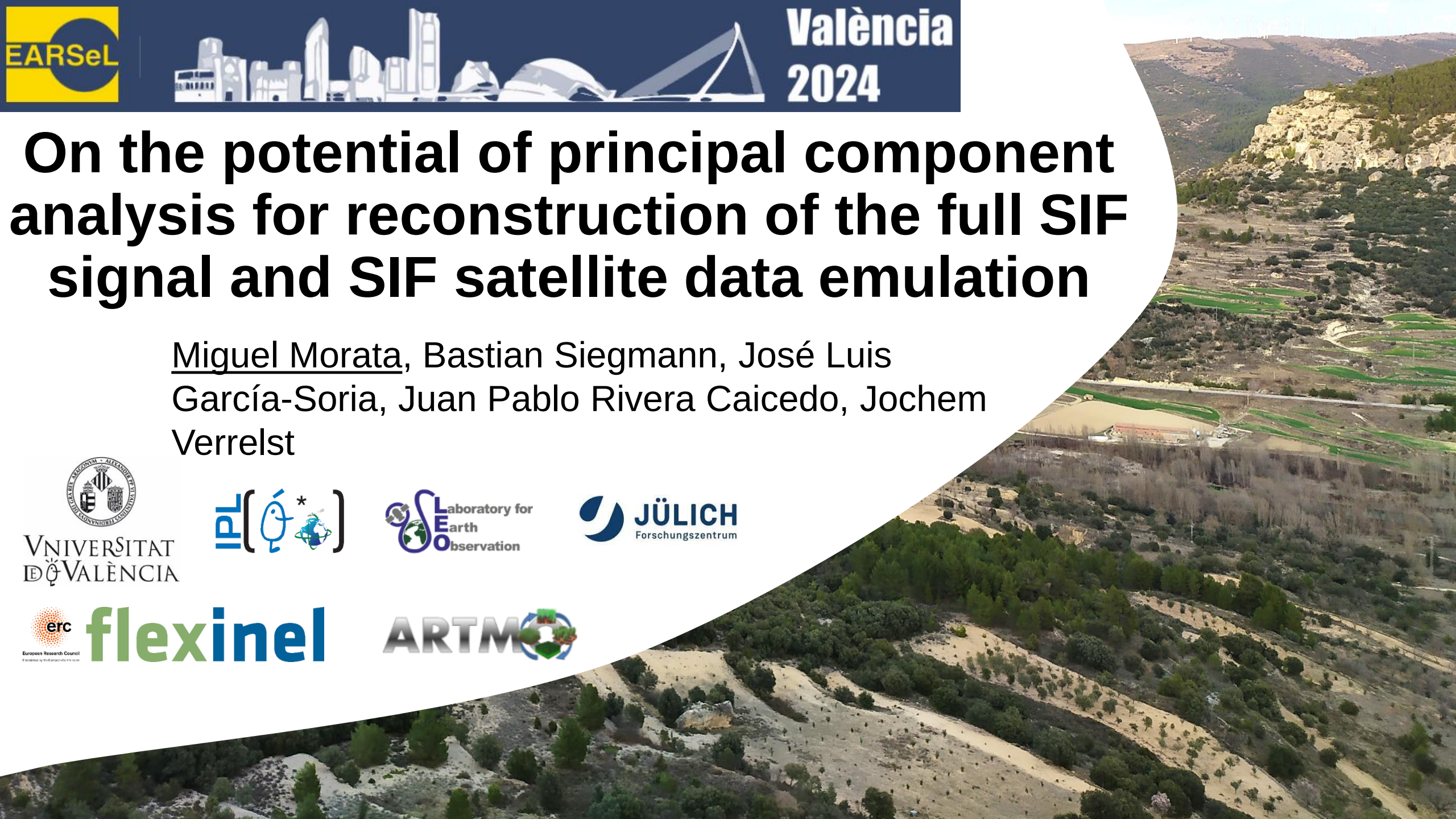
EARSeL València 2024

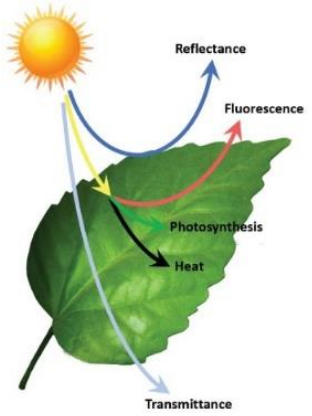
On the potential of principal component analysis for reconstruction of the full SIF signal and SIF satellite data emulation

Miguel Morata, Bastian Siegmann, José Luis García-Soria, Juan Pablo Rivera Caicedo, Jochem Verrelst



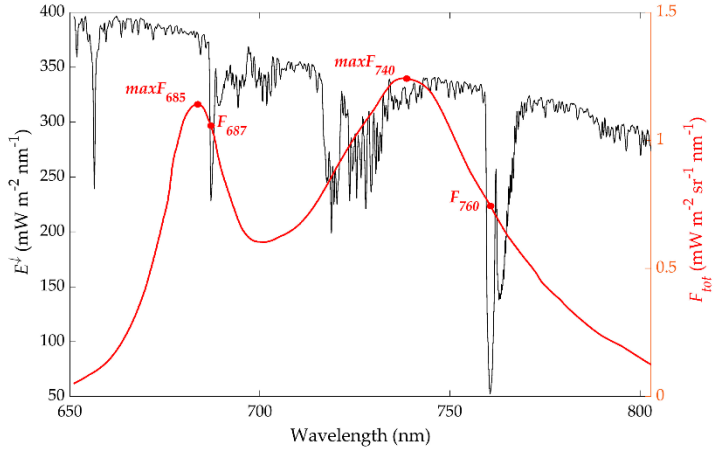
Introduction

- **Photosynthesis**, a critical process sustaining life on Earth, releases solar-induced fluorescence (SIF) as a byproduct.



- Methods like the **Spectral Fitting Method (SFM)** retrieve SIF only in narrow spectral absorption features (O_2A, O_2B).

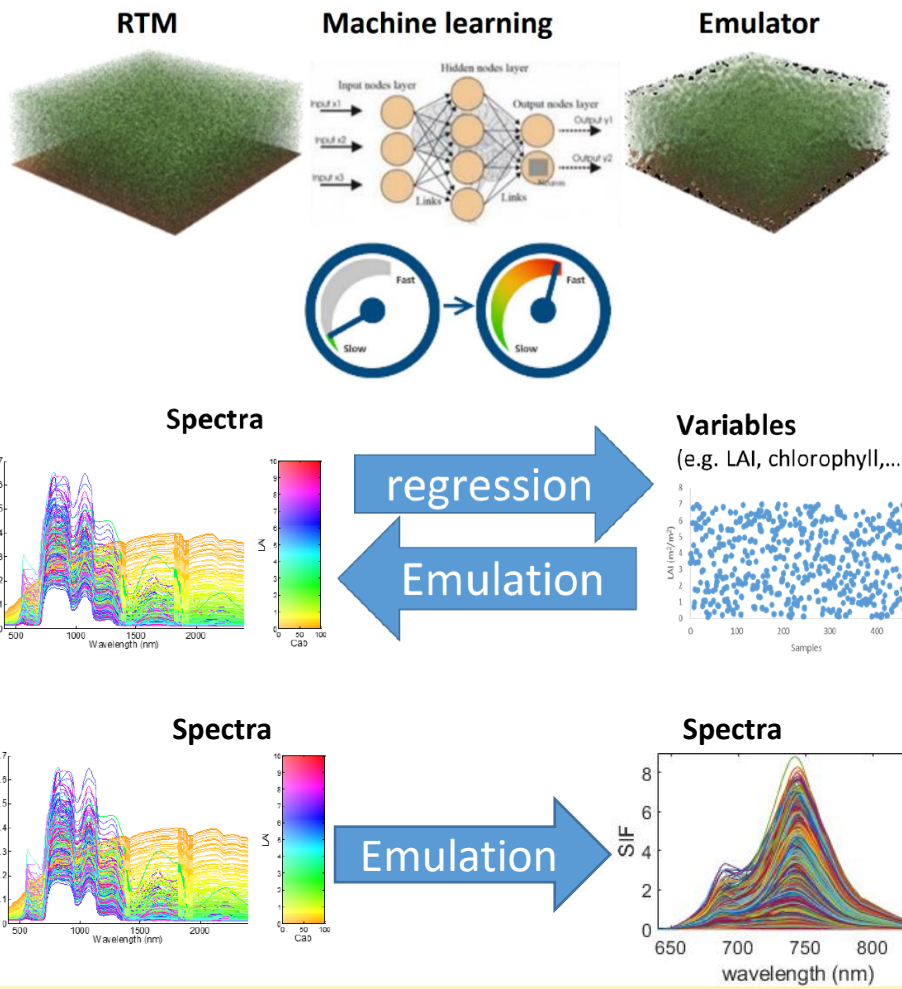
- Approach to reconstructing the **full-spectral SIF** emission from SIF_{687} and SIF_{760} using **SCOPE** and **Principal Component Analysis (PCA)**.



- Use of PRISMA hyperspectral images to emulate **Solar Induced Fluorescence (SIF)**.

Emulation

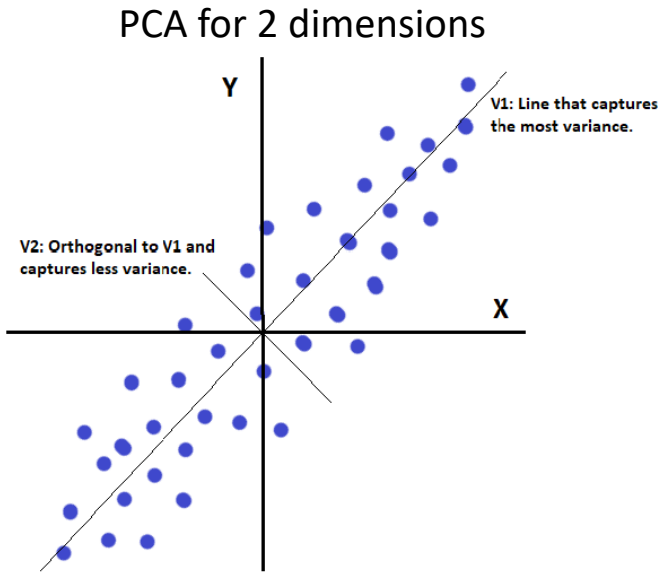
- **Emulators** are **surrogate** statistical models that are able to approximate a complex processing at a **fraction of the computational cost**.
- The principle of emulation is approximating the input-output relationships by means of a **statistical learning model**, also referred to as an emulator.
- Multi-input and multi-output from emulation allow for obtaining **non-linear relations** between different spectra.

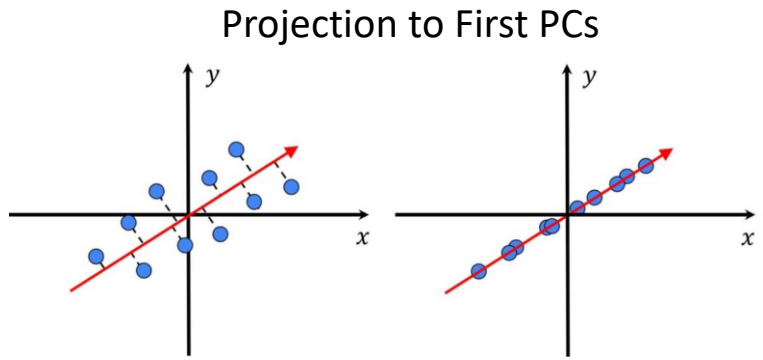


Rivera et al. (2015). An Emulator Toolbox to Approximate Radiative Transfer Models with Statistical Learning. Remote Sensing. p. *7*, 9347-9370.

Principal Component Analysis (PCA)

- PCA as a **linear Dimensionality Reduction** (DR) technique.
- If the variables/dimensions are **highly correlated**, we can reduce the dimensionality keeping most of the original information.
- **Project** the original spectra onto a set of vectors known as **Principal Components** (PCs).
- The **eigenvectors matrix** is then used as a projection matrix.
- **Reconstruct** the signal in original feature space.

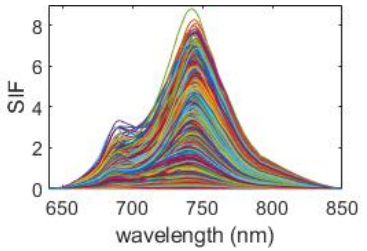




SCOPE

- Soil Canopy Observation, Photochemistry and Energy fluxes (SCOPE) is a widely used RTM for simulating the radiative transfer in vegetation.
- Based on **physical** calculations, it can simulate the behavior of light in complex systems.
- **Full-spectrum-SIF** for each combination of input variables.
- 640nm to 850nm at 1nm resolution to a total of 211 bands.
- **LUT of 100,000** spectra employing Latin Hypercube Sampling (LHS).

SCOPE LUT

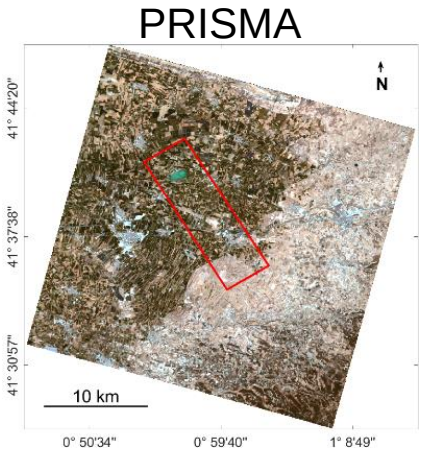


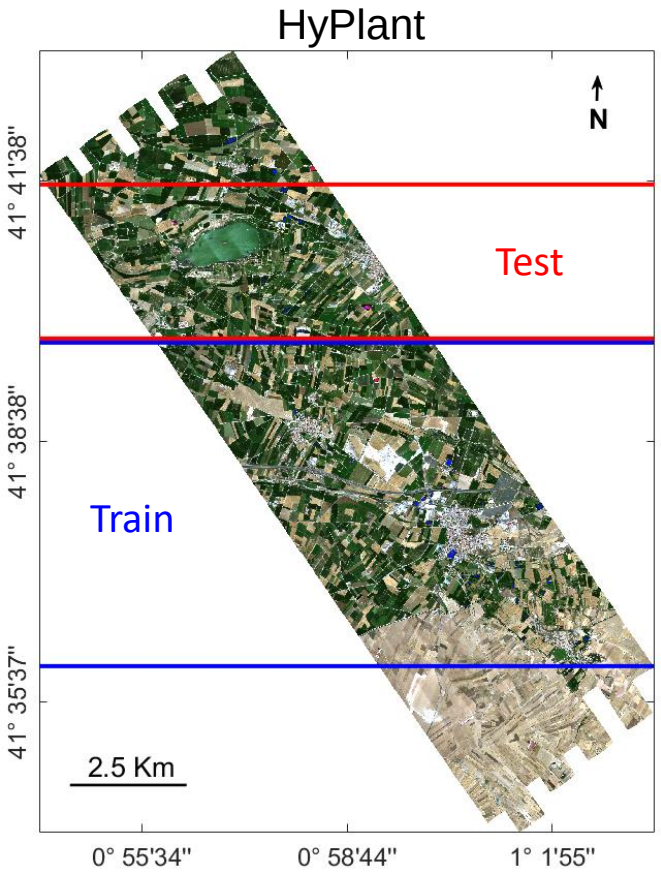
Type	Biophysical parameter	Min.	Max.
Weather	Incoming shortwave radiation (W/m^2)	0	1000
Weather	Air temperature (ºC)	0	40
Weather	Atmospheric vapor pressure (hPa)	0	100
Weather	CO_2 concentration in the air (ppm)	100	800
Leaf	Chlorophyll ($\mu g/cm^2$)	0	50
Leaf	Dry matter (g/cm^2)	0	0.05
Leaf	Fraction of senescent material	0	0.3
Biochemistry	Maximum carboxylation capacity ($\mu molm^{-1}s^{-1}$)	0	150
Biochemistry	Ball-Berry stomatal conductance	5	15
Soil	BSM model parameter for soil brightnes	0	0.9
Canopy	Leaf area index (m$^2/m^2$)	0	7
Canopy	Vegetation height (m)	0.05	100
Canopy	Leaf width (m)	0.01	2
Canopy	Leaf inclination distribution factor a	-1	1
Canopy	Leaf inclination distribution factor b	-1	1
Geometry	Solar zenith angle (º)	20	60

Verrelst et al. (2015). Global sensitivity analysis of the SCOPE model: What drives simulated canopy-leaving sun-induced fluorescence? Remote Sensing of Environment, 166, 8-21.

Materials & Methods

Used data: HYFLEX

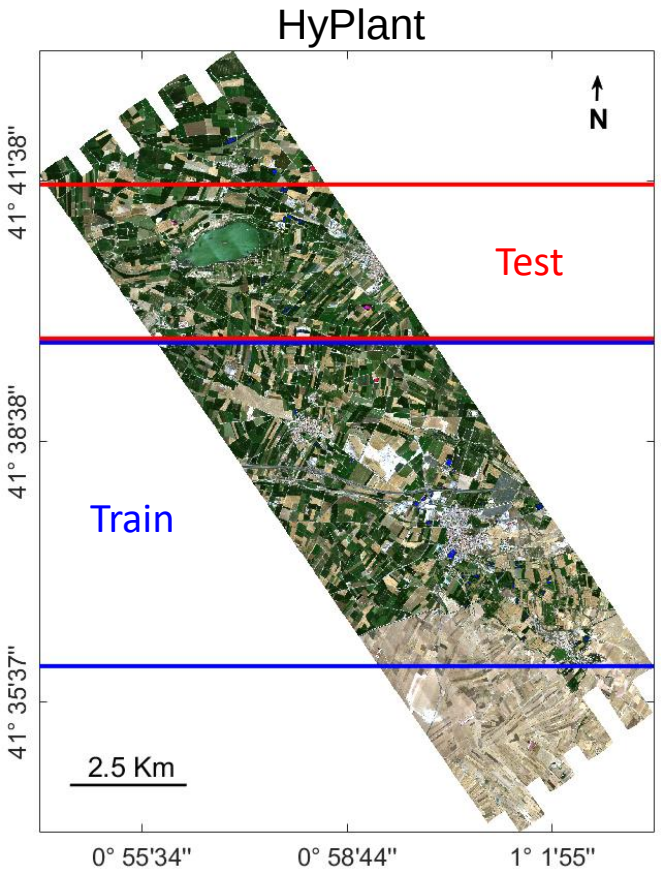


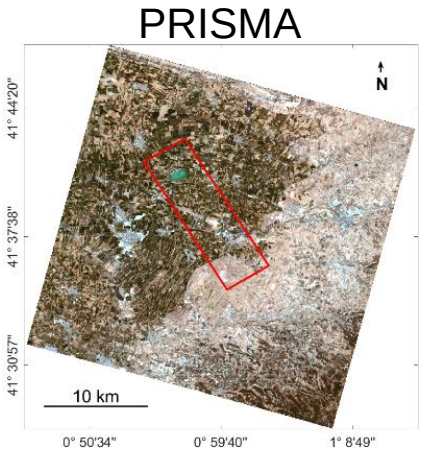


- **Study site:** West of Lleida in northeast Spain. Border area between an irrigated and rain-fed semi-arid landscape. Flight campaign of 9 HyPlant flight lines mosaicked to one image. The image was split into two subsets, Train and Test.
- **Sensors: PRISMA** hyper-spectral spaceborne image as input, and **HyPlant reconstructed full-spectrum SIF** product as output.
- **Date acquisition:** 17th of July 2021.

Materials & Methods

Used data: HYFLEX



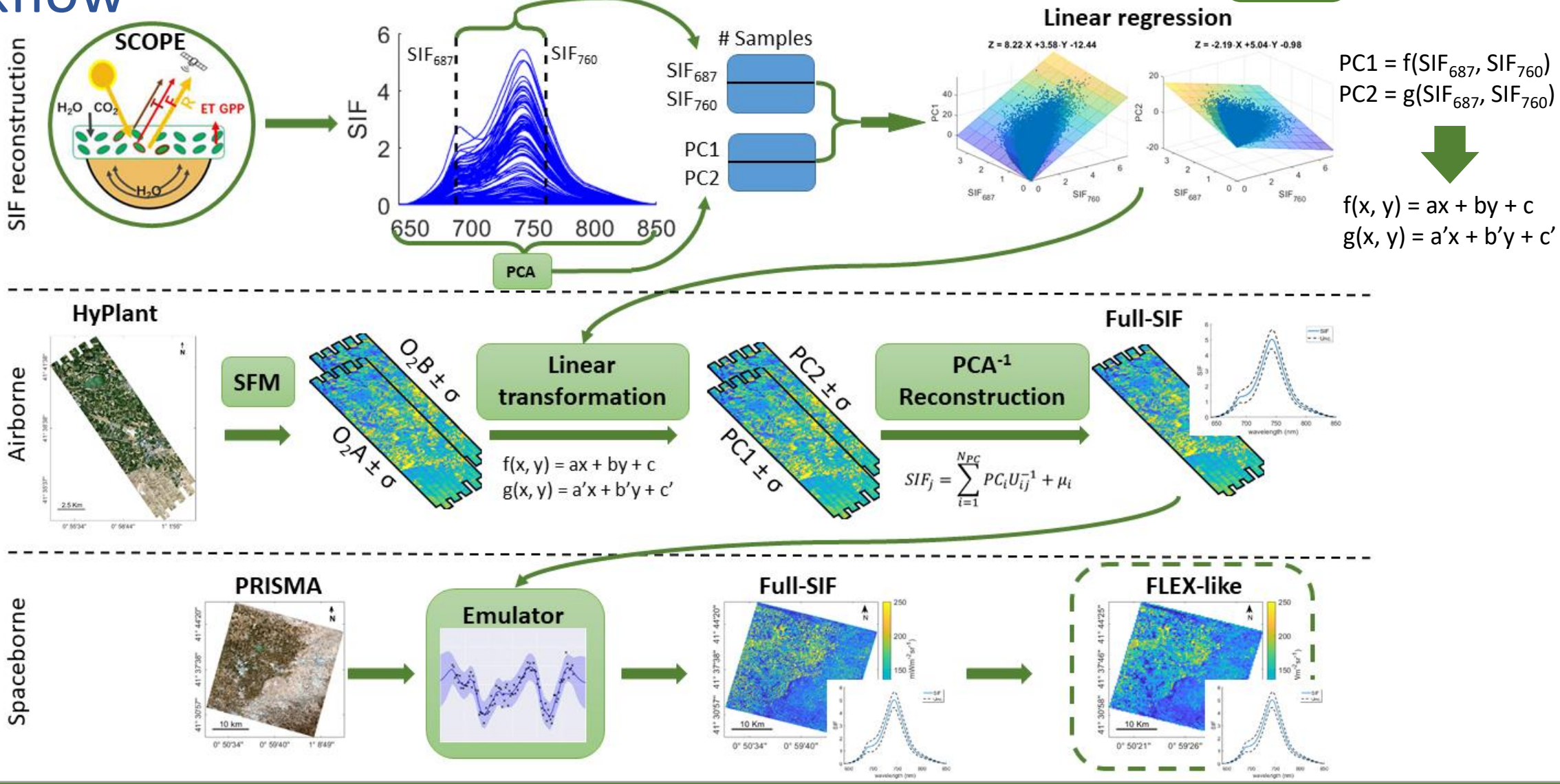


Images:

- **HyPlant:** SIF-SFM product of SIF_{687} and SIF_{760} with uncertainties. Spatial resolution of 1.7m.
- **PRISMA:** Reflectance data (L2D) with 231 bands in the range of 402 to 2497nm. Spatial resolution of 30m.

Materials & Methods

Workflow



Uncertainty SIF reconstruction

- **Linear-based reconstruction**.
- In case SFM products have **uncertainty**, we can **propagate** it through the full-spectrum SIF reconstruction.
- Error propagation using the **partial derivative method**.
- Propagation error in **linear transformation** obtained by linear regression.
- Propagation error in **inverse PCA** transformation.

$$U^{-1} \equiv PCA^{-1}$$

$$A, B \equiv SIF_{760}, SIF_{687}$$

$$\varepsilon_Y = \sqrt{\sum_{k=1}^{N} \left(\frac{\partial Y}{\partial X_k}\right)^2 \varepsilon_{X_k}^2}$$

$$PC_i = a_i A + b_i B + c_i$$

$$\varepsilon_{PC_i} = \sqrt{(A\varepsilon_{a_i})^2 + (a_i\varepsilon_A)^2 + (B\varepsilon_{b_i})^2 + (b_i\varepsilon_B)^2 + \varepsilon_{c_i}^2}$$

$$SIF_j = \sum_{i=1}^{N_{PC}} PC_i U_{ij}^{-1} + \mu_i$$

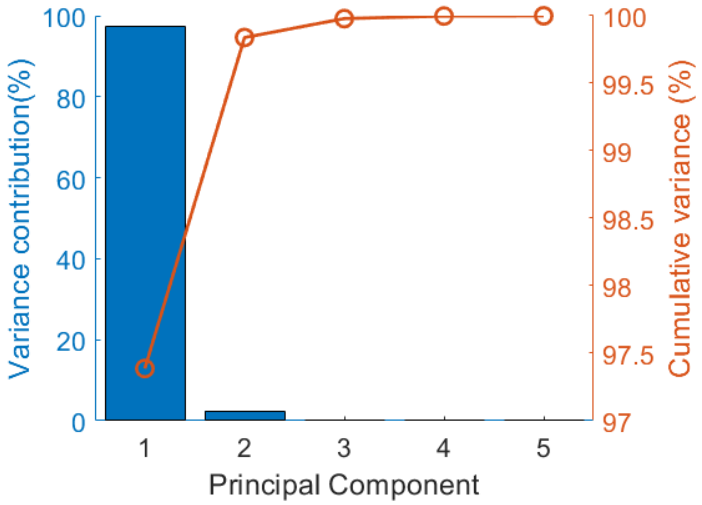
$$\varepsilon_{SIF_j} = \sqrt{\sum_{i=1}^{N_{PC}} \left(U_{ij}^{-1}\varepsilon_{PC_i}\right)^2}$$

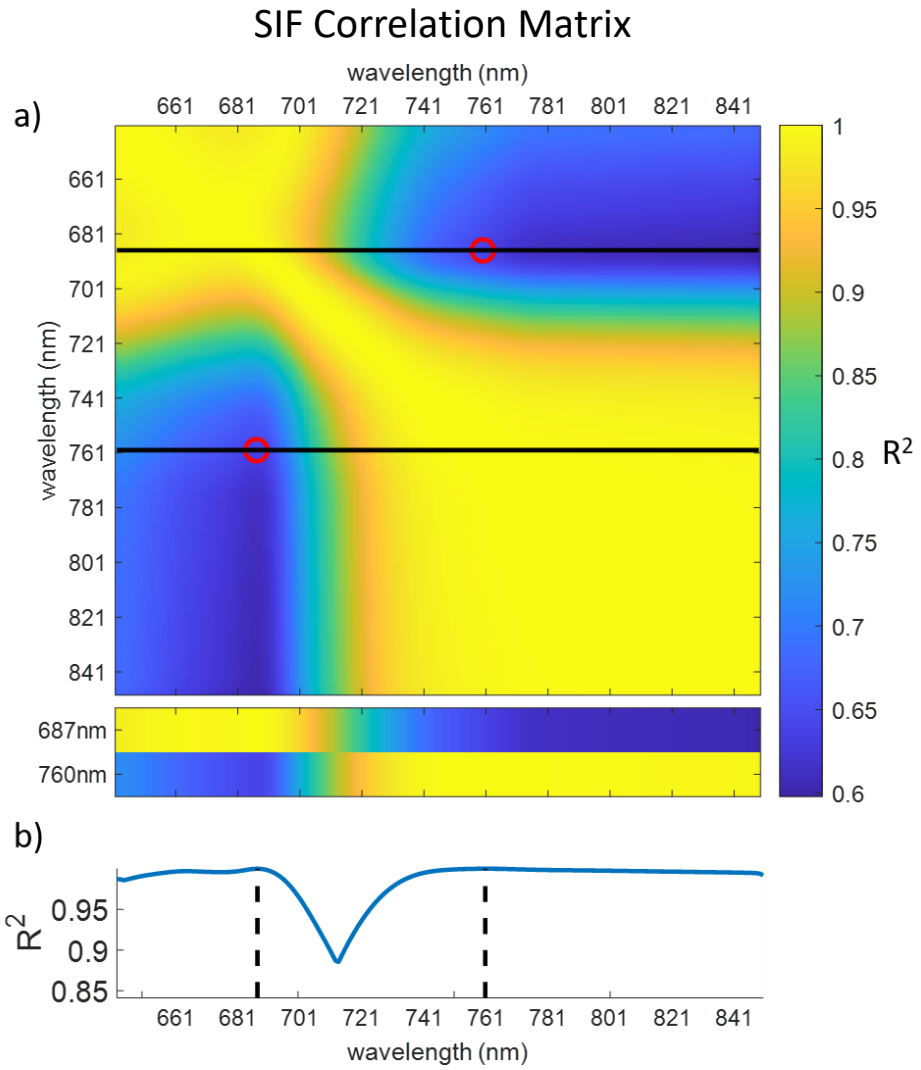
Morata et al. (2023). Neural Network Emulation of Synthetic Hyperspectral Sentinel-2-Like Imagery With Uncertainty. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. vol. 16, pp. 762-772.

Results: Theory

Correlation matrix

- **SIF Correlation matrix** between SIF 211 bands.
- 687 and 760nm are **highly correlated** with the rest of the bands, together R^2 higher than 0.85.
- Two PCs represent the **99.84%** variance.
- Low correlation between SIF_{760} and SIF_{687}.
- **Independent** variables

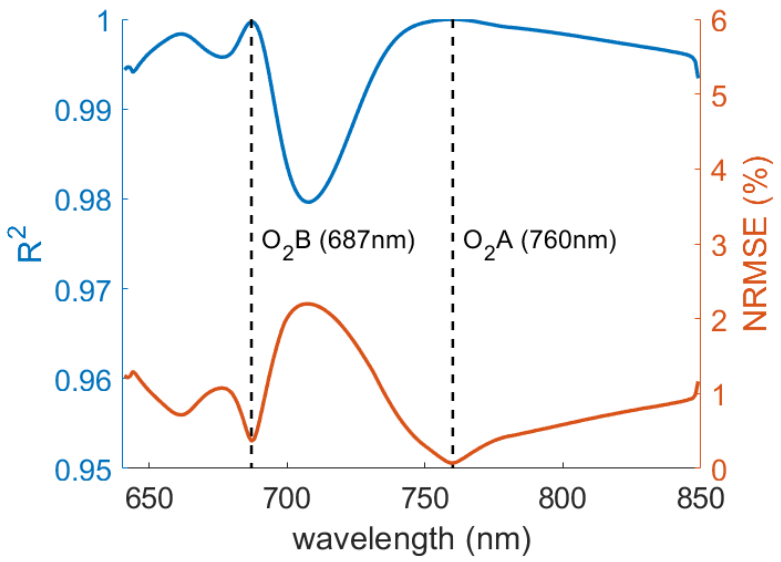


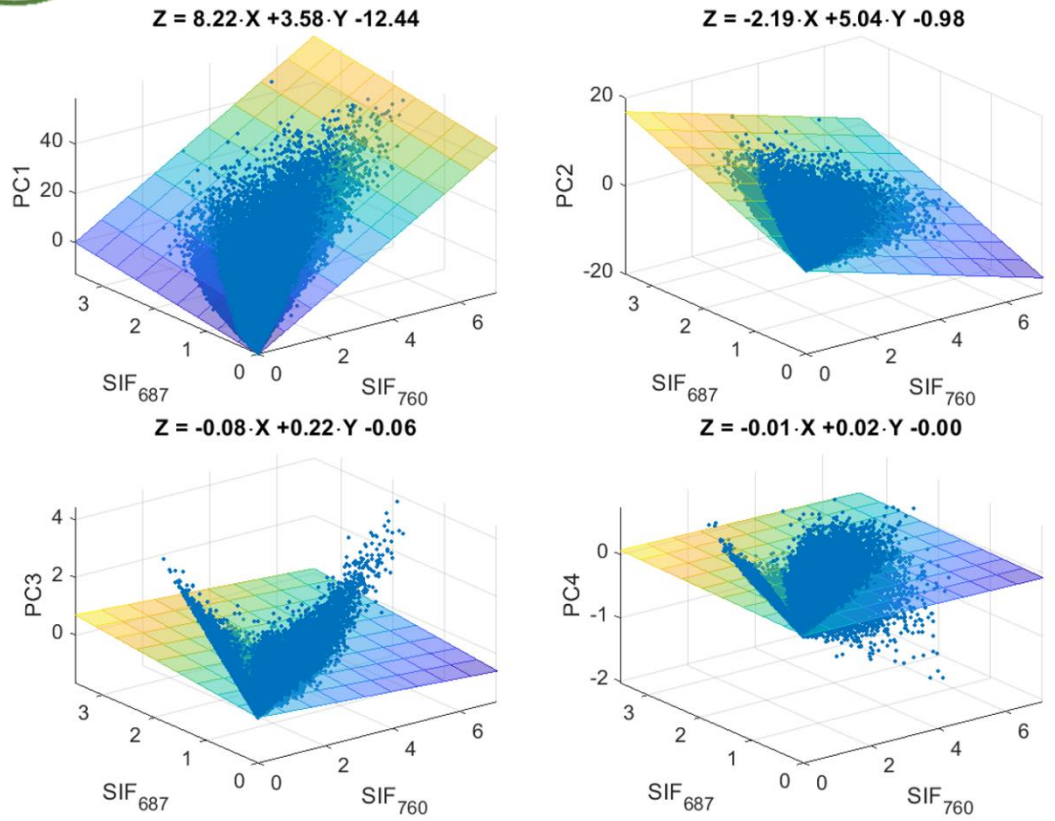


Results: Theory

Full-spectrum SIF Reconstruction

- Linear regression to obtain the f(SIF_{760},SIF_{687}) and g(SIF_{760},SIF_{687}).
- The prediction statistics of the first components are very significant.
- PC3 is **non-linear**.
- We use only the **two first PCs**.
- Reconstruct **Full-spectrum SIF** and obtained statistics.



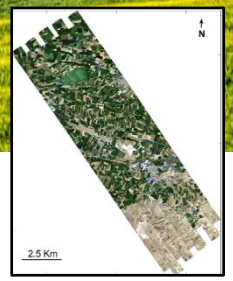


Prediction statistics

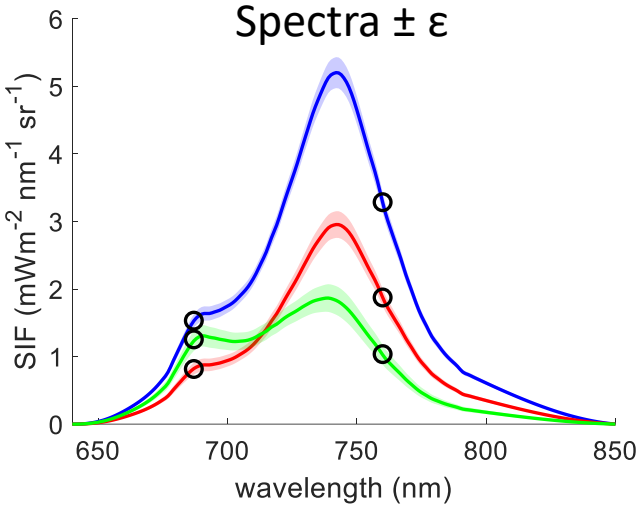
PCs	RMSE	R^2	NRMSE (%)	SLOPE (1:1)
PC1	0.47	0.998	0.78	0.998
PC2	0.27	0.969	1.43	0.969
PC3	0.37	0.031	6.00	0.031
PC4	0.14	0.002	4.89	0.002

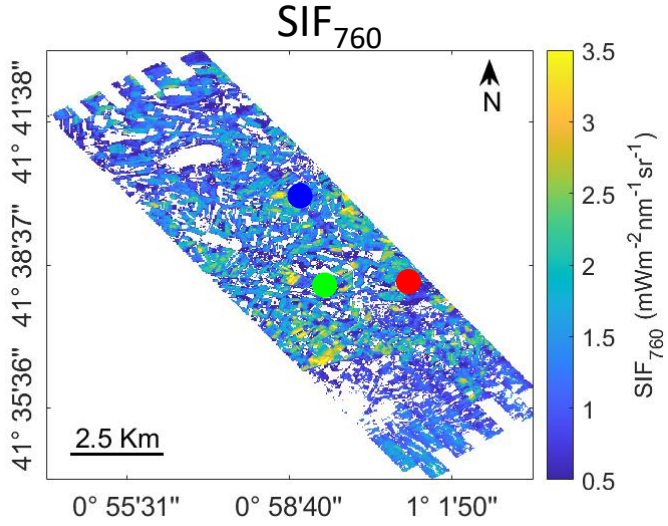
Results: Airborne

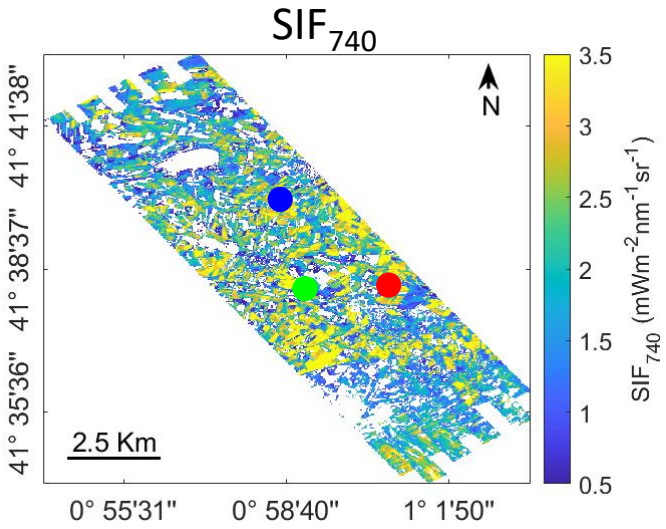
Full-spectrum SIF HyPlant image

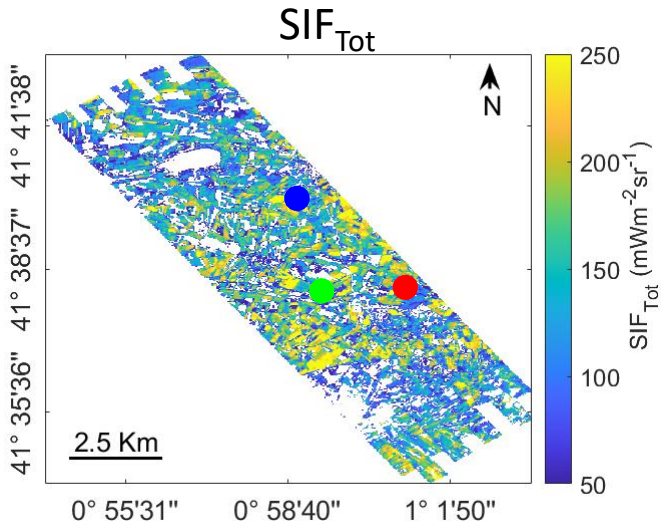


- Implement the novel technique we had developed.
- Obtaining a **full-spectrum SIF datacube** from the original HyPlant image.
- As the technique is only based on SCOPE-LUT, it's applicable to every HyPlant image.

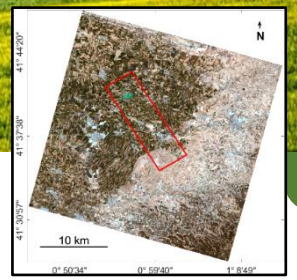






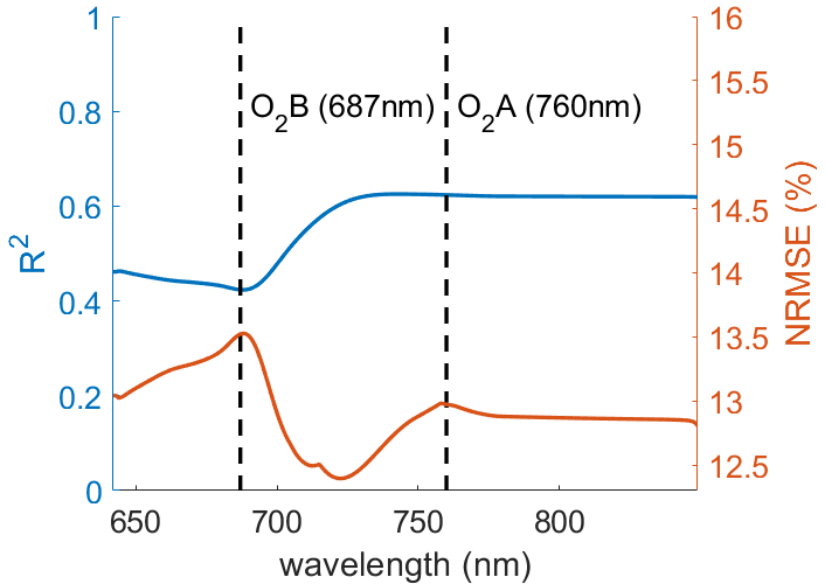


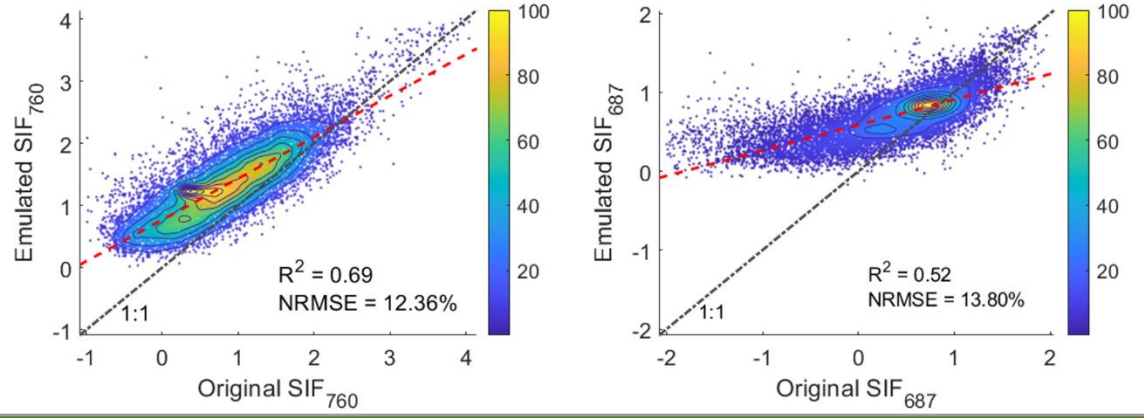
Results: Spaceborne



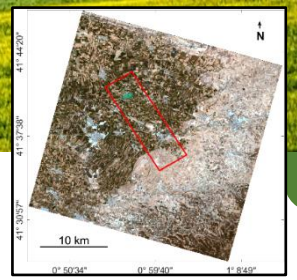
Full-spectrum SIF PRISMA emulation

- **Emulation** with PRISMA at the input and reconstructed SIF at the output.
- PCA as DR at the input PRISMA spectra. 27PCs represent 99.0% of the radiance PRISMA information.
- **KRR** with bootstrapping, 30PCs at input, 2PCs at output, 30,000 randomly samples from Train subset and 5 k-fold cross-validation.
- Validation with Test subset.





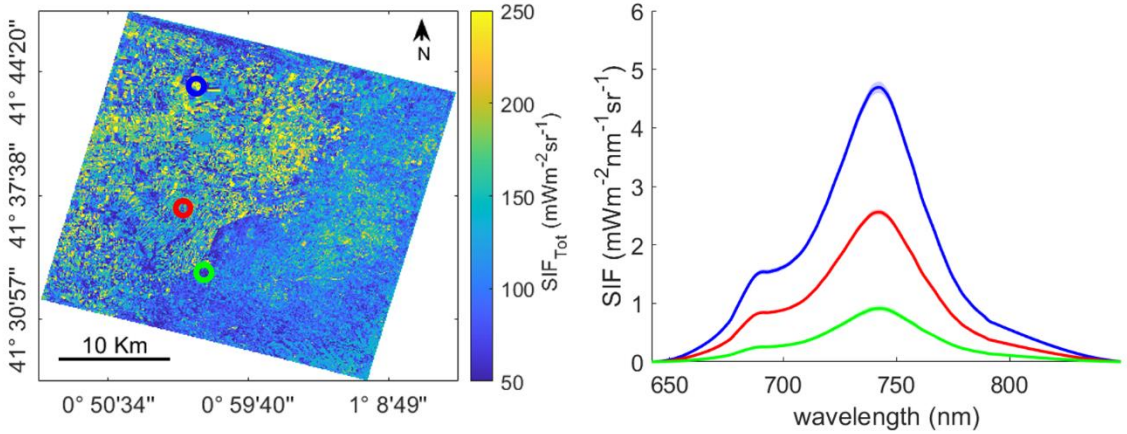
Results: Spaceborne



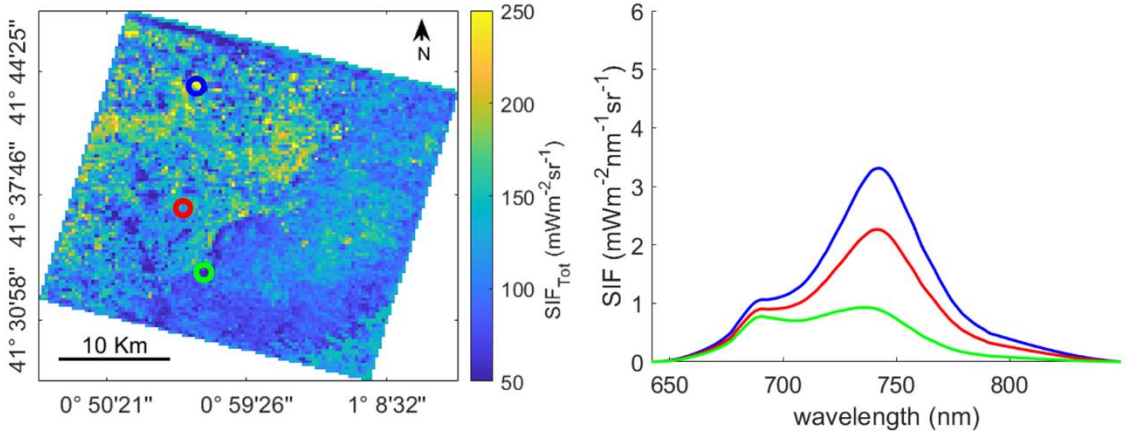
Full-spectrum SIF PRISMA emulation

- Application of emulator to a complete **PRISMA tile**.
- A standard computer took a **few minutes** to generate the **full-spectrum SIF synthetic scene**.
- PRISMA tile datacube with **full-spectrum SIF synthetic spectra at 211 bands and uncertainty in all bands**.
- Resampled to **FLEX** spatial resolution.

PRISMA Full-spectral SIF (30m)



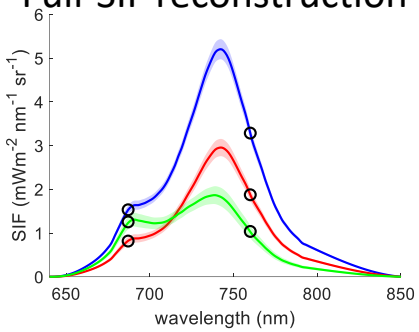
FLEX-like Full-spectral SIF (300m)



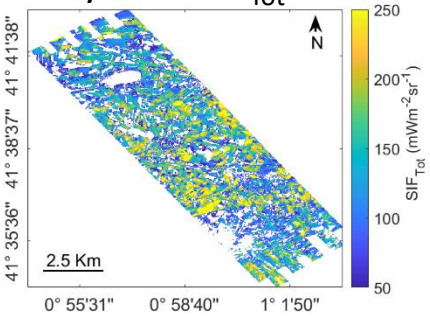
Conclusions

- **A novel technique to reconstruct full-spectrum SIF** from retrieved values at O_2A and O_2B.
- Reliable **uncertainty propagation** during all processing, obtaining uncertainty for the **full-spectrum SIF** emission.
- **Powerful** and reliable **fast** generation of **full-spectrum SIF** images covering a large area for every sensor that provides SIF in O_2A and O_2B regions.
- **Emulation** of synthetic **full-spectrum SIF** for spaceborne sensors.

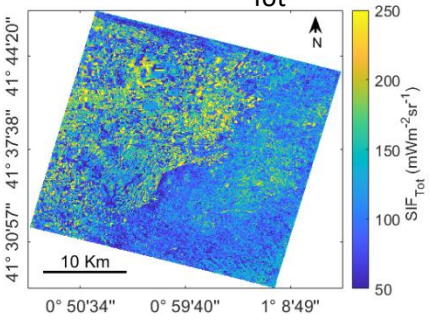
Full-SIF reconstruction



HyPlant $SIF_{Tot} \pm \varepsilon$



PRISMA $SIF_{Tot} \pm \varepsilon$



Many thanks for your attention!

Questions

flexinel

https://leoipl.uv.es/flexinel/

Email: miguel.morata@uv.es

Image Processing Laboratory (IPL)
University of Valencia
Valencia, Spain