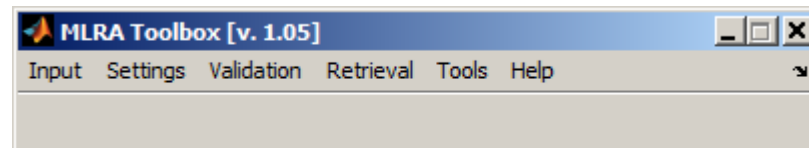


Machine Learning Regression Algorithms (MLRA) toolbox - v.1.05

Tutorial: Training/Validating & Mapping with User data



Jochem Verrelst

Jochem.verrelst@uv.es - <http://ipl.uv.es/artmo/>

8-12 September 2014 – University of Valencia – Valencia, Spain



ARTMO v3.03

ARTMO [v. 3.03]

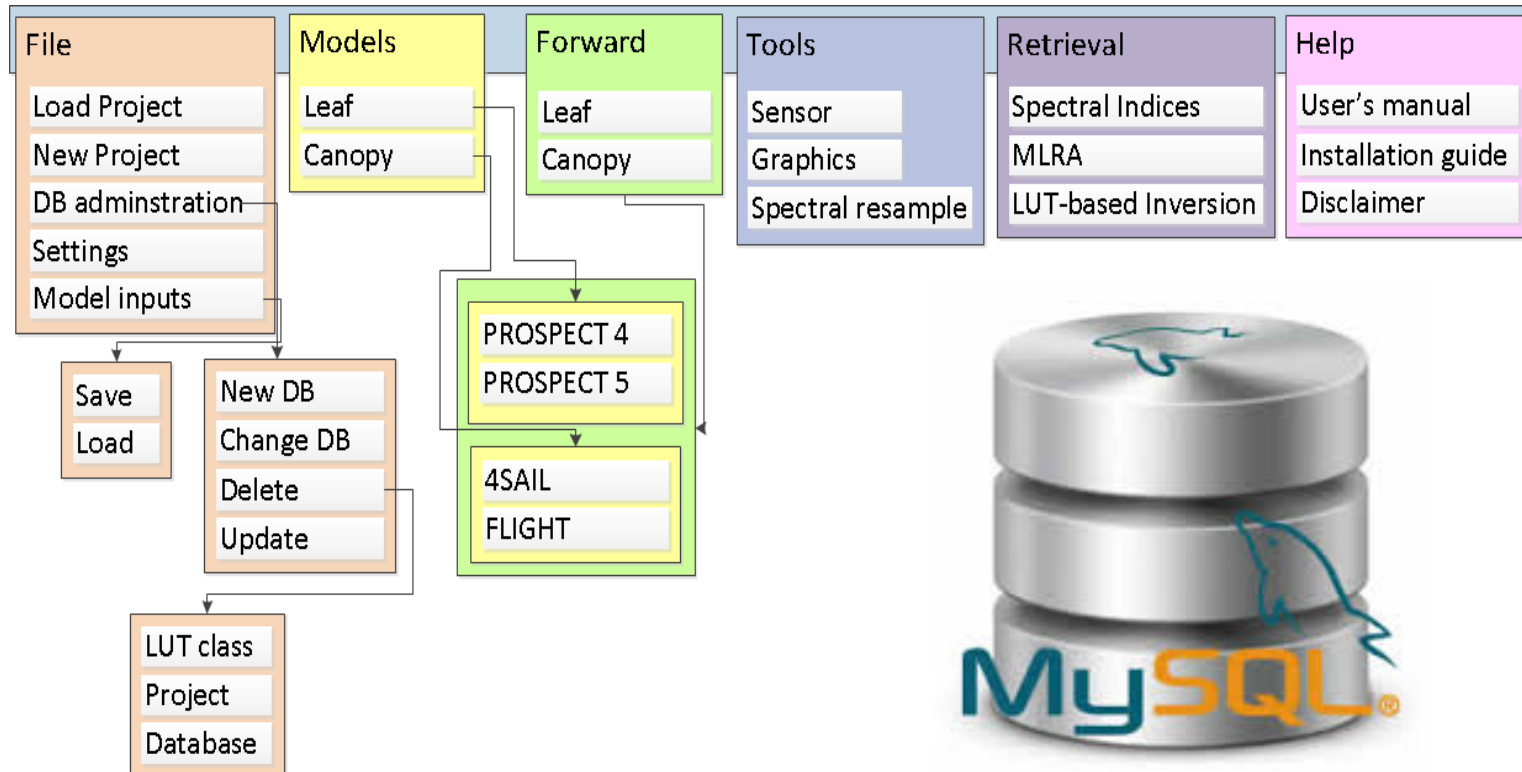
File Models Forward Tools Help

Project Description

Project Name:

Comment:

Sensor:



Website: <http://ipl.uv.es/artmo/>

The screenshot shows the ARTMO website interface in a web browser. The browser's address bar displays `ipl.uv.es/artmo/`. The website features a navigation menu with links: Home, Radiative Transfer Models, Retrieval Toolboxes, Tools, Publications, Join Us!, and Contacts. A search bar and 'Log in'/'Register' buttons are located in the top right. The main content area welcomes visitors and provides information about the software, including a registration notice and a description of its capabilities. An inset image shows the ARTMO v. 3.0.3 main window, which includes a menu bar (File, Models, Forward, Tools, Help) and several functional panels: Project Description, Models (Leaf, Canopy), Forward (Leaf, Canopy), Tools (Sensor, Graphics, Spectral resample), Retrieval (Spectral indices, MLRA, LUT-based Inversion), and Help (User's manual, Installation guide, Disclaimer). A flowchart at the bottom illustrates the software's architecture, showing the relationships between these components and the LUT class.

Home

ARTMO

Search...

Log in Register

Home Radiative Transfer Models Retrieval Toolboxes Tools Publications Join Us! Contacts

Welcome to the ARTMO website!

— Please register to freely download ARTMO. By registering you agree with the License agreement. We will verify whether you are a non-commercial user and activate your account. If successful you will be informed by email. The Download page will appear after logging in. —

Plant Radiative Transfer models (RTMs) have become important tools for the analysis of optical Earth observation data, providing meaningful links between radiometry and environmental applications, such as ecological processes, environment and precision agriculture. However, for the broader community these models are still often perceived as excessively complicated tools and are not always easily accessible.

The in-house developed **Automated Radiative Transfer Models Operator (ARTMO)** Graphic User Interface (GUI) is a software package that provides essential tools for running and inverting a suite of plant RTMs, both at the leaf and at the canopy level. **ARTMO** facilitates consistent and intuitive user interaction, thereby streamlining model setup, running, storing and spectra output plotting for any kind of optical sensor operating in the visible, near-infrared and shortwave infrared range (400-2500 nm).

ARTMO's main window.

Principally, **ARTMO** allows:

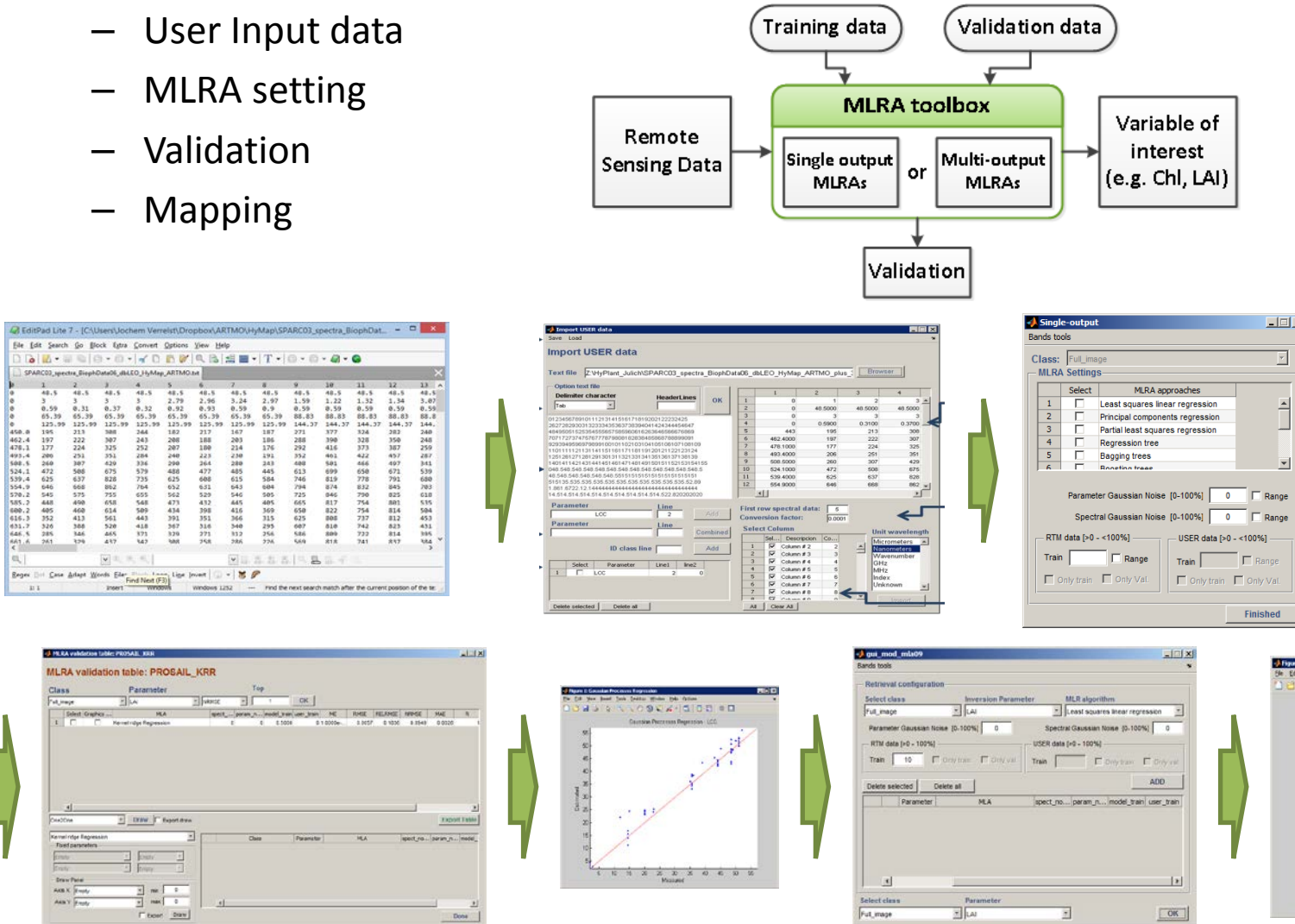
1. To configure and run leaf and canopy RTMs, independently or combined, in an intuitive way through various GUIs with input options to insert single values, value ranges, or imported external datasets.
2. To simulate and store a massive quantity of spectral output based on a look-up table (LUT) approach in a relational database.
3. To plot groups of simulated spectra in the same plotting window with color gradients as a function of input parameters.
4. To export simulated spectra and associated meta-data to a text file for further processing.
5. To analyze and apply retrieval techniques in order generate maps of biophysical parameters from optical remote sensing imagery.

```
graph TD
    subgraph File
        LoadProject[Load Project]
        NewProject[New Project]
        DBAdmin[DB administration]
        Settings[Settings]
        ModelInputs[Model inputs]
        Save[Save]
        LoadDB[Load]
        NewDB[New DB]
        ChangeDB[Change DB]
        Delete[Delete]
        Update[Update]
    end
    subgraph Models
        Leaf[Leaf]
        Canopy[Canopy]
    end
    subgraph Forward
        LeafF[Leaf]
        CanopyF[Canopy]
    end
    subgraph Tools
        Sensor[Sensor]
        Graphics[Graphics]
        SpectralResample[Spectral resample]
    end
    subgraph Retrieval
        SpectralIndices[Spectral indices]
        MLRA[MLRA]
        LUTBasedInversion[LUT-based Inversion]
    end
    subgraph Help
        UsersManual[User's manual]
        InstallationGuide[Installation guide]
        Disclaimer[Disclaimer]
    end
    subgraph LUTClass [LUT class]
        PROSPECT4[PROSPECT 4]
        PROSPECT5[PROSPECT 5]
        DLM[DLM]
        4SAIL[4SAIL]
        FLIGHT[FLIGHT]
    end
    File --> Models
    Models --> Forward
    Forward --> Tools
    Tools --> Retrieval
    Retrieval --> Help
    Tools --> LUTClass
    Models --> LUTClass
    Forward --> LUTClass
```

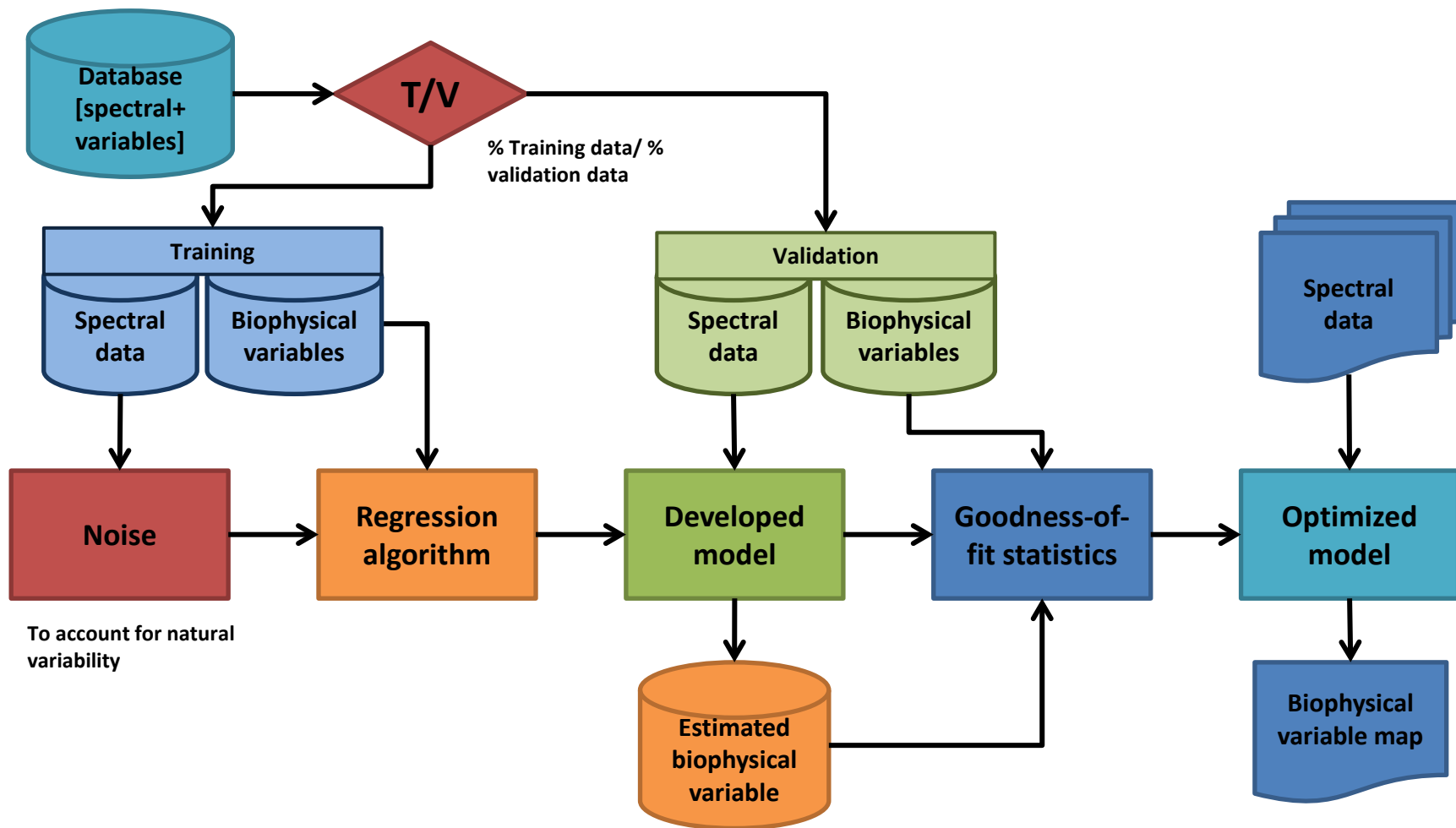
Outlook:

1. MLRA mapping based on User data

- User Input data
- MLRA setting
- Validation
- Mapping



Schematic overview for systematic evaluation of nonparametric regression models to estimate biophysical variables

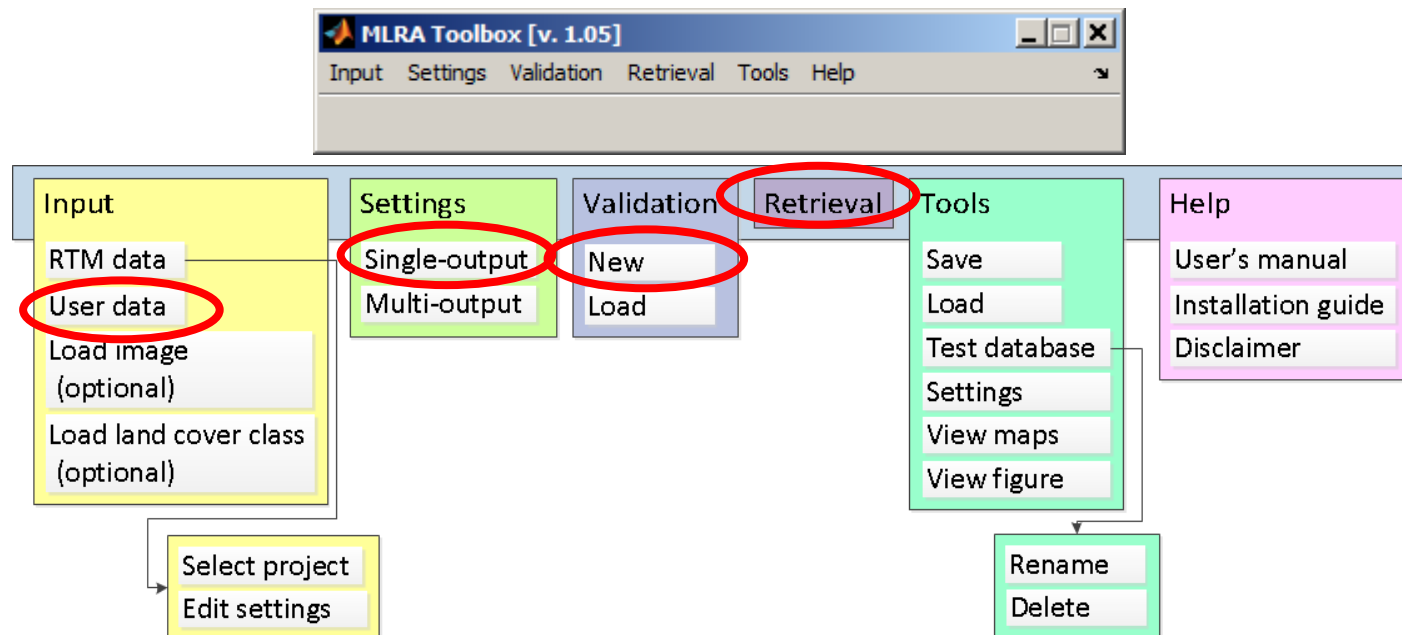


MLRA mapping based on User data

The **example** will demonstrate how to train and validate single-output MLRA models and apply it to an image using a field dataset for training. Consult the **Manual** for more details.

The procedure will be as follows:

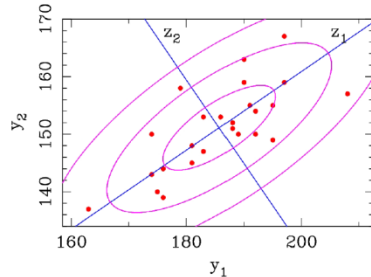
1. **User data:** Insert field data for training and validation
2. **Single-output:** Choose single-output MLRA models and define training/testing partitioning
3. **Validation:** Validate the defined MLRA strategies
4. **Retrieval:** Apply the best one to a remote sensing image.



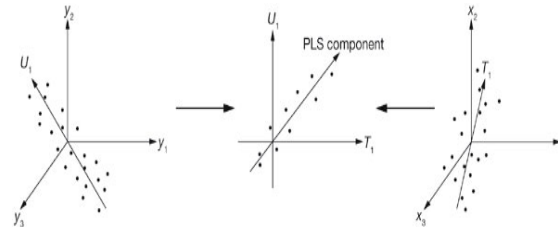
• Non-parametric models

- **SimpleR** [Camps-Valls et al., 2013]
- <http://www.uv.es/gcamps/code/simpleR.html>

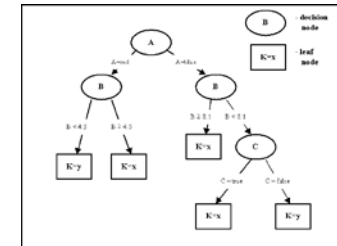
Principal component regression– PCR



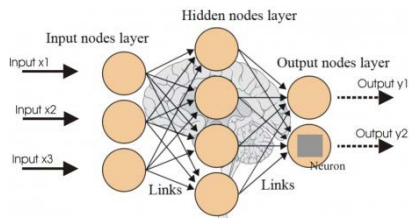
Partial least squares regression– PLSR



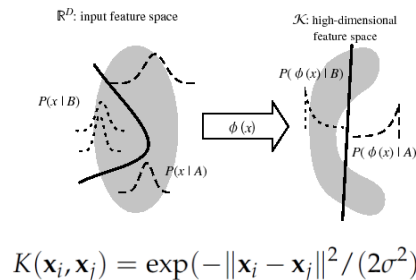
Decision Trees – DT



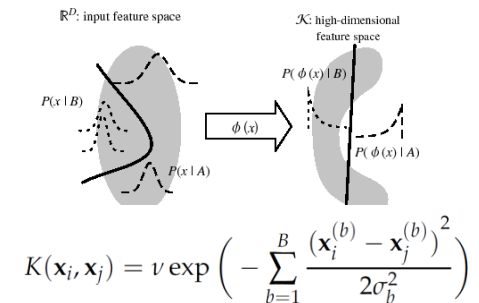
Neural networks – NN



Kernel ridge regression – KRR



Gaussian processes regression - GPR



Also:

- Elastic Net (ELASTICNET)
- Bagging trees (BAGTREE)
- Boosting trees (BOOST)
- Neural networks (NN)

- Extreme Learning Machines (ELM)
- Support Vector Regression (SVR)
- Relevance Vector Machine (RVM)
- Variational Heteroscedastic Gaussian Process Regression (VHGPR)

Input: User data (e.g. field data)

User data for training and validation requires one input file, including:

1. Biophysical parameters (e.g., LAI, chlorophyll content,...)
2. Associated spectra (e.g., obtained from a remote sensing image)

User data need to be organized in a matrix format in plain text file, according to example below:

Wavelengths

	1	2	3	4	5	6	7	8	9	10	11	12	13
0	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5
0	3	3	3	3	2.79	2.96	3.24	2.97	1.59	1.22	1.32	1.34	3.07
0	0.59	0.31	0.37	0.32	0.92	0.93	0.59	0.9	0.59	0.59	0.59	0.59	0.59
0	65.39	65.39	65.39	65.39	65.39	65.39	65.39	65.39	88.83	88.83	88.83	88.83	88.8
0	125.99	125.99	125.99	125.99	125.99	125.99	125.99	125.99	144.37	144.37	144.37	144.37	144
450.0	195	213	308	244	182	217	167	187	271	377	324	282	240
462.4	197	222	307	243	208	188	203	186	288	390	328	350	248
478.1	177	224	325	252	207	180	214	176	292	416	373	387	259
493.4	206	251	351	284	240	223	230	191	352	461	422	457	287
508.5	260	307	429	336	290	264	280	243	408	501	466	497	341
524.1	472	508	675	579	488	477	485	445	613	699	650	671	539
539.4	625	637	828	735	625	608	615	584	746	819	778	791	680
554.9	646	668	862	764	652	631	643	604	794	874	832	845	703
570.2	545	575	755	655	562	529	546	505	725	846	790	825	618
585.2	448	490	658	548	473	432	445	405	665	817	754	801	535
600.2	405	460	614	509	434	398	416	369	650	822	754	814	504
616.3	352	413	561	443	391	351	366	315	625	808	737	812	453
631.7	326	388	520	418	367	316	340	295	607	810	742	823	431
646.5	285	346	465	371	329	271	312	256	586	809	722	814	395
661.6	261	329	437	342	308	258	286	226	569	818	741	837	384

Input parameters

Associated spectra

- Make sure to fill up the whole Matrix! In case of empty cells, use **NaN** and remove those samples in the following step.
- Make sure that wavelengths are the same as the remote sensing image! They need to match. A band selection or band transformation can be later done in **Settings**.

Input

- **RSM data**
- **User data**
- Load image (optional)
- Load land cover class (optional)

[illegible]

A sample of the input text file is visualized.

Chosen input parameter and corresponding column. Parameters can be combined (product)

*Inserted input
parameters*

Selected
input data:
parameters
on top (rows)
and spectra
below
(columns).

Starting line
spectra. Convert
units if needed.

Option to
remove
samples.

1. **Browser:** Import User data file.
2. Inspect if right data in left panel. By clicking on **OK** data will appear in right panel.
3. Define a row with a parameter to its line. Click on **Add**. Multiple parameters can be define by repeating this step. Parameters can be combined.
4. **Define the row where spectra starts.**
5. If needed, **convert** spectral data.
6. Option to **remove** samples.
7. Configured input data can be **saved** and **loaded** as .m file.
8. Finally, click on **Import**.

Single-output Settings

Option to apply a band subselection or transformation

If a land cover map has been provided, per class can be configured.

Multiple regressors can be selected

If RTM data is inserted it can serve for training or validation

Single-output

Bands tools

Class: Full_image

MLRA Settings

Select	MLRA approaches
<input type="checkbox"/>	1 Least squares linear regression
<input checked="" type="checkbox"/>	2 Principal components regression
<input type="checkbox"/>	3 Partial least squares regression
<input type="checkbox"/>	4 Regression tree
<input type="checkbox"/>	5 Bagging trees
<input type="checkbox"/>	6 Boosting trees

Parameter Gaussian Noise [0-100%] 0 ☐ Range

Spectral Gaussian Noise [0-100%] 0 ☐ Range

RTM data [≥0 - <100%]
Train ☐ Range
☐ Only train ☐ Only Val.

USER data [≥0 - <100%]
Train ☐ Range
☐ Only train ☐ Only Val.

Finished

Settings
Single-output
Multi-output

gui_mod_mla07

Parameter noise - Range:[0-100]

Min: 0 Max: 20 Step: 2

OK

Range options:

- Step
- Distribution

Options to add noise to parameters and spectral data

Options to control the training/validation partitioning for user data

1. Select the **nonparametric regression models** to be trained and validated.
2. The option to add Gaussian **noise** is provided. A range of noise scenarios can be applied.
3. Select the User data **training/validation partitioning**. This will randomly partition the input data in a training and validation dataset. **Make sure to keep some data for validation (thus < 100% training)**. Also a range of training/validation partitioning scenarios can be applied. **If no validation is required, go directly to Retrieval.**
4. Click on **Finished**.

Multi-output Settings

Settings

Single-output

Multi-output

When multiple input variables have been selected, the following regression algorithms provide multiple-outputs with the same single model:

- partial least square regression (PLSR)
- neural networks (NN)
- kernel ridge regression (KRR).

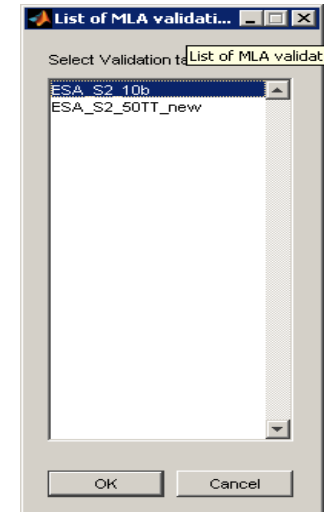
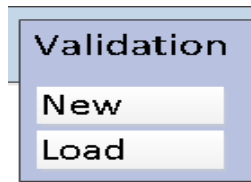
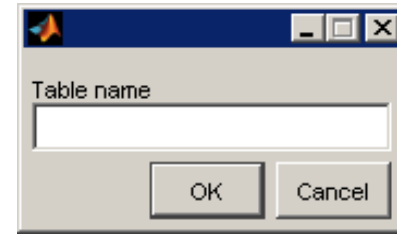
The screenshot shows a software window titled "Multi-output" with a "Bands tools" header. Below the header, there is a "Class:" dropdown menu set to "Full_image". The main section is titled "MLRA Settings" and contains a table with three rows for selecting MLRA approaches. Below the table, there are input fields for "Parameter Gaussian Noise" and "Spectral Gaussian Noise", both set to 0, with "Range" checkboxes. At the bottom, there are two sections for "RTM data" and "USER data", each with a "Train" input field, a "Range" checkbox, and checkboxes for "Only train" and "Only Val.". A "Finished" button is located at the bottom right.

	Select	MLRA approaches
1	<input type="checkbox"/>	Partial least squares regression
2	<input type="checkbox"/>	Neural Network
3	<input type="checkbox"/>	Kernel ridge Regression

The same options as Single-output are provided

Validation

- Start a **New** validation: provide a name
- All MLRA scenarios will be trained using training data and validated against validation data according to goodness-of-fit indicators:
 - R , R^2 , RMSE, RELRMSE, NRMSE, ME, MAE (see manual)
- Results will be automatically stored in a MySQL table.
- When finished, an overview table will appear (*see next slide*). Such overview table can also be consulted when selecting: **Load**. A window with generated validation results will appear: (see next slide)



MLRA validation

Options to organize statistics per Class, Parameter, stat. and # of best results

Overview of best performing MLRAs. The check boxes allow to select a MLRA model for retrieval, the graphics for graphics outputs.

Graphic options: 1:1-line, sigmas (GPR), and 2D correlation matrices.

Options to control graphics properties.

MLRA validation table: CHRIS_PROSAIL_USER_LAI_converted

MLRA validation table:

	Class	Parameter	Top	
	Full_image	LAI	1	OK

Select	Graphics	Model	spect_no...	param_n...	model_train	used_train	ME	RMSE	RELRMSE	NRMSE	MAE	R	
1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Kernel ridge Regression	20	0	0.1000	0	0.1017	1.2031	46.8297	20.2876	0.9536	0.7494
2	<input type="checkbox"/>	<input type="checkbox"/>	Partial least squares regression	20	0	0.2000	0	-0.3863	1.2921	50.2973	21.7898	0.9377	0.7346
3	<input type="checkbox"/>	<input type="checkbox"/>	Gaussians Processes Regression	20	0	0.1000	0	-0.2049	1.4884	57.9376	25.0998	1.0609	0.5971
4	<input type="checkbox"/>	<input type="checkbox"/>	Linear Regression	20	0	0.3000	0	-0.2859	1.6949	65.9738	28.5813	1.2989	0.6682

One2One Draw Export draw Export Table

Linear Regression

Fixed parameters

Empty 0

Empty 0.1

Draw Panel

Axis X spect_noise min 0

Axis Y model_train max 50

☒ Reset settings ☐ Export Settings Draw

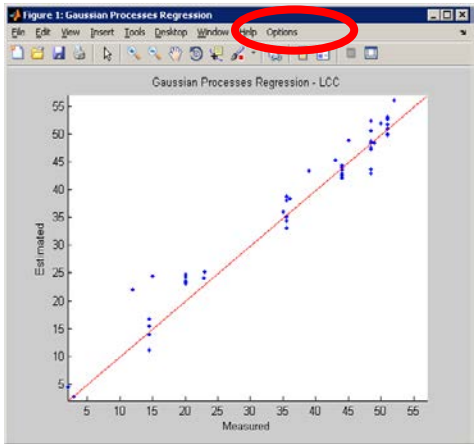
Option to export a table to a .txt file.

Selected MLRA model. With Done it will be moved to Retrieval.

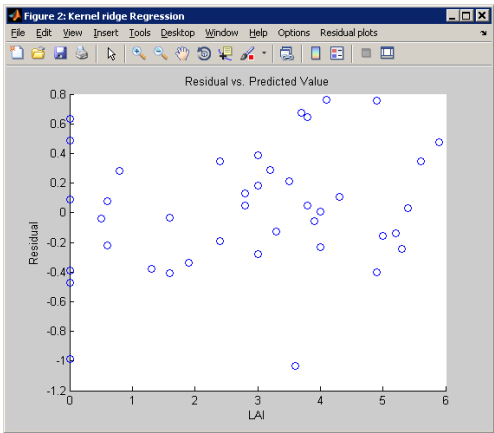
1. Choose how to **sort outputs**, according to parameter, statistic and number of top results per regressor. Click on **OK**.
2. **Select a MLRA scenario for retrieval** (e.g. the top performing one). It will move to lower panel. When clicking on **Done** it will move to the Retrieval window (slide 11).
3. Select a MLRA scenario for **Graphics plottings: 1:1-line measured vs. predicted**. For GPR additional band relevance information will be provided. **Make sure to have User data loaded, because the selected model will be regenerated.**
4. In case ranges were introduced (noise, training/validation partitioning), validation results can be plotted in a **2D-matrix**. Results are plotted according to selected parameter and statistic.

Examples of validation results

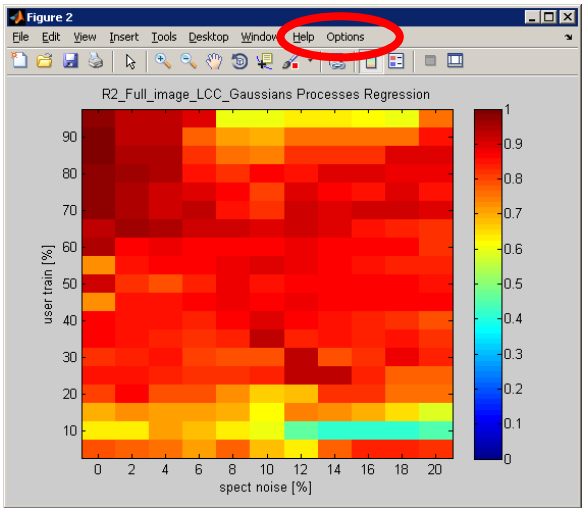
Measured data (validation) vs.
estimated data along 1:1-line



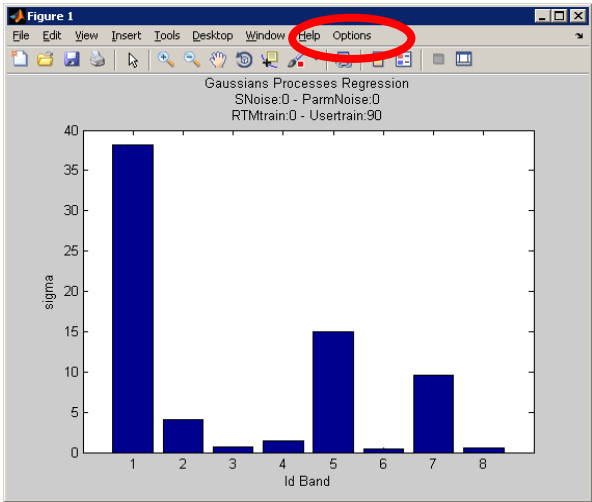
Residuals



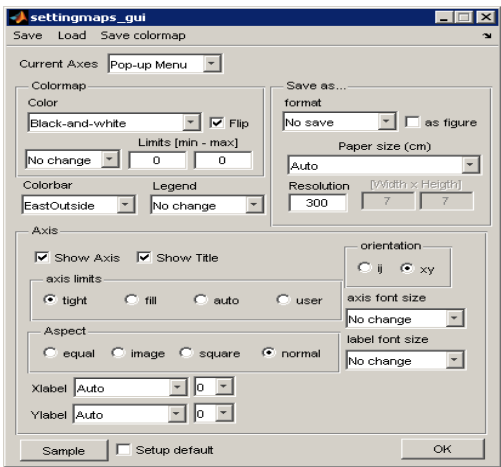
2D-matrix with validation results according to
selected parameter and statistic.



For GPR sigmas (band relevance) are provided. The
lower the sigma, the more important the band.



When clicking on **Options**, options are provided to
control the figure properties and export it.



Retrieval

Options to organize statistics per Class, Parameter, stat. and # of best results

Select per land cover class, parameter and MLRA.

Options to add noise and select training %. Here, 100% training can be chosen.

Selected MLRA model.

Mapping options

gui_mod_mla09

Bands tools

Retrieval configuration

Select class: Full_image

Inversion Parameter: LCC

MLR algorithm: Gaussian Processes Regression

Parameter Gaussian Noise [0-100%]: 0

Spectral Gaussian Noise [0-100%]: 0

RTM data [0-100%]: 0

USER data [0-100%]: 100

Train: []

Train: 100

Delete selected

Delete all

ADD

	Class	Parameter	MLRA	spect_no...	param_n.
1	Full_image	LCC	Gaussian Processes Regressi...	0	

Select class: Full_image

Parameter: LCC

Bands: Select option:

OK

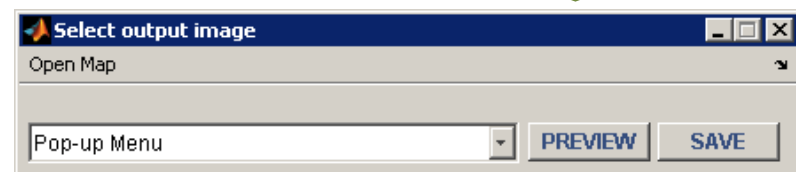
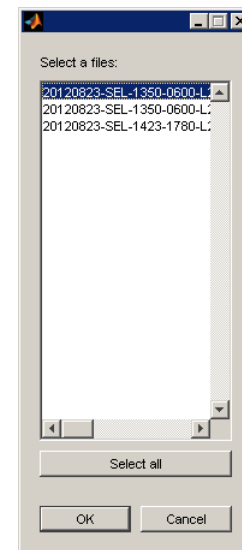
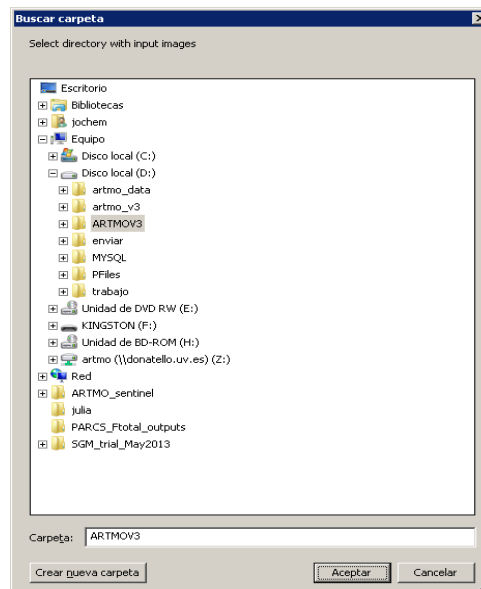
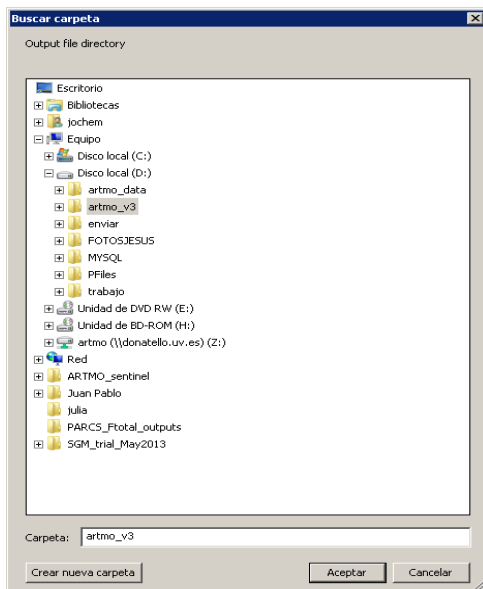
Instead of going through the validation procedure, **one can also choose to immediately train a regression model and apply it to a remote sensing image**. User data has for training to be first inserted (see slides 4 & 5).

1. Select the **parameter** and **the regression model**.
2. Optionally **noise** can be added.
3. Select the **training partitioning**. Here 100% training data can be applied. The configuration need to be **ADDED** and chosen model will appear in the down panel. **In case a model has been selected during the validation step, it will directly appear in that panel.**
4. Band tools options are provided (e.g. spectral subset, PCA).
5. When clicking on **OK**, the mapping procedure will start (see next slide).

Retrieval

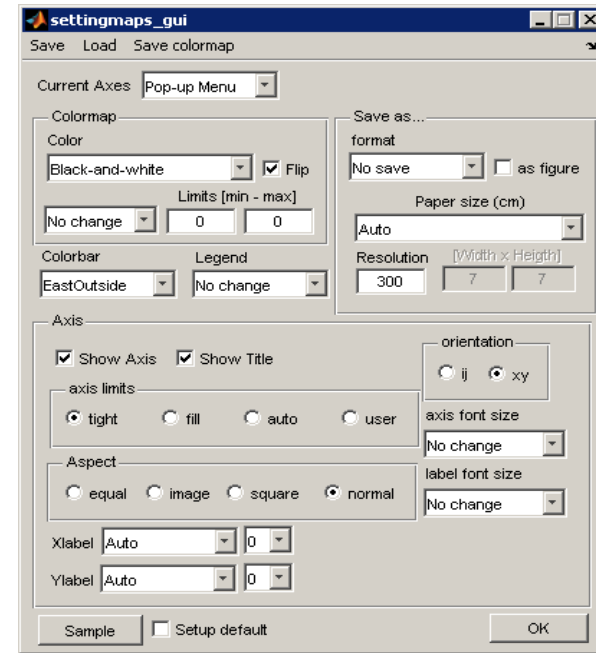
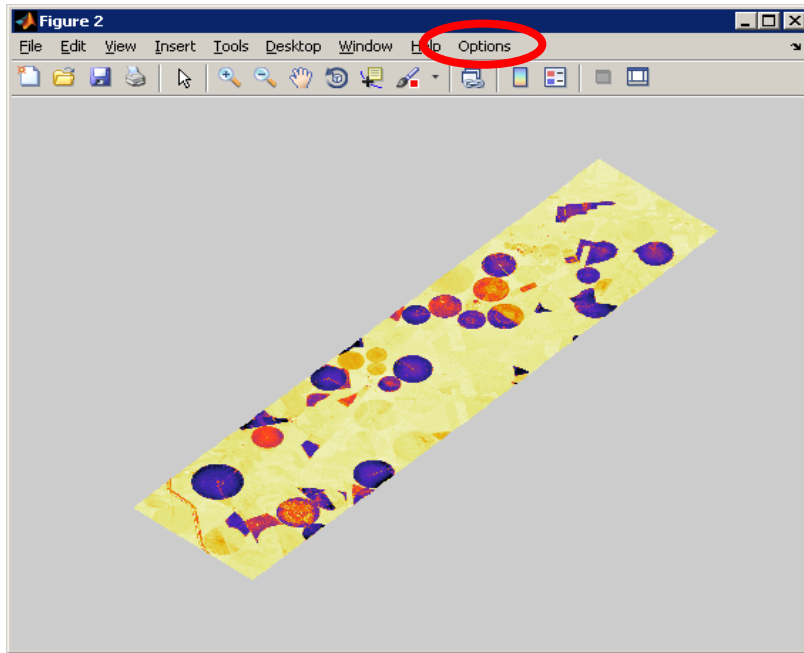
The mapping of selected biophysical parameter requires the following steps:

1. Select the **Output directory**
2. Select the directory with **Input images**
3. Images according to **ENVI file format** (including .hdr file) will be identified and listed. **Multiple images can be selected.** They will be processed one-after-another.
4. When the processing is done, the output maps can be viewed. Select one through **Open Map** and click on **PREVIEW**.



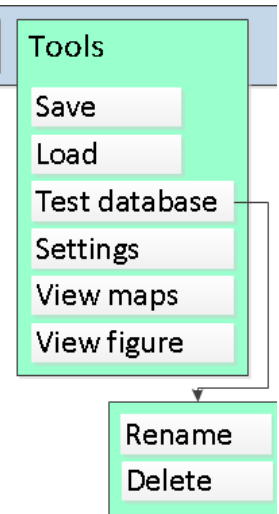
A drop-down list will show the provided output layers. One output map can then be previewed.

Final maps

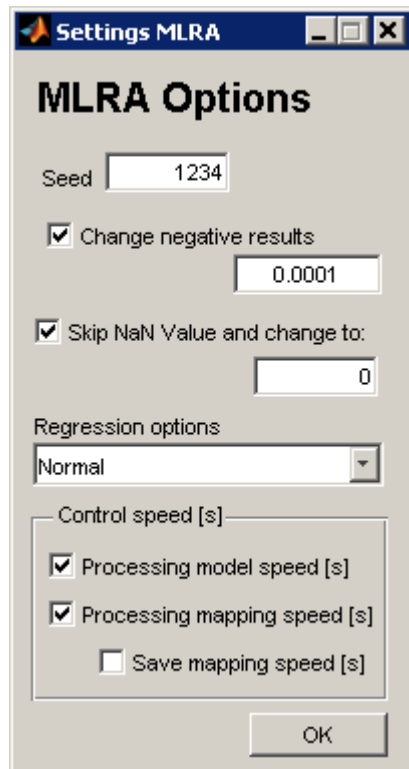


- Visualization of an output layer. In **Options**, map properties can be controlled (e.g. color scale, color table).
- Make sure to orient the map according to **ij** for **correct orientation**.
- The map can be **saved** according to various vector or bitmap formats. Redundant white space around the figure will be automatically removed.
- Settings can be set as **default** – will be automatically applied to subsequent maps.
- Click on **Sample** to visualize the map. Click on **OK** to save it away.

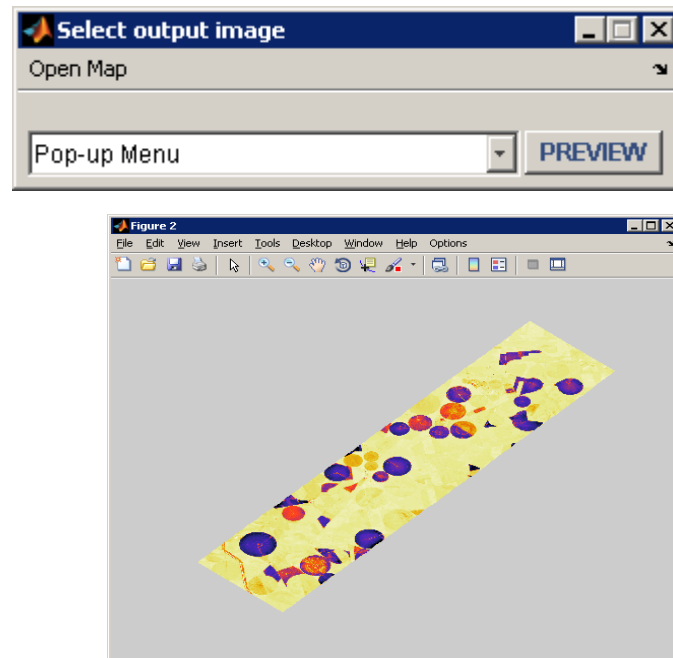
Tools



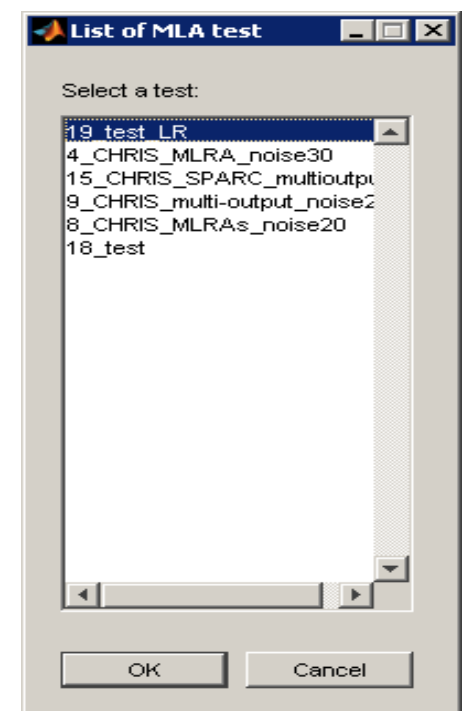
Options



View a map – select a band



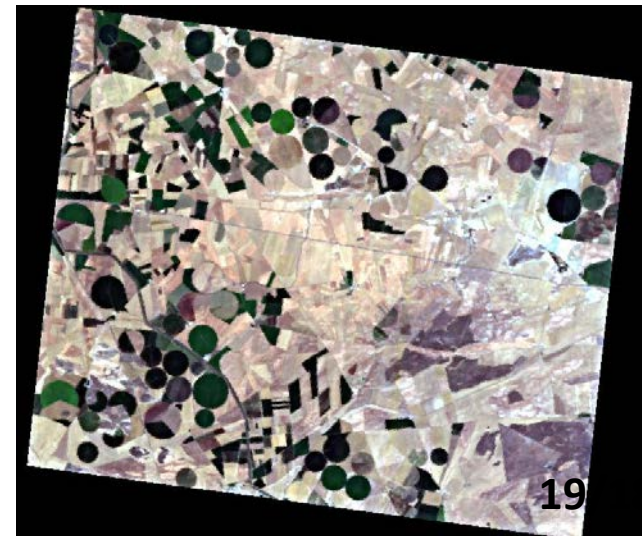
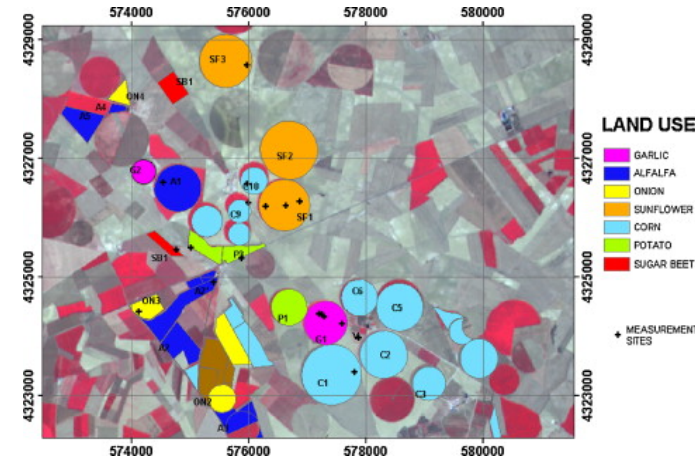
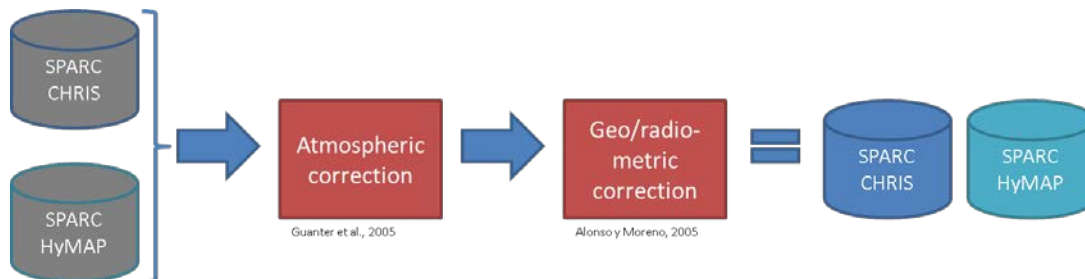
Delete validation tables



- **Field data (135 points):**

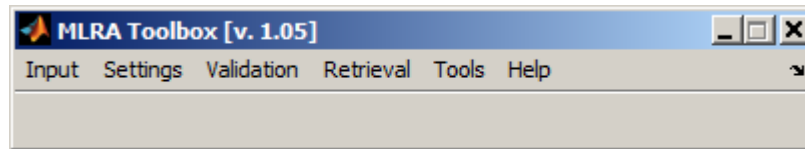
- **Leaf Chl** measured with CCM-200
- **LAI** measured with LiCor LAI-2000
- **FVC** measured with hemispherical photographs
- 30 additional **bare soil samples**

- **CHRIS mode 1** (62 bands; 34m) nadir spectra (July 2003). Data has been resampled to Sentinel-2 (20 m: 8 bands)
- **HyMAP** (125 bands; 5 m)



Exercise

- ✓ Evaluate the performance of MLRAs using a field dataset and remote sensing data.
- ✓ Apply the best performing regression algorithm to a RS image.



1	2	3	4	5	6	7	8	9	10	11	12	13
48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0

