ARTMO'S NEW SPECTRAL INDICES (SI) MODULE TO RAPIDLY EVALUATE A MULTITUDE OF SI FOR MAPPING OF BIOPHYSICAL PARAMETERS

Jochem Verrelst[1], Juan Pablo Rivera[1], Adrian Guadalajara[1], Jesus Delegido[1], Jose Moreno[1]

1. University of Valencia, Image Processing Laboratory, Valencia, Spain; jochem.verrelst@uv.es

ABSTRACT

As part of the scientific Automated Radiative Transfer Models Operator (ARTMO) Graphic User Interface (GUI) software package, a new module, being 'Spectral Indices Module' has been developed. The toolbox provides a seamless link between inputs and outputs required for running a suite of reflectance models both at the leaf level and at the canopy level. ARTMO facilitates consistent and intuitive user interaction, thereby streamlining model setup, running, storing and output plotting for any kind of optical sensor operating in the VNIR range. In this work ARTMO version 3 (V.3) is presented. It differs from earlier versions that it is completely redesigned in a modular architecture. As such, new models and modules can be easier implemented into the toolbox. Specifically, the 'Spectral Indices Module' enables to analyze the sensitive power of established and generic SIs in an automated manner. Multiple options have been implemented, e.g., controlling calibration/validation data partitioning, adding noise, and the possibility to select all kinds of curve fittings (e.g., linear, exponential, power). As a showcase, here for ground-based data coming from the SPARC campaign (Barrax, Spain) all 2-band combinations according to SR and NDVI formulation were analyzed for the hyperspectral CHRIS sensor [410-1000nm]. Regions with high accuracies are to be found when combining bands in the red edge and in the PRI region. Systematic analysis, e.g. as automatically done by the Spectral Indices Module, is recommended to ensure accurate estimations.

INTRODUCTION

With the forthcoming superspectral Sentinel-2 and Sentinel-3 missions and the planned EnMAP and PRISMA imaging spectrometers, operational Earth observation (EO) is reaching a state of maturity. This unprecedented data availability requires processing techniques that are easy and fast to apply to obtain information about the plants' growth or health status. The easiest way to obtain maps of biophysical parameters such as leaf area index, chlorophyll content, fractional vegetation cover, etc., is still through the calculation of simple vegetation indices. While a panoply of spectral indices (SI) have been developed, it remains nevertheless unknown whether the most appropriate SI has been applied until all of them have been tested. Traditionally, a spectral index such as the conventional NDVI is linked to a biophysical parameter by using a linear regression (e.g., 1). Alternative regression models, such as exponential, power, etc., are also commonly applied (e.g., 2). In statistics, this approach belongs to the category of parametric regression, and is characterized by the introduction of various constraints in order to achieve a predetermined form to link from spectral data to the parameter of interest. In fact, constraints are being introduced in SI-based regression at three levels:

1. Band selection: Typically a SI consists of only 2 or 3 bands. However, how do we know whether the most sensitive bands were selected? This question is particularly relevant in view of hyperspectral data.
2. Formulation: How do we know whether the used formulation is the most appropriate one? For instance, why would it be that a formulation such as (b2-b1)/(b2+b1) is most appropriate. Also, given hyperspectral data, there is no reason to believe that a formulation based on only 2 bands would lead to best relationships.

3. Curve fitting. Typically a linear regression is applied. However, how do we know whether a linear regression would be most optimal? In fact, as saturation effects are common, particularly at dense canopies, exponential or power functions can be more successful.

Hence, having constraints introduced at three levels within the regression function, it is doubtful whether the conventional NDVI-kind of linear regression will lead to most accurate biophysical parameter estimations. More likely, given the countless number of possible band combinations a suboptimal approach has been applied. In order to find out which parameter strategy is most successful, therefore, there is a need for a rapid and systematic evaluation of all possible band combinations, SI formulations and curve fittings prior to applying the most appropriate one to an imagery. This is not a trivial task, not in the least in view of hyperspectral sensors where a large number of combinations needs to be evaluated for each retrievable parameter to achieve an optimized NDVI-type of relationship.

To facilitate and automate this task, here we present a recently developed 'Spectral Indices Module' that allows systematically analyzing SI-based relationships at the aforementioned three levels of constraints. This module is being implemented within the innovative toolbox called ARTMO: "Automated Radiative Transfer Models Operator" (3).

The following sections will first briefly describe the latest status of the ARTMO toolbox, and then the most important components of the new 'Spectral Indices Module' will be introduced. Subsequently the used data is presented that has been applied to demonstrate the utility of the toolbox through some showcase. A conclusion closes this paper.

ARTMO V.3

A first version of ARTMO has been presented at the 7th EARSeL Imaging Spectrometry Workshop 2011 (Edinburgh, UK). In short, ARTMO brings multiple leaf- and canopy-RT models together along with essential tools required for semiautomatic retrieval of biophysical parameters in one graphical user interface. The toolbox, developed in Matlab, permits the user: i) to choose between various invertible leaf and canopy RTMs with varying complexity (e.g., PROSPECT-4, PROSPECT-5, 4SAIL, SLC, FLIGHT), ii) to choose between spectral band settings of various air- and space-borne sensors or defining new sensor settings, iii) to simulate a massive amount of top-of-canopy (TOC) reflectance spectra of any sensor in the range of 400 to 2500 nm based on look-up tables (LUT) which are then stored in a database, and finally, iv) to run various retrieval strategies.

In comparison to the first version, an updated version is presented here, being ARTMO 3 (V.3). Various major changes and new modules have been introduced in this version. The most important ones are briefly listed below:

- ARTMO has been completely redesigned and is now organized in a modular way. The modular architecture makes possible that easily RTM models can be added or removed. The idea behind this modular design is that the existing models can be seamless coupled with new or other types of models. For instance, it is foreseen to couple canopy models with atmospheric models so that to top-of-atmosphere radiance can be simulated and inverted.
- Internally, the MySQL database has been reorganized in a more efficient manner to support the modular architecture, to avoid redundancy and to speed up processing.
- Various new retrieval modules have been designed, based on parametric regression, non-parametric regression and physically based inversion. This led to the development of 1) 'Spectral Indices module', 2) 'Machine learning module' and, 3) 'LUT-based inversion', respectively.

Figure 1 presents ARTMO V.3's main window. Compared to earlier versions, the main window has been considerably simplified. Now in the main window a new project can be initiated, a sensor chosen and a comment added, but all modules are accessible through drop-down menus at the top bar. These drop-down menus depend on the modules and tools found within the ARTMO folder

and can thus easily be expanded. A systematic overview of the drop-down menu is provided in Figure 2. The focus of this paper is on 'Spectral Indices module'.
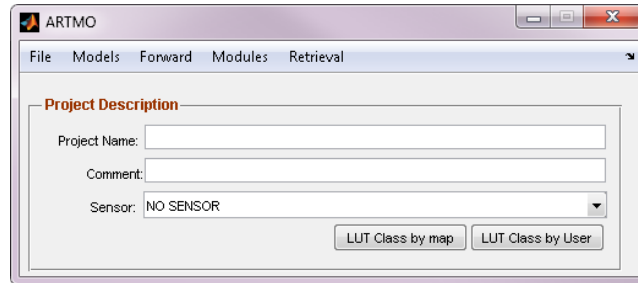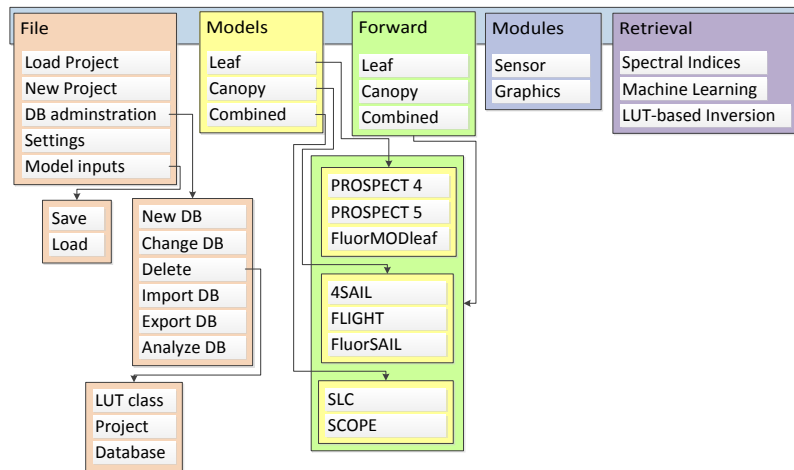

Figure 1: ARTMO V.3's main window

Figure 2: ARTMO's main modular architecture.

1. Spectral Indices Module

The 'Spectral Indices Module' is again designed in a modular manner. Figure 3 displays the module as displayed by Matlab, while a schematic overview of the dropdown menu is shown below. In short, the Spectral Indices module enables to load and apply any kind on spectral index. For each index and each band a wavelength range can be defined. As such, all bands of a sensor within this range can be evaluated search for the most sensitive bands. At the same time, it is also possible to evaluate all bands of a sensor according to the formulation. Second, a variety of curve fittings have been implemented and new ones can be easily added. Two sorts of input data can be loaded into the module for calibration or validation of the relationships. On one hand, data can come from the RTMs. In this way an earlier generated LUT with simulated spectra can be imported. On the other hand, user-defined data can be imported, e.g. as collected during field campaigns. Options are available to merge and partition both datasets, e.g. calibration on the basis of RTM data and validation on the basis of user-defined field data. Furthermore, if a land cover map is loaded then for each land cover class different SI optimization strategies can be defined. Most importantly, when having validation data available, then all band combinations, curve fittings, etc. can be analyzed against the validation and statistical results (e.g. R^2, RMSE, NRMSE) are stored in a relational database. The best performing strategy can then be loaded and applied to an image. In the following sections the most important modules will be explained in a bit more detail. These are: 1) the tool to 'Add a spectral index', 2) 'SI settings', 3) 'Test results', and 4) 'Retrieval'.
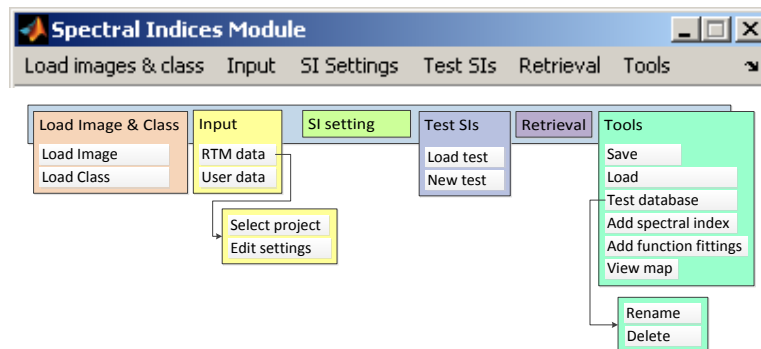
Figure 3. ARTMO's 'Spectral Indices Module' architecture.

2. Add spectral index

Starting with 'Add spectral index' (Figure 4) this option can be found within Tools and allows the user to add a new index to the list. Spectral indices can be manually created through the GUI or alternately can be imported from other databases through the 'DB tools'. When selecting the latter, then automatically SIs from other DBs will be detected and can be loaded accordingly. Spectral indices are organized per 'Group' and enables to group similar spectral indices, e.g. broadband, narrowband, pigments, water, etc. For instance, 'Broadband Greenness' consists of conventional SR, NDVI, EVI indices. A new SI Group can always be created, and within a Group, a new SI can be created. To do so, the following input is required: add a name and an Acronym, and then define the equation, e.g., an NDVI formulation (b2-b1)/(b2+b1). When clicking on 'Eval' the equation will be evaluated and bands are recognized. In the lower panel for each band a default value or a minimum and maximum boundary range can be inserted. When these boundaries are defined then automatically all bands within this range will be evaluated in the 'Test results' module for a given a selected sensor. When instead a zero value is inserted then the whole range of bands of a given sensor is evaluated according to the defined formulation.
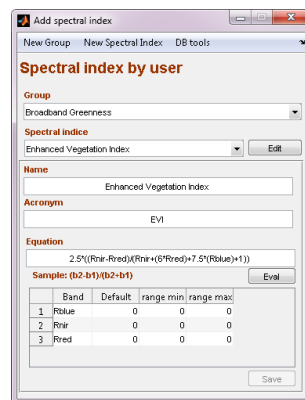

Figure 4: GUI to add a spectral index.

3. SI settings

The following step addresses the analysis of multiple SI-based retrieval strategies. A first step to do is inserting RTM LUT-data or ground truth data through the Input module. Required input data refers to retrievable biophysical parameters and associated spectra. Once this is done the 'SI settings' module is activated (Figure 5).

This GUI configures the parametric regression strategies through five successive steps. First, if multiple land cover classes have been defined (within the 'Load Image and Class' window) then retrieval strategies can be configured per land cover class. Second, a group of spectral indices has

to be selected. The included indices are then listed below. Multiple spectral indices can be selected within the same group. Note that only within one group indices can be selected. Instead, within the 'Add spectral index', as earlier described, it is possible to create a generic group that includes all indices. Third, multiple curve fittings can be selected, being 'linear', 'exponential', 'logarithmic', 'power', 'polynomial second degree'. In Tools new curve fitting functions can be added. Fourth, options to add Gaussian noise are provided. Noise can be added both on the parameters, as on the spectra. Here, a range of noise can be configured, so that multiple noise scenarios can be evaluated. The injection of noise can be of importance to account for environmental and instrumental uncertainties when synthetic spectra from RTMs are used for calibration. Fifth, the cal/val data partition can be controlled by setting the percentage how much data from a RTM or user-defined is assigned to calibration or to validation. Both datasets can also be merged by selecting a portion of both datasets for calibration and validation. Also here multiple cal/val partitions can be evaluated.
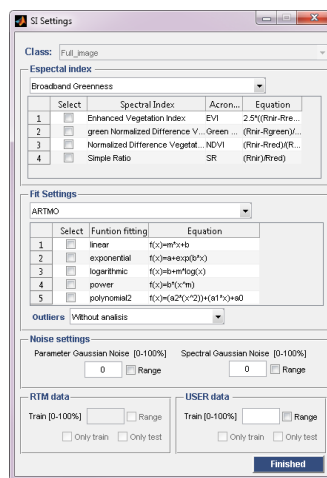

Figure 5: GUI to configure the SI settings.

Once having the cal/val data partitioning defined and SI settings configured, then those scenarios can be run. This is done through naming a new test in the 'Test SI'. Subsequently, one-by-one all retrieval strategies over the band ranges and configuration ranges are analyzed. The analysis is executed in two consecutive steps:
1. Each SI is for all selected bands evaluated against the calibration data through correlation (R^2). A correlation matrix can be obtained and best evaluated scenario can be consulted.
2. Each obtained relationship subsequently leads to estimation of the targeted parameter according to a curve fitting. These estimations are then evaluated against the validation dataset through a wide range of evaluation statistics, being R^2, RMSE, NRMSE, MAE, ME. For a single selected relationship the curve fitting and a scatterplot of estimated results against validation measurements can be plotted.

In case data is partitioned in a calibration and validation set then those outputs are twice generated; first for the calibration dataset, and second for the validation dataset.

4. Test results

All results are automatically saved in a MySQL database. This has the advantage that a large number of results can be stored in a systematic manner and that results can be easily queried. In this tabular GUI best results are selected per class, parameter, cal/val, and then ordered according to a statistic. Moreover, all kinds of plotting options are provided to show the results, e.g. as displayed in the Results section. Finally, by clicking on 'Retrieval' an analyzed regression function

can be selected for each retrievable parameter (e.g. the best one). Such regression function can then be applied to an image.

Figure 6: GUI with results of tested configurations.

5. Retrieval

In the Retrieval GUI t is also possible to directly configure a relationship and apply to an image to map a parameter. This can be useful in case a relationship is already known (e.g. coming from literature) or when no validation data is available. Hence, the user can select the required land cover class (if available), the retrievable parameter, the SI group and SI, the required curve fitting and required parameters. Similarly, noise can be added to the spectra or parameters and the size of the calibration data can be selected. Multiple regression strategies can be added, e.g. for each retrievable parameter another one. Finally, by clicking on OK, the input images can be loaded and the output maps will be written away in an ENVI format.


Figure 7: SI retrieval GUI

USED DATA FOR SHOWCASES

A diverse field dataset, covering various crop types, growing phases, canopy geometries and soil conditions was collected during SPARC (Spectra bARrax Campaign). The SPARC-2003 and SPARC- 2004 campaigns took place in Barrax, La Mancha, Spain (coordinates 30º3´N, 28º6´W, 700 m altitude). In the 2003 campaign, carried out on 12-14 July, biophysical parameters were measured within a total of 113 Elementary Sampling Units (ESU) among different crops. ESU refers to a plot size compatible with a pixel size of about 202 m. In the 2004 campaign, carried out on 15-16 July, the same field data were collected within a total of 18 ESUs among different crops. Leaf chlorophyll content (LCC) was derived by measuring within each ESU about 50 samples with a calibrated CCM-200 Chlorophyll Content Meter. LCC values obtained in the SPARC 2003 campaign show good agreement with these obtained in the SPARC 2004 campaign. Green LAI was derived from canopy measurements made with a LiCor LAI-2000 digital analyser. Each ESU was assigned to a LAI value, which was obtained as a statistical mean of 24 measures (8 data readings x 3 replications) with standard errors between 5 and 10%. For both years, we have a total of 9 crops (garlic, alfalfa, onion, sunflower, corn, potato, sugar beet, vineyard and wheat), with field-measured values of LAI that vary between 0.4 and 6.3 and LCC between 2 and 55 µg/cm^2. Further details on the measurements can be found in (4). Additionally, 60 random spectra over bare soils, man-made surfaces and water bodies were added to broaden the dataset to non-vegetated samples (i.e., with a biophysical LCC and LAI value of zero).

During the campaign hyperspectral CHRIS images were acquired. CHRIS provides high spatial resolution hyperspectral data over the VNIR spectra from 400 to 1050 nm. It can operate in different modes, balancing the number of spectral bands, site of the covered area and spatial resolution because of on-board memory storage reasons. We made use of nominal nadir CHRIS observations in Mode 1 (62 bands, maximal spectral information) for the four SPARC campaign days, where field measurements of surface properties were measured in conjunction with satellite overpasses. CHRIS Mode 1 has a spatial resolution of 34~m at nadir. The spectral resolution provides a bandwidth from 5.6 to 33 nm depending on the wavelength. The images were geometrically corrected followed by atmospheric correction (see (4) for details).

SHOW CASES

Since the focus of this work was on the development of a SI module, here only a few showcases based on hyperspectral CHRIS data (62 bands) are shown. To start with established index formulations, first results using the formulation SR (*b2*/*b1*) and NDVI ((*b2-b1*)/(*b2+b1*)) are presented. Accordingly, all 2-band combinations are first calculated and then correlated with the retrievable parameters (i.e., LAI & LCC). For this first exercise all input data was used for calibration in order to obtain full understanding on the impact of band combinations. First the R^2 correlation matrices are shown (Figure 8).
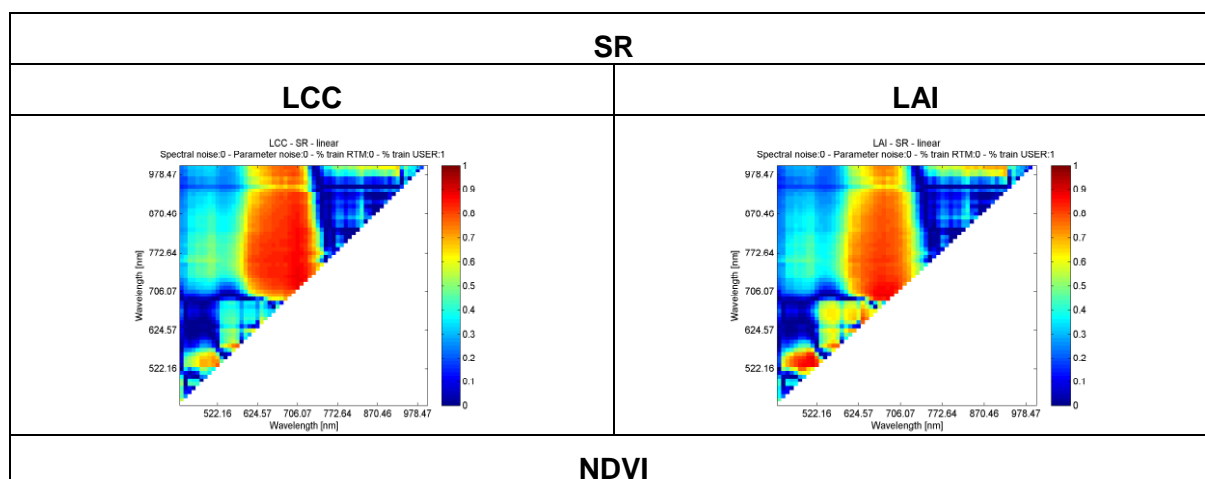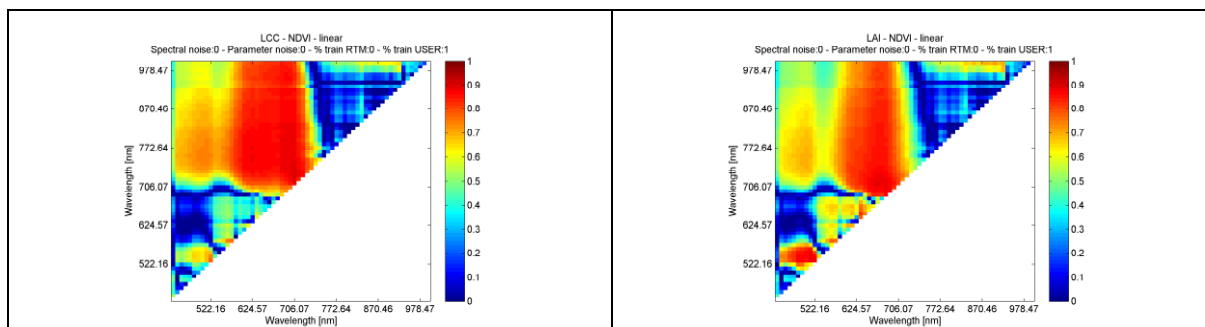
Figure 8: R^2 results between SI and calibration dataset for SR [top] and NDVI [bottom] and LCC [left] and LAI [right]

Table 1: Band combinations of best evaluated results R^2 matrices Figure 8.

SI	Parameter	Band combination	R^2
SR	LCC	706.07 - 731.23	0.89
	LAI	511.53 - 542.36	0.89
NDVI	LCC	706.07 - 731.23	0.89
	LAI	674.42 - 712.17	0.89

When inspecting these correlation matrices the following observations can be made. First, in each of the matrices the region at the red edge arises as most successful in estimating LCC or LAI. This sensitive region continues up to in the NIR, but most predictive bands are found within red edge (LCC) or combination with red (LAI). Second, SR and NDVI provided very similar results. While NDVI responds relatively more successful in the blue and NIR, the most successful regions are similar for both indices. This is also visible in Table 1 where the band correlation leading to best R^2 is shown. Both SR and NDVI formulation yielded the same best results for LCC. Third, also LCC and LAI yielded quite similar patterns. This similarity can be attributed by sampling on a heterogeneous agroecosystem that is controlled by irrigated parcels. As such, for a large number of crop types both parameters are related to each other; i.e. mature crops that are characterized by high biomass (LAI) and high LCC as opposed to juvenile crops characterized by low biomass and low LCC. Fourth, it is noteworthy that a small region in the green (around 511 and 540 nm) also leads to strong correlations (see bottom of triangle). Remarkably, this is the region where PRI is to be found. Typically, the PRI (531 with 570 nm) is sensitive to changes in carotenoid pigments (e.g. xanthophyll pigments) in live foliage. Carotenoid pigments are indicative of photosynthetic light use efficiency, or the rate of carbon dioxide uptake by foliage per unit energy absorbed (5).

Having obtained a first insight in the predictive power of calculated SIs through correlation, the next step is calculating the estimations through curve fitting. To analyze the success of the regressor, these estimations were subsequently evaluated against the validation dataset. In case 100% of the input data is used for calibration then validation occurs against the data. Various output statistics are provided, as also shown in Figure 9. For instance, the normalized RMSE (NRMSE [%] / range validation data * 100) allows comparing the predictive power across the different retrievable parameters. Users typically require an error threshold of 10%. NRMSE results are displayed below.

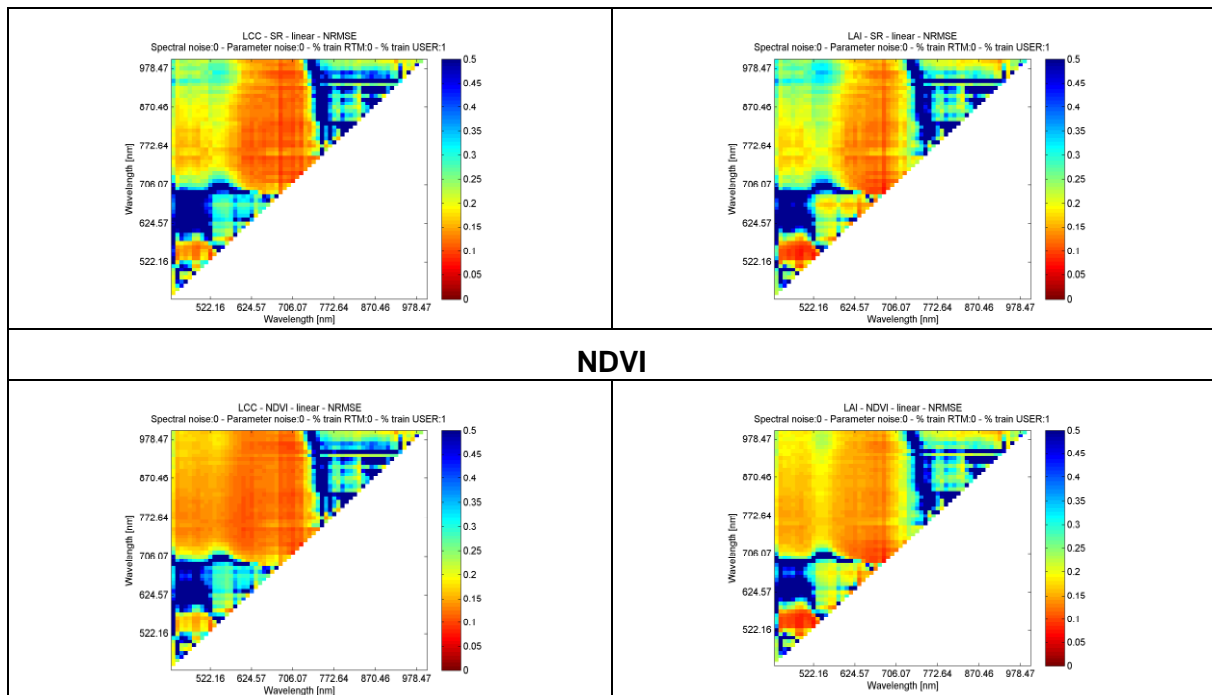| SR	
LCC	**LAI**

Figure 9: NRMSE results between SI and calibration dataset for SR [top] and NDVI [bottom] and LCC [left] and LAI [right].

Table 2: Band combinations of best evaluated results NRMSE matrices Figure 9.

SI	Parameter	Band combination	NRMSE [%]
SR	LCC	718.37 - 724.74	8.13
	LAI	688.41 – 674.19	7.45
NDVI	LCC	718.37 - 724.74	8.40
	LAI	522.16 – 542.36	7.45

The NRMSE figures confirm the earlier observations that the differences between LCC and LAI are rather small. Also SR and NDVI perform alike, with this difference that NDVI has a much wider range in the blue and NIR region with relatively good performances. More importantly, overall the same patterns as the correlation matrices appeared, though small local variations took place. In fact, it can be observed that there is one band that leads particularly good performances over the NIR, being at 706 nm. Nevertheless, the best band combination for LCC was more in the middle of the red-edge region, whereas for LAI strongest relationships were evaluated in the PRI zone.

When it comes to mapping it needs to be decided on which basis the best performing relationship is selected. What matters hereby is how well the relationships perform when comparing against validation data, e.g. by using NRMSE. Therefore, results from table 2 seem to be more indicative. Figure 10 displays maps over Barrax region according to the NRMSE obtained relationships. While results are indicative, it was noted that some striping effects remaining in the CHRIS image were propagated to the map. Also LAI results seem overestimated, especially over the dried-out fallow land (non-irrigated surfaces). Hence, while being a fast and convenient method care is required when interpreting the map. In fact, the main disadvantage of the SI method is that no associated per-pixel uncertainties are provided. Alternative methods such as Look-up table (LUT)-based inversion or through some machine learning regression algorithms (MLRA) do provide such uncertainties at the expense of more computational demanding.
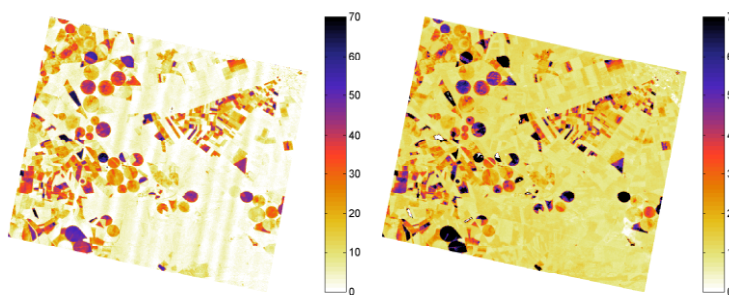
Figure 10: LCC (left) and LAI (right) maps as obtained by the NDVI formulation of Table 2.

More results can be easily generated with the 'Spectral Indices Module'. For instance, the performance alternative curve fittings (e.g., exponential, power) or RTMs can be evaluated, or new type of index formulations can be tested across all bands. We have tested 3-band formulations and they led to slightly improved accuracies (results not shown for brevity) as compared to the evaluated 2-band formulations. This leads us to the conclusion that existing vegetation indices are most likely suboptimal performing, especially in view of hyperspectral data where hundreds of bands are available. Thus, it is very likely that still more powerful SIs can be developed. With the SI module we offer the tools to seek for more powerful formulations. The toolbox is freely accessible at Valencia University.

CONCLUSIONS

In this work ARTMO version 3 (V.3) is presented. It is designed in a modular architecture and consists of various new modules. Specifically, the 'Spectral Indices Module' enables to analyze the sensitive power of established and generic SIs to map biophysical parameters in an automated manner. Multiple options have been implemented, e.g., controlling calibration/validation data partitioning, adding noise, and the possibility to select all kinds of curve fittings (e.g., linear, exponential, power). As such, all kinds of scenarios can be automatically evaluated and best performing relationship can be applied to an image.

ACKNOWLEDGEMENTS

REFERENCES

1. Elvidge C D & Z Chen, 1995. Comparison of broad-band and narrow-band red and near-infrared vegetation indices. Remote Sensing of Environment, 54 (1), 38-48.

2. Carlson T N & D A Ripley, 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. Remote Sensing of Environment, 62 (3), 241-252.

3. Verrelst J, J P Rivera, L Alonso & J Moreno, 2011. ARTMO: an Automated Radiative Transfer Models Operator toolbox for automated retrieval of biophysical parameters through model inversion. Proceedings of EARSeL 7th SIG-Imaging Spectroscopy Workshop, Edinburgh, UK.

4. Verrelst J, J Muñoz, L Alonso, J Delegido, J P Rivera, G Camps-Valls & J Moreno. 2012. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel- 2 and -3. Remote Sensing of Environment 118, 127–139.

5. Gamon J A, L Serrano & J S Surfus, 1997. The Photochemical Reflectance Index: An Optical Indicator of Photosynthetic Radiation Use Efficiency across Species, Functional Types, and Nutrient Levels. Oecologia. 112, (4), 492-501.