VNIVERSITAT DE VALÈNCIA IPL

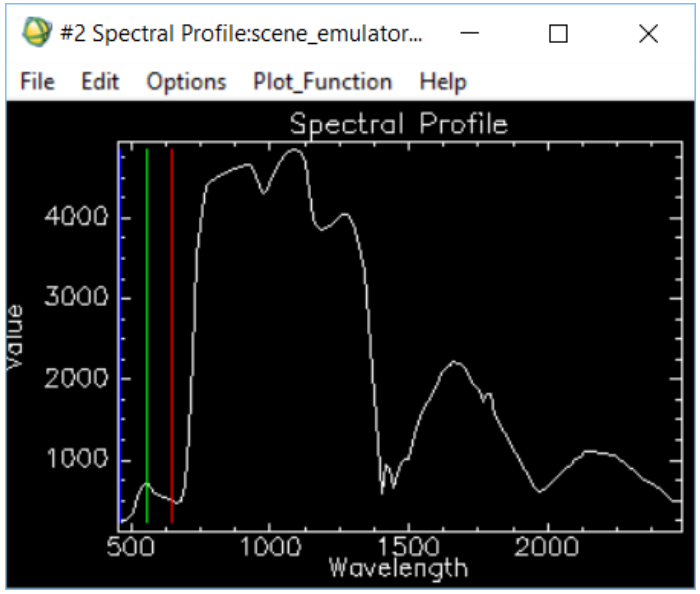
Laboratory for Earth Observation

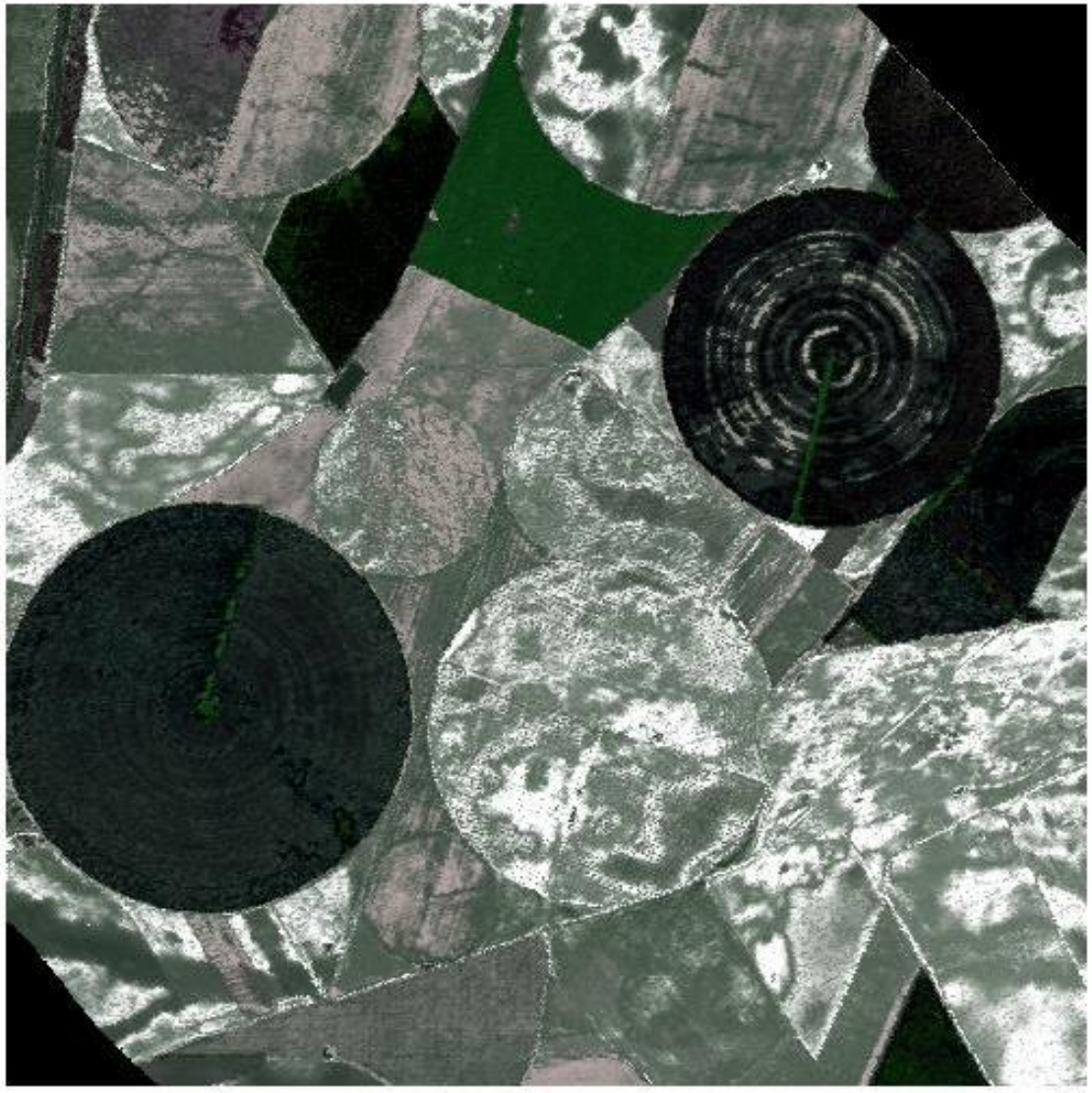
APPROXIMATING EXPERIMENTAL VEGETATION SPECTROSCOPY DATA THROUGH EMULATION

Jochem Verrelst, Juan Pablo Rivera, Jorge Vicent, Pablo Morcillo & Jose Moreno

IPL Image Processing Laboratory, University of Valencia (Spain)

Hyperspectral image (subset Barrax)

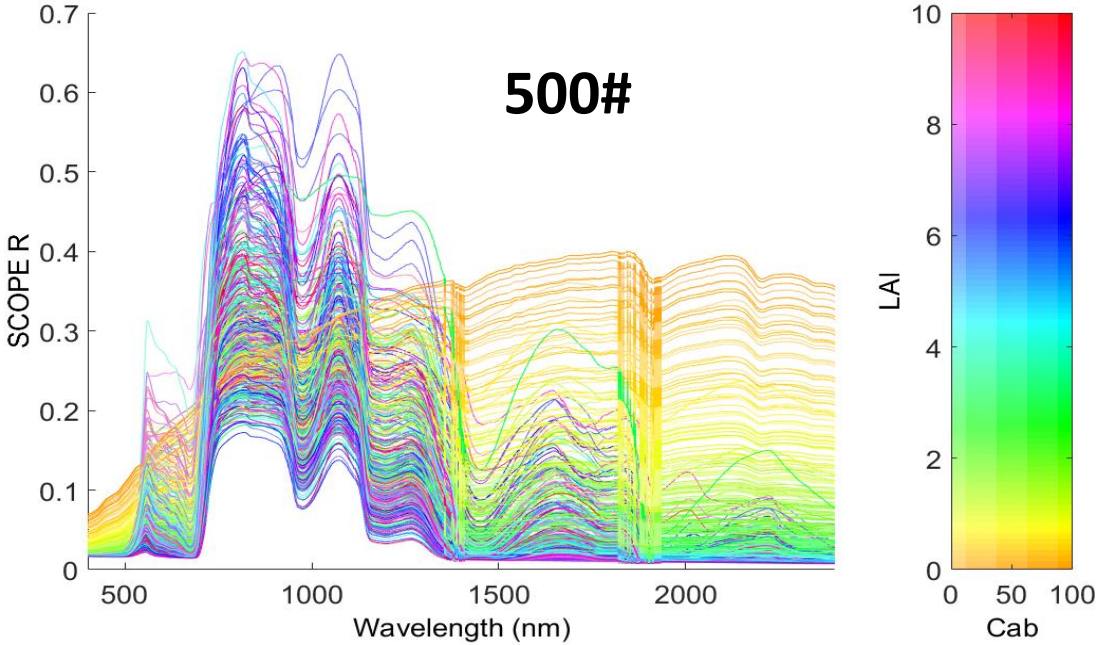




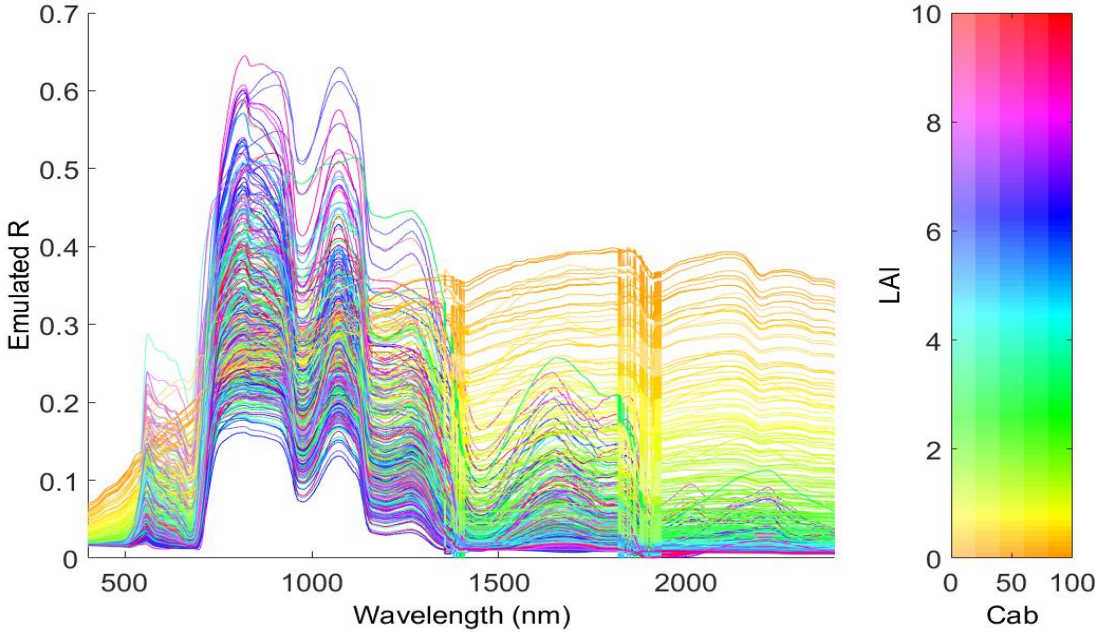
Any difference? Which model would you choose?

37 min

0.2 s



SCOPE



Emulator

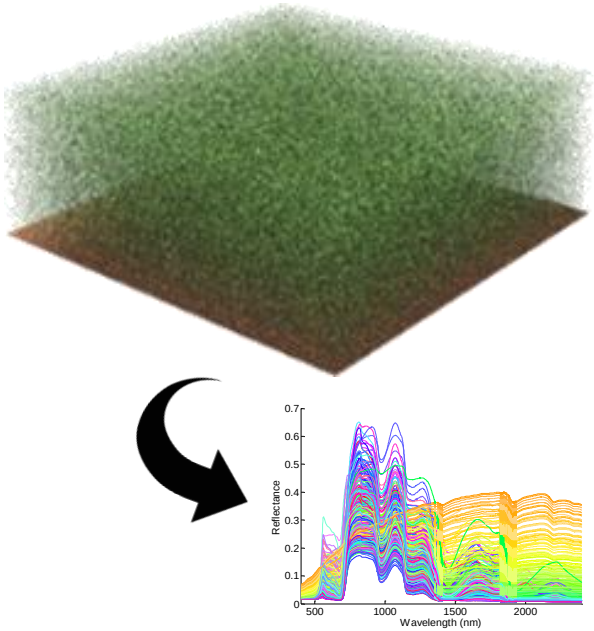
(emulated SCOPE)

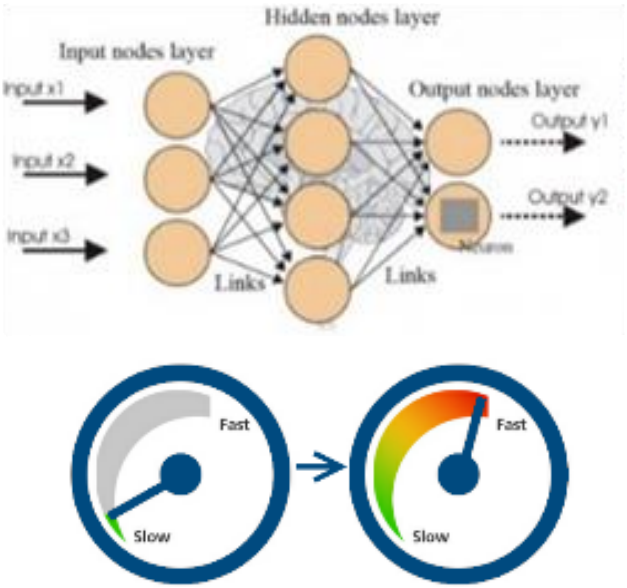
Emulation of RTMs

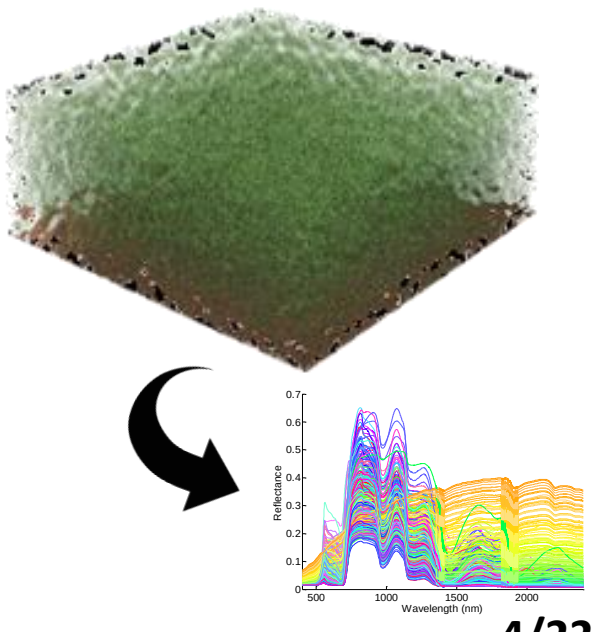
Emulators are statistical models that approximate the processing (input-output) of a physical model (e.g. RTM) - at a fraction of the computational cost:

making a statistical model from a physical model

RTM | **Machine learning** | **Emulator**

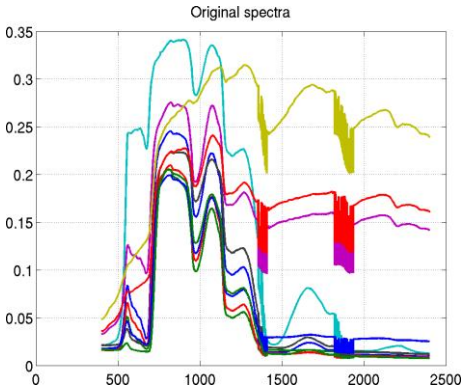




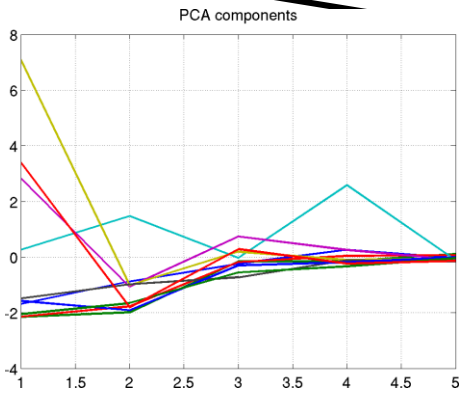


Processing steps emulation

PCA on spectra



$$Sc = U \cdot X$$

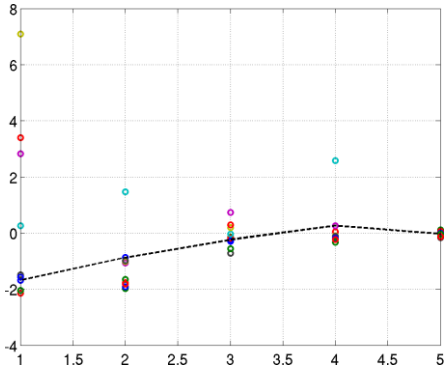


MLRA training looping over components

$$W = (Y + \lambda I)^{-1} \cdot Sc$$

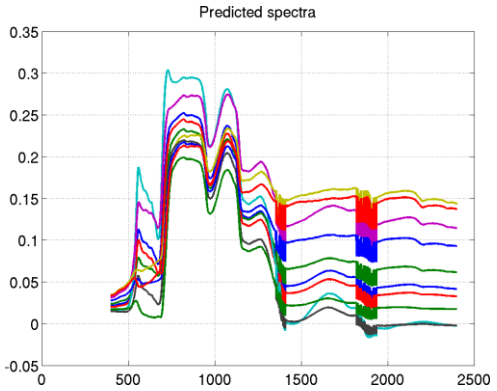
Prediction of components

$$Sp = Sc \cdot W$$



Reconstruction of spectra

$$Xr = U^{T} \cdot Sp$$

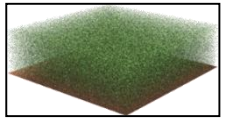


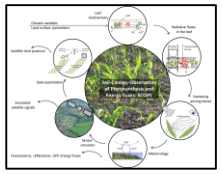
Emulator toolbox

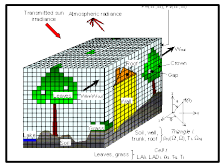
http://ipl.uv.es/artmo/

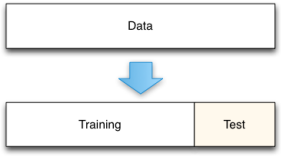
With ARTMO's emulation processing chain any RTM can be converted into an emulator.

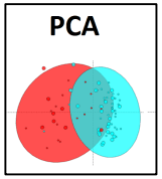
Input (variables + spectra) → Splitting into training/validation → PCA on spectra → MLRA training looping over components → Prediction of components → Reconstruction of full spectrum → Validation → Emulator

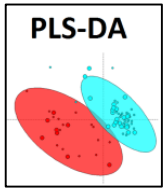


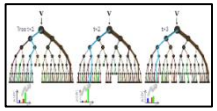


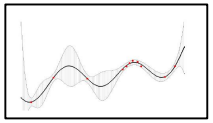


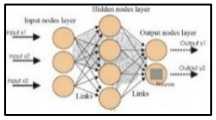


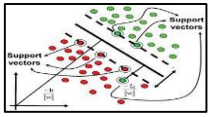


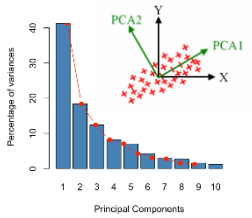








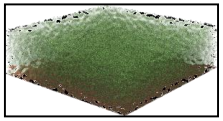


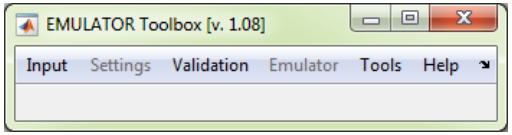


$$Xr = U^{T} \cdot Sp$$

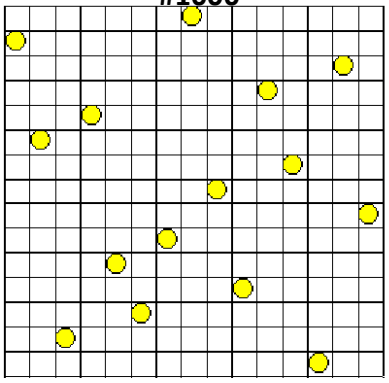
$$RMSE = \sqrt{\sum \frac{(y_{pred} - y_{ref})^2}{N}}$$

$$NRMSE = \frac{1}{Y_{max} - Y_{min}} \sqrt{\frac{\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}{N}},$$





Latin Hypercube Sampling (LHS) #1000



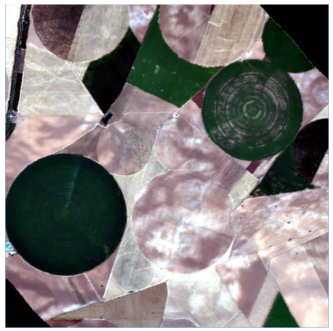
Input → Settings → Validation → Emulator

- **Input**: RTM data, Txt data → Select project, Edit settings
- **Settings**
- **Validation**: New, Load
- **Emulator**: RTM vs Emulator, LUT Emulator, Txt Emulator, Scene Emulator (NEW!)
- **Tools**: Save, Load, Manage tests, Options, View figure, Plot LUT, Residual analyzer, Statistical analyzer, Import, Scene comparison → Delete, Rename
- **Help**: Show Log, User's manual, Installation guide, Disclaimer

Rivera, J.P., Verrelst, J., Gómez-Dans, J., Muñoz-Marí, J., Moreno, J., Camps-Valls, G. (2015). An Emulator Toolbox to Approximate Radiative Transfer Models with Statistical Learning. Remote Sensing. p. 7, 9347-9370.

Emulation of experimental spectral data



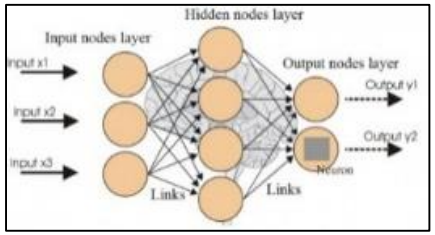
Can we use emulation to predict <u>noisy data</u> such as:

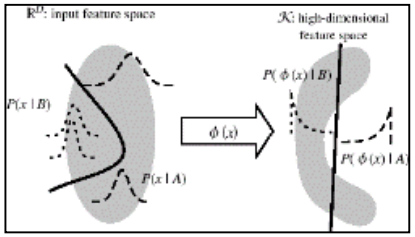
1. ***Field measurements?***
2. ***Images?***

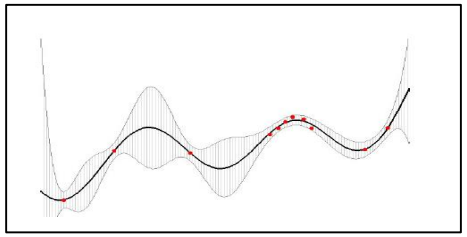
Various open questions:

Emulators great idea… what about accuracy?

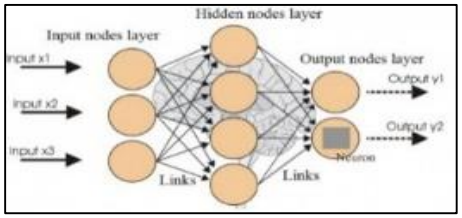
1. Role of machine learning regression algorithm?



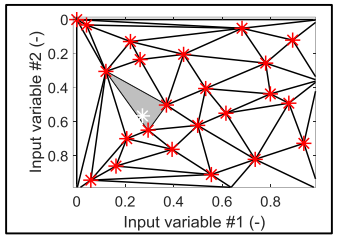




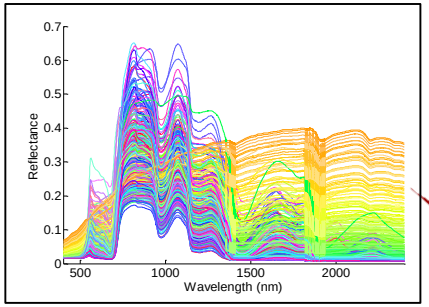
2. Emulation is same as interpolation?



vs



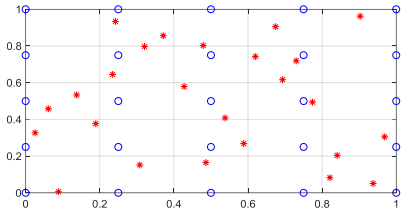
3. Role of data type?



vs.

deterministic versus stochastic?

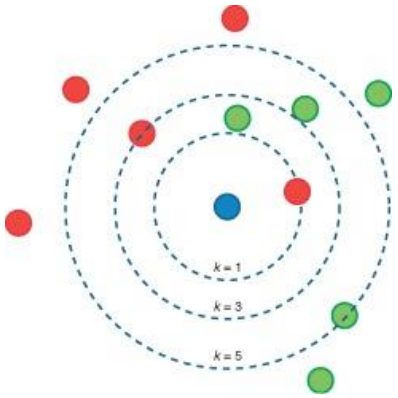
Interpolation of experimental data: Scattered methods



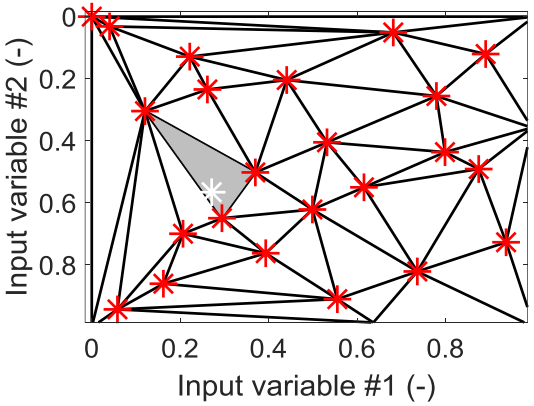
Scattered (*) vs
Gridded(o) LUTs

- Value at query point depends on distance to LUT nodes
- No tuning hyper-parameters
- Multi-output: spectra (***K***-dim)
- Only few interpolation methods allow interpolating in scattered ***D***-dim parameter space:

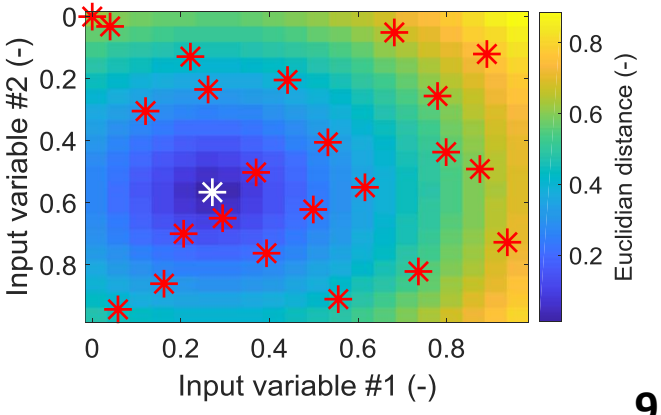
Nearest neighbour



Linear interpolation

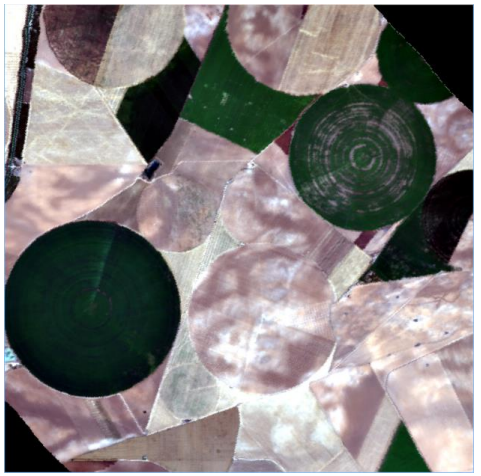


Inverse Distance Weighting (IDW)



Rationale & setup experimental SPARC dataset

Can we use emulation to predict experimental spectroscopic data?



SPARC data set (July 2003; **Barrax**, Spain)

- **Field data (135 samples), 6 variables:**
  - **Leaf Chl** measured with CCM-200
  - **LAI** measured with LiCor LAI-2000
  - **FVC** measured with hemispherical photographs
  - **Biomass**
  - **Leaf water content**
  - **Canopy water content**
- **Spectral data**:
  - **HyMap (125 bands)**

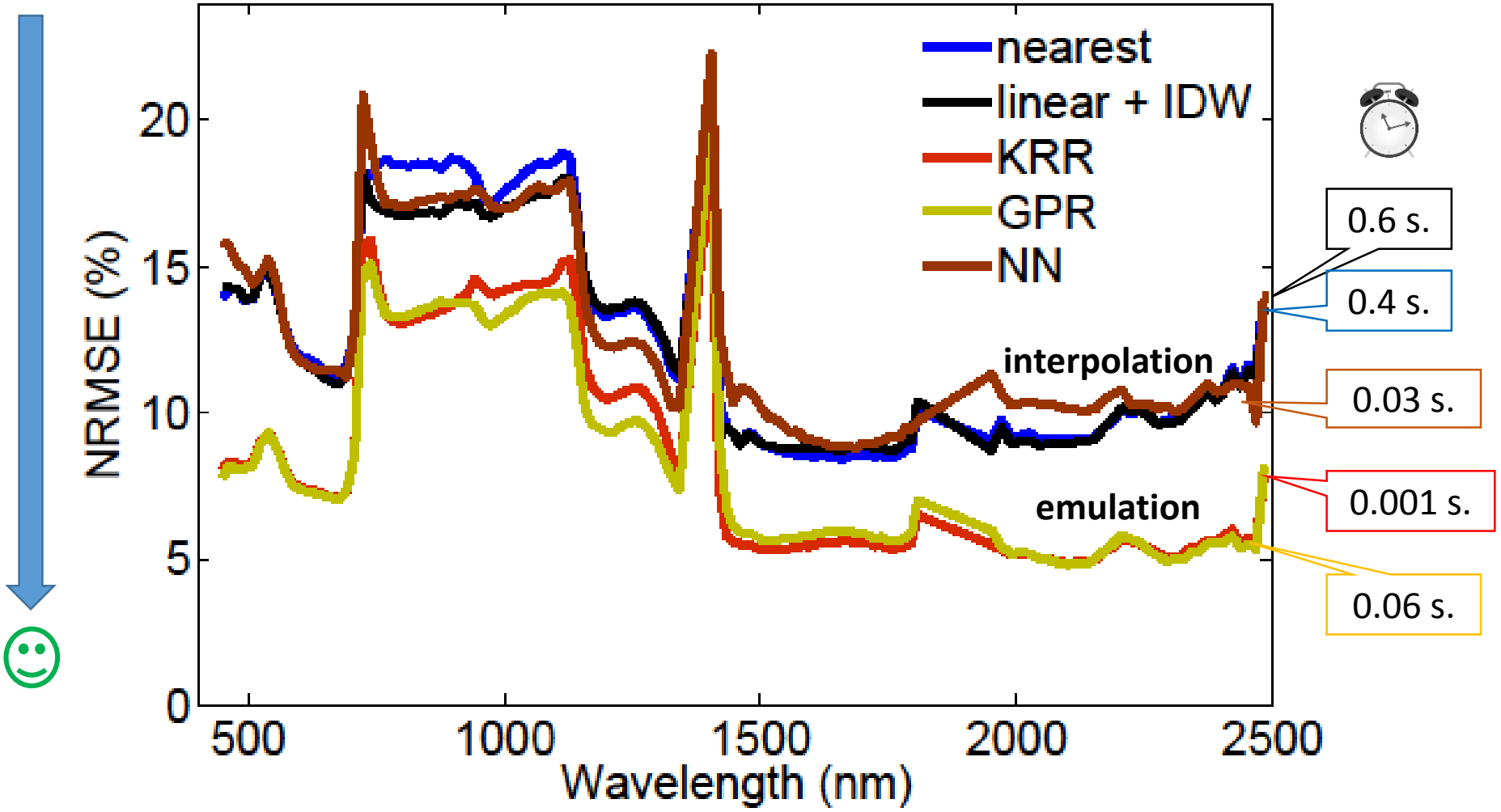
80% Training

(MLs: 80% training – 20% testing)

20 PCA

20% Validation

NRMSE (%) results interpolation and emulation validated against remaining 20% SPARC data.



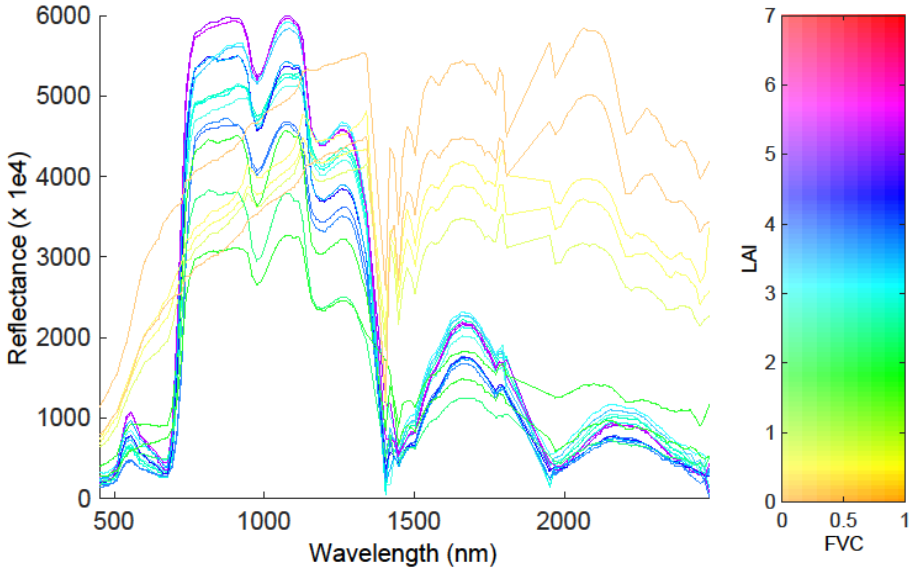
Emulation methods more accurate and faster than conventional interpolation methods.

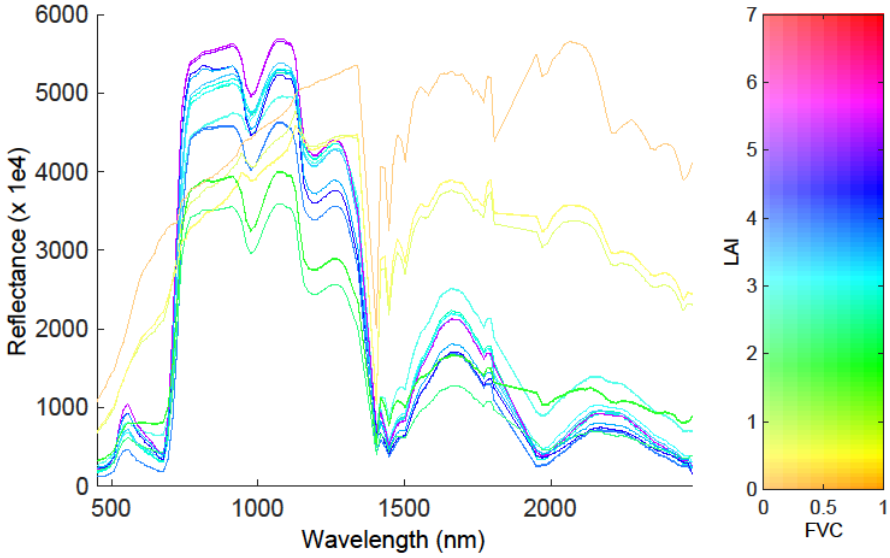
Visual comparison reference data vs. emulated data GPR emulator

20% SPARC dataset

GPR emulator

0.02 seconds





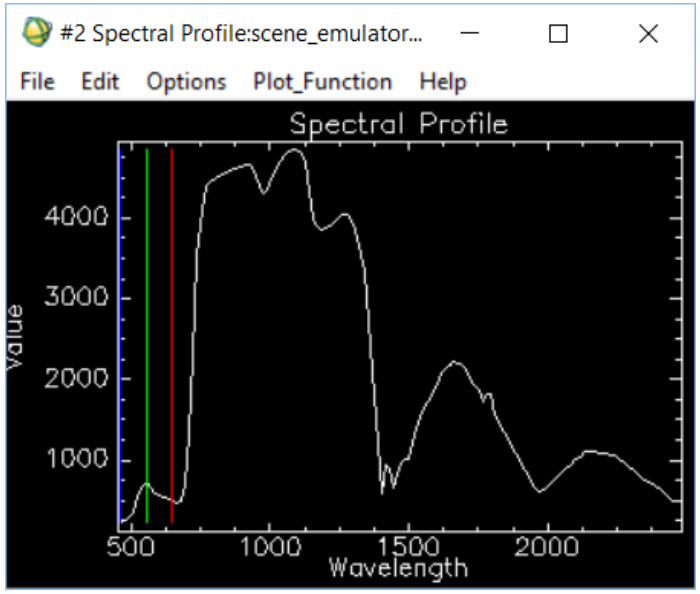
- **Somewhat less variation emulator, because bare soils (variables=0) produce only 1 output spectra.**

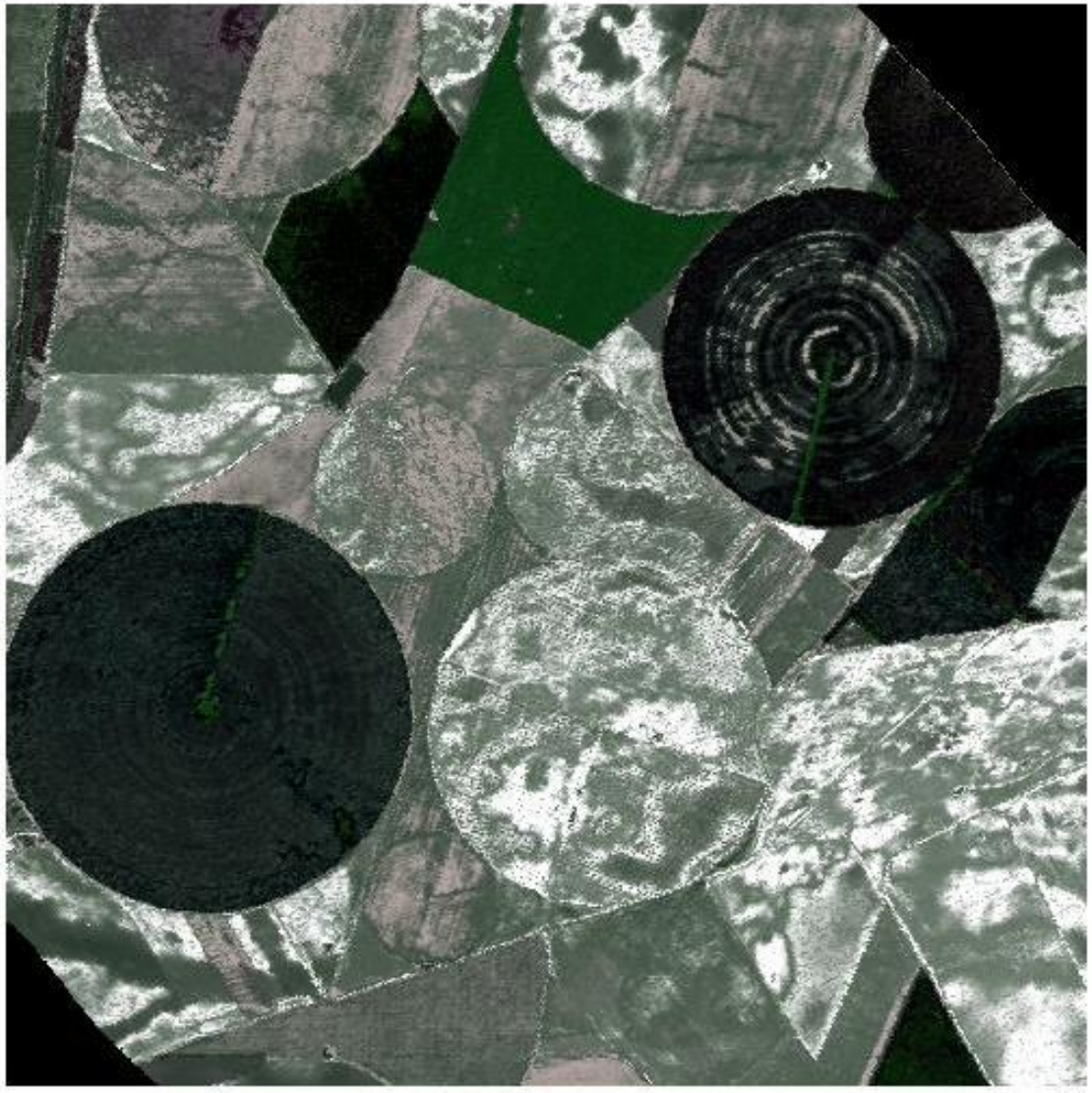
- **Inclusion of a soil variable can solve this issue, e.g. soil moisture**

Example of #500 emulated SPARC spectra based on varying all 6 variables

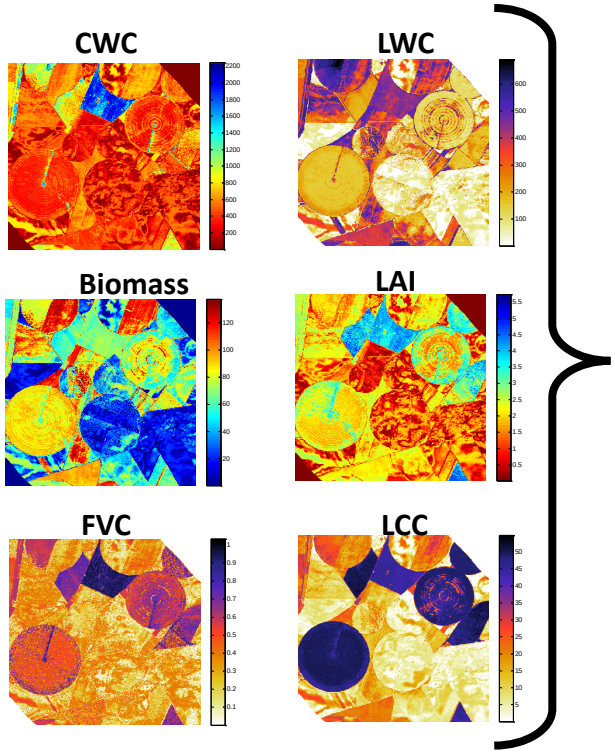
0.1 seconds

Hyperspectral subset: image?



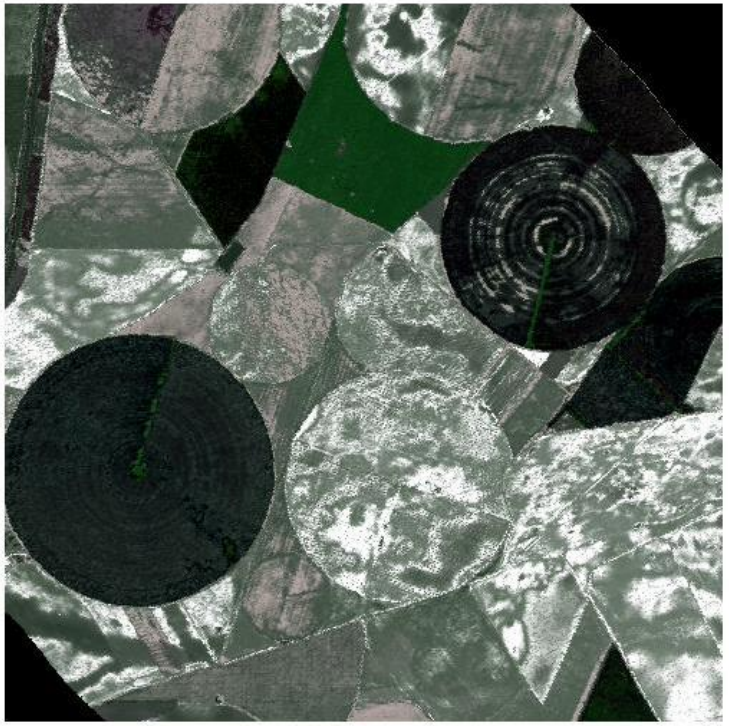


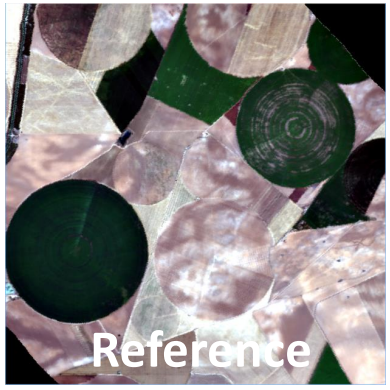
Using emulator to reconstruct hyperspectral image (125 bands, 500 x 500)



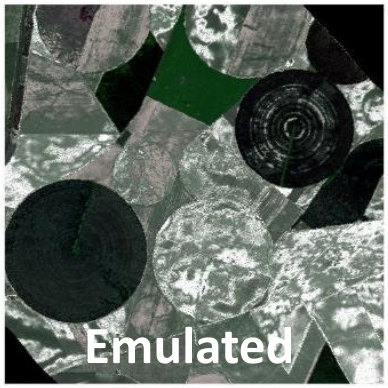
Emulator

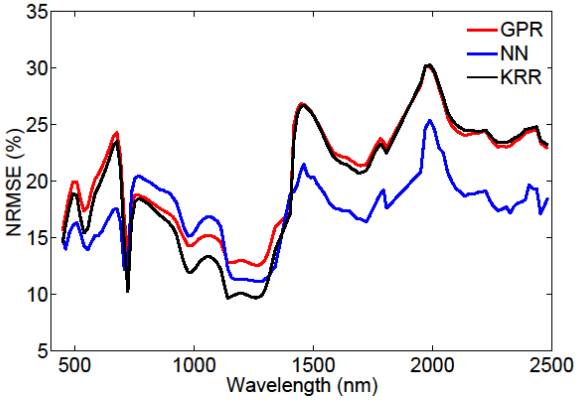
2 seconds





NRMSE (%)

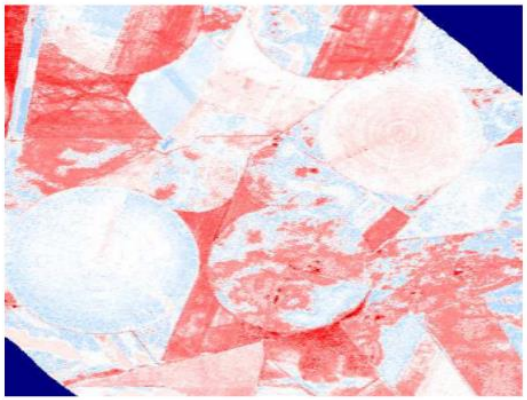




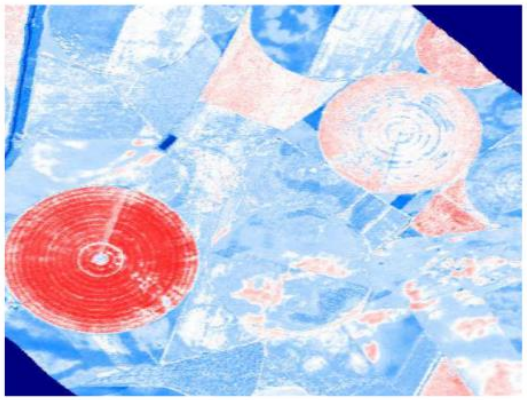
Evaluation NN emulator

(125 bands, 500 x 500)

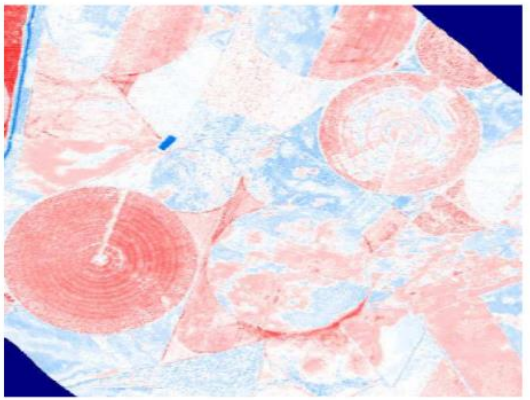
570 nm



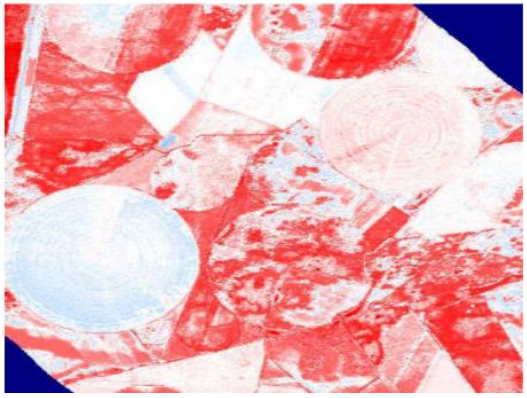
738 nm



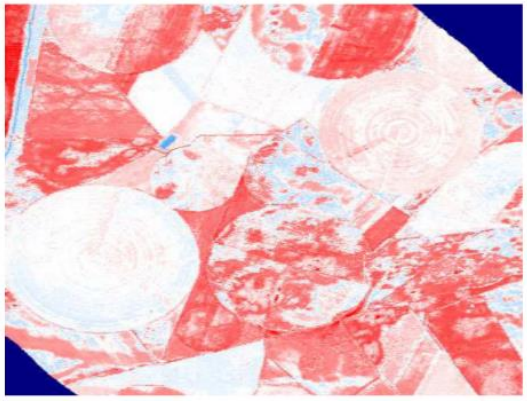
1257 nm



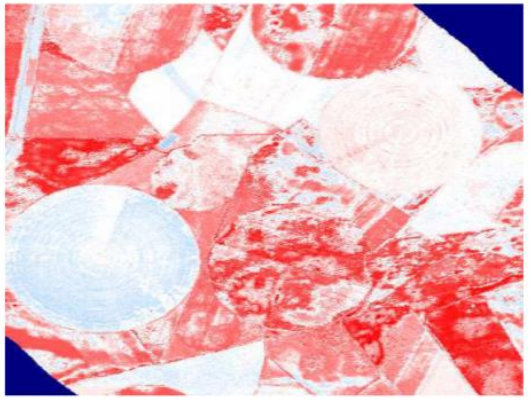
1503 nm

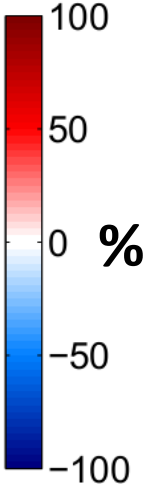


1723 nm



2154 nm





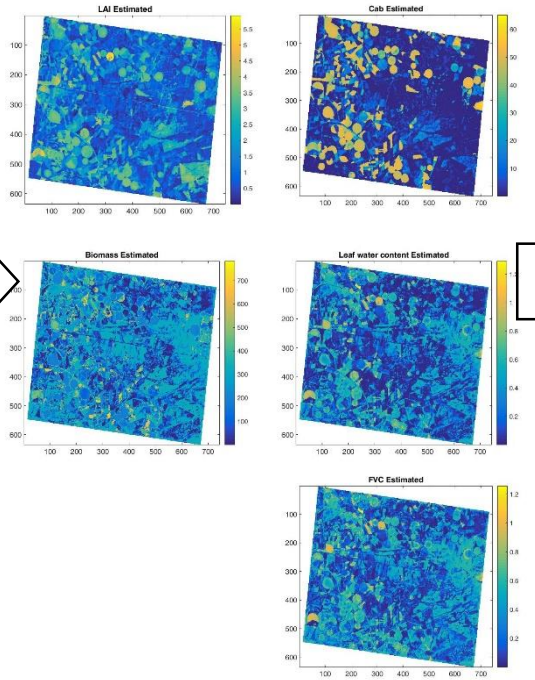
Some areas perfectly emulated, however, also significant differences: soil spectral variability poorly emulated (because not trained for it).

Using emulator to reconstruct a complete CHRIS image

CHRIS image (62 bands), Barrax, Spain

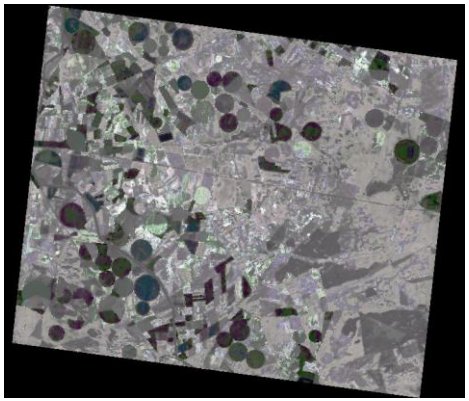
Maps

Maps of biophysical variables

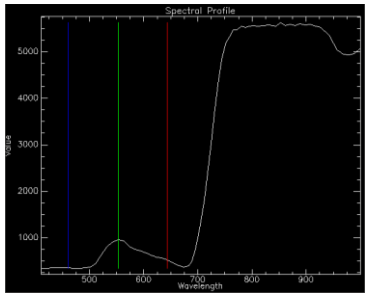


Emulator

Original CHRIS image (62 bands)



22 seconds

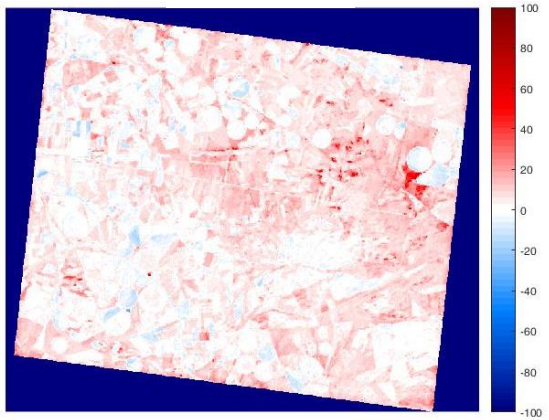


- **Fast rendering of full hyperspectral image.**
- **Vegetated surfaces adequatly emulated.**
- **A variable for controlling spectral variability bare soil needed (e.g. soil moisture)**

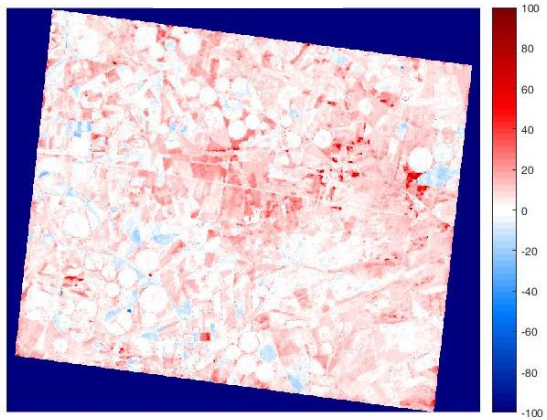
Size of GPR model: 0.51Mb

Relative error maps

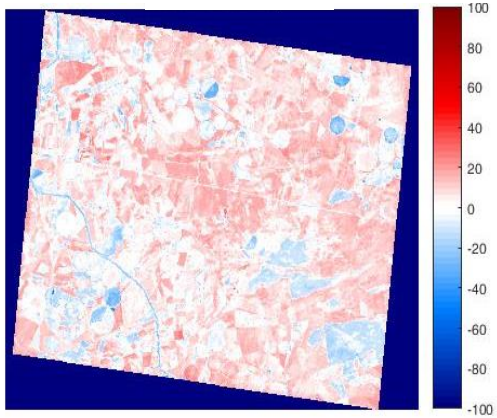
553 nm

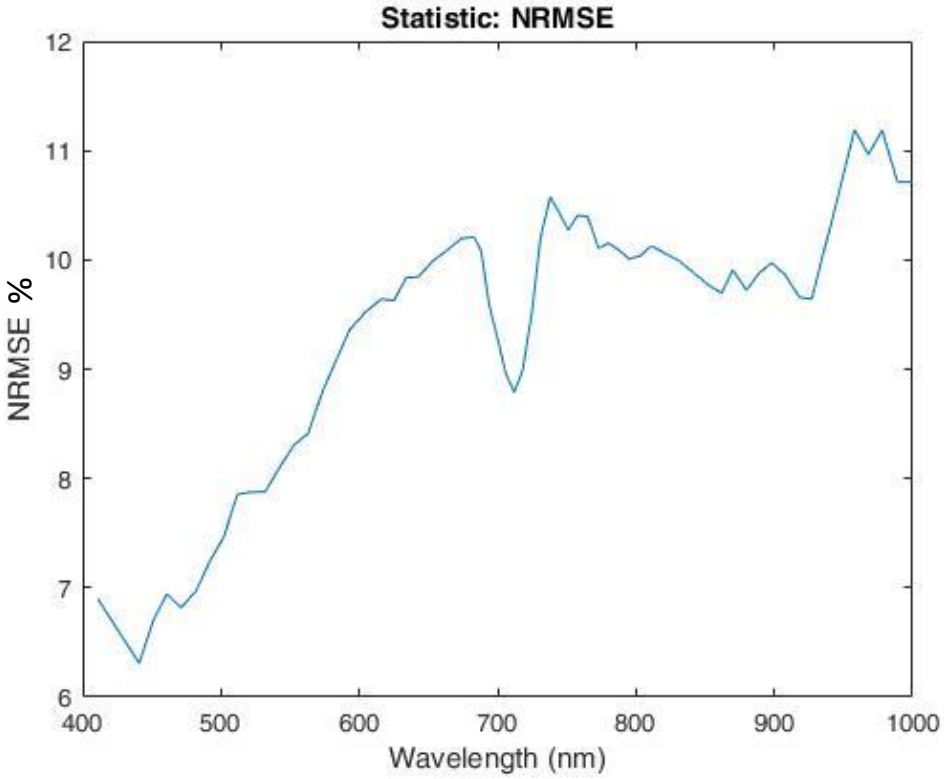


643 nm



844 nm

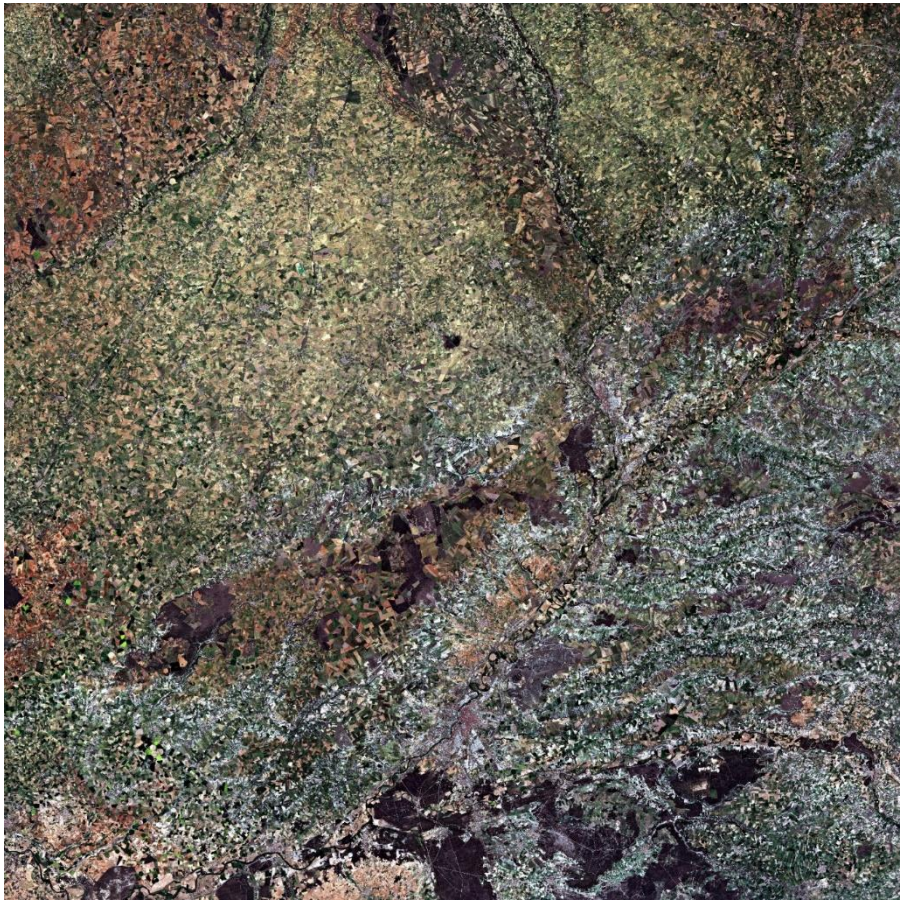




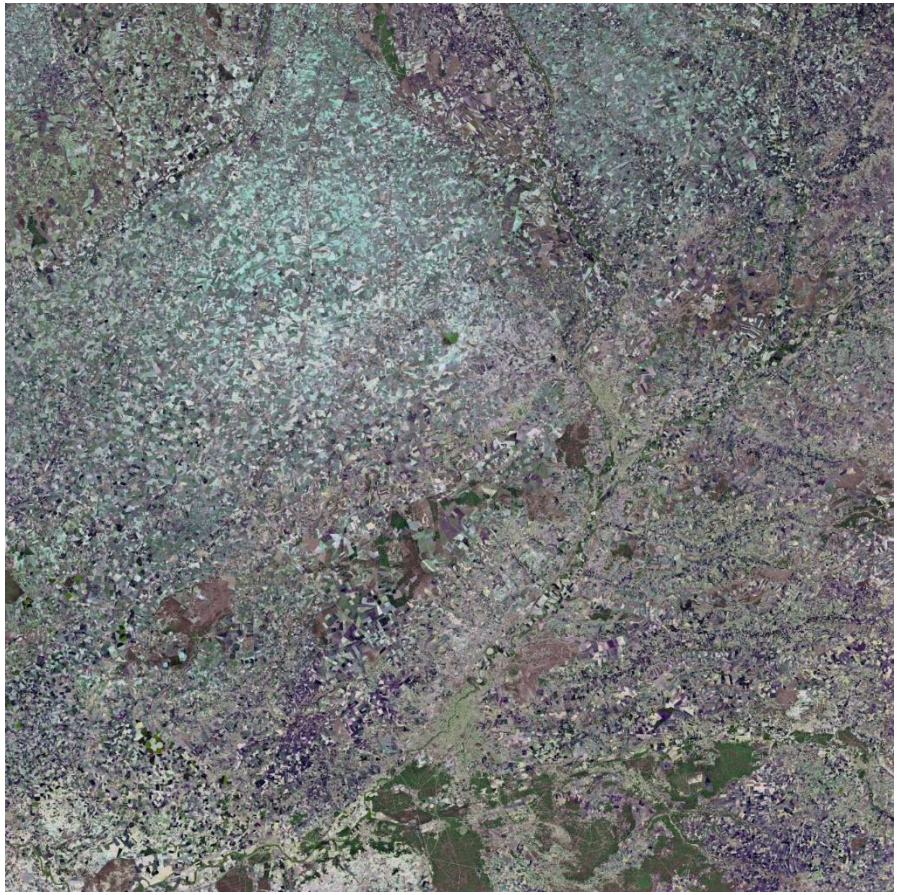
Visible part better emulated than NIR.

Emulation of a S2 subset

L2A S2 subset
(10 bands), Valladolid, Spain



GPR reconstructed S2 subset

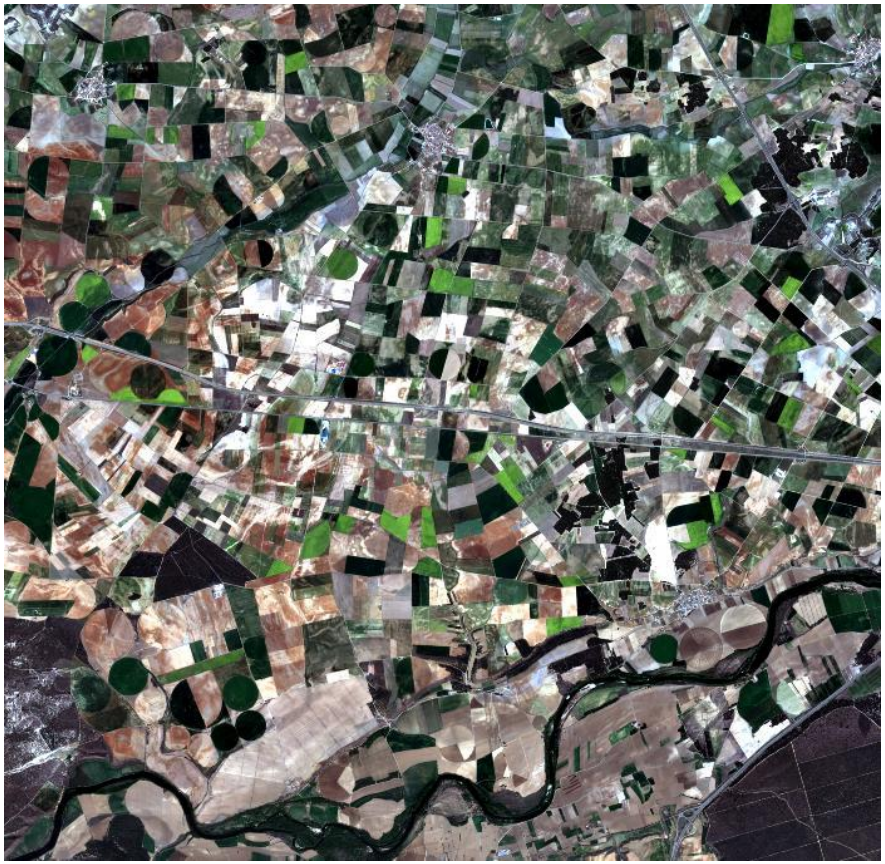


313 seconds

Size of GPR model: 1.8Mb

Emulation of a smaller S2 subset

L2A S2 subset
(10 bands), Valladolid, Spain

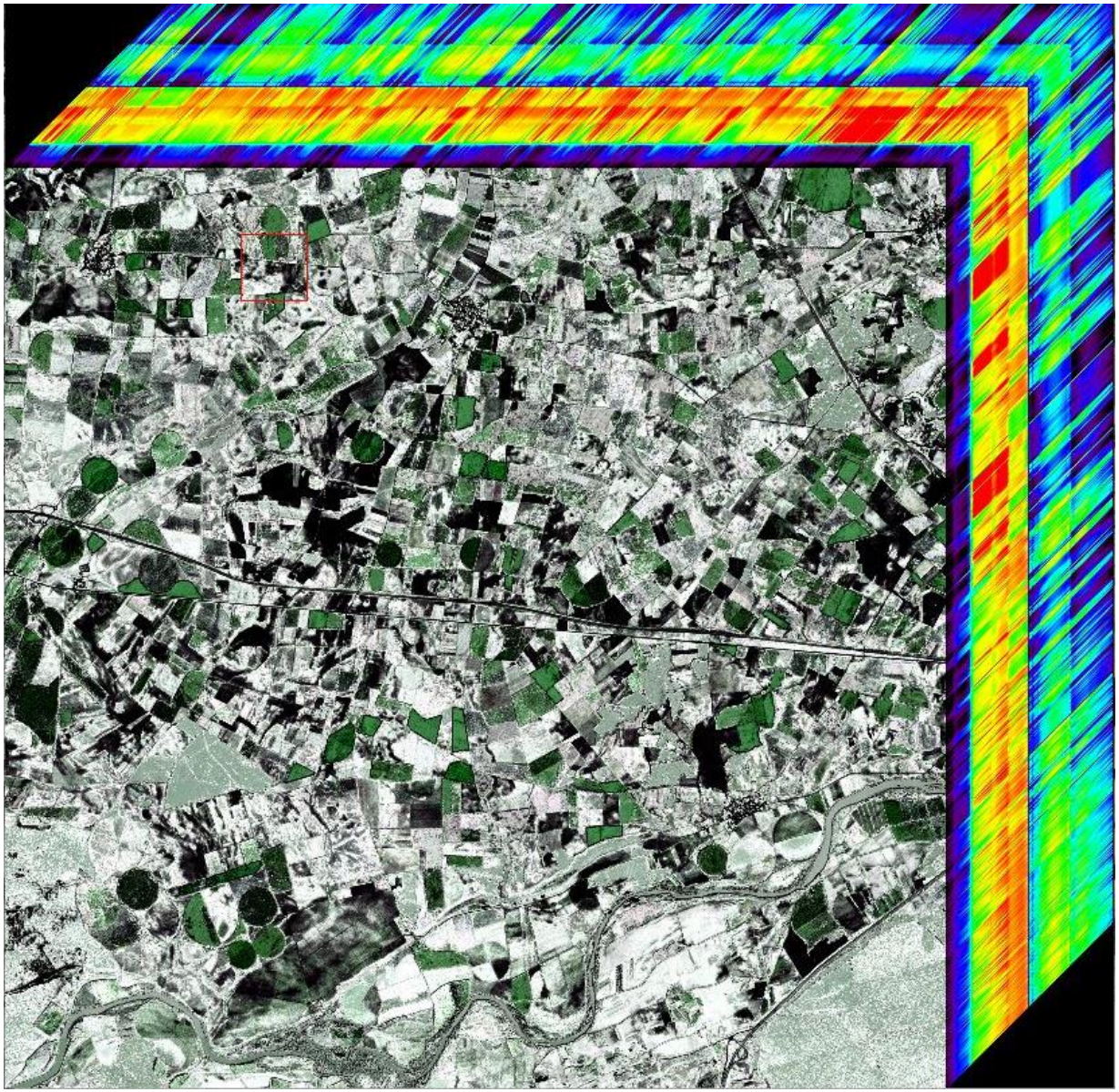


GPR reconstructed S2 subset

23 seconds

Size of GPR model: 0.12Mb

Emulation of S2-like hyperspectral image

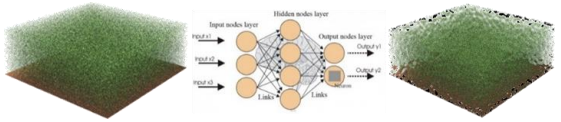


125 bands

18 seconds

Size of GPR model: 0.4Mb

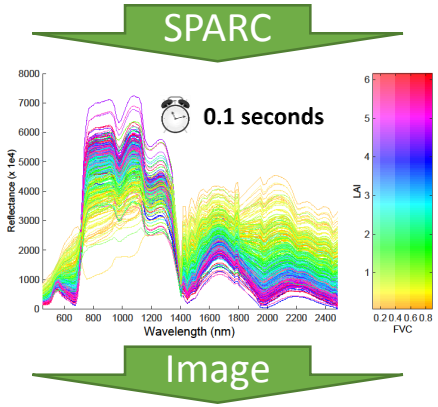
Take home messages

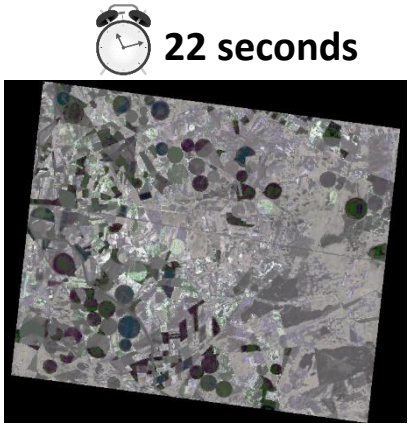


- ✓**Emulation can** be used to rapidly **reconstruct sensor-like (hyper)spectral data with sufficient accuracy**.

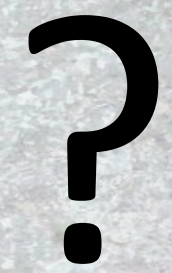
- ✓**Emulation is more accurate and faster than** conventional **interpolation techniques**.

- ✓**Emulation can generate simplified (hyper)spectral scenes** in the order of seconds.





Thanks!



More about emulation:

Daniel Heestermans: WE2.R7.5 **MULTIOUTPUT AUTOMATIC EMULATOR FOR RADIATIVE TRANSFER MODELS** (12:30-12:50)