# RS indicator vs. Yield regression calculator

Michele Meroni, FOOD SEC Action, MARS Unit, H04, Institute for the Environment and Sustainability, Joint Research Centre. Email: michele.meroni@jrc.ec.europa.eu

## Objective: to correlate Remote Sensing (RS) metrics to yield

Make use of:

* Department map\*
* Crop mask (actually crop area fraction mask, BYTE values from 0 to 100% expressed with DN 0-200)\*

\* spatially registered with the RS product and all at VGT resolution

To:

* compute the crop area fraction weighted RS indicator for each region and year, regress it regionally and globally with yield

The output of the code is represented by the coefficients of the linear regression to be used for the forecast (and associated error statistics). Be aware that in order to apply such coefficients, the department level aggregation of the RS products has to be carried out in a consistent way (see aggregation options below).

## Basic Instructions

### 1. Build the RS product stack file

Here you have to build your own stack file starting from the data you have. You can use ENVI for this purpose. The stack file must have a “bil” interleave and floating point format.

### 2. Extract the dekads to be used for regression

The objective is to build the stack file of the RS product (n band = n years) to be regressed against the yield. In the case of CNCT approach these are all the possible yearly time series of dekad X (dekad 12 – last dekad of April, being found the most correlated with yield). Here be care that these yearly time series are named “dX”. When X > 36, the calendar dekad is X-36.

|  |
| --- |
| Set user parameters in the procedure “dec\_files\_job\_handler2.pro” and run it |

### 3. Prepare yield and surface (optional) data in the required format

Yield and surface data have to be saved as comma separated value files with given format (3 columns without header: department id, year, and yield). Missing data are set to -999. Files must be stored in a sub directory of the base\_path (defined by user in “job\_handlers.pro”) named “\yield\_data” and they must be named surface\_+CROP\_BASE\_NAME.csv and yield\_+CROP\_BASE\_NAME.csv where, for Tunisia, CROP\_BASE\_NAME can assume the following values: “durum\_wheat”, “soft\_wheat”, and “barley”.

### 4. Regress RS indicator vs. Yield

The objective is to perform the linear regression analysis between yield data and a set of remote sensing indicators. Yield data are provided at department level.

RS indicators are aggregated at the department level. Options for this aggregation and filtering are set in the procedure “set\_options.pro”.

|  |
| --- |
| Set options in “set\_options.pro”  - Set aggregation options  - List the department to be considered in the regression  - List the departments or years to be excluded from the analysis (my suggestion is always exclude the first RS band – year 1998 – which is the first VGT year)  - Decide if the surface (for computing production) is to be computed from the crop mask or read from file  - Set plotting and warning options  - List the RS indicators to be regressed against Yield |

The whole process is managed by a so called ‘job handler’ which is an IDL procedure found in job\_handler.pro.

In this procedure the user must set the environment info (directory, filenames) and image characteristics.

|  |
| --- |
| - Set user parameters in the procedure “job\_handler\_tunisia in job\_handlers.pro”  - run job\_handler\_tunisia once for each of the analysed crop (changing the CROP\_BASE\_NAME each time) |