

1. Heading and logos

Practice on **Crop yield prediction** recommended by **SRI NASU-SSAU and UN-SPIDER RSO in Ukraine**.

2. Contact details

Main contact:

Nataliia Kussul — nataliia.kussul@gmail.com

Sergii Skakun — serhiy.skakun@ikd.kiev.ua; serhiy.skakun@gmail.com

Oleksii Kravchenko — oleksiy.kravchenko@gmail.com

Andrii Shelestov — andrii.shelestov@gmail.com

Andrii Kolotii — andrew.k.911@gmail.com

Alternative Contact(s):

3. Workflow

General description of the workflow:

(1) Data acquisition: MODIS NDVI, vector data on administrative regions and yield statistics.

(2) Identification of a trend in the crop yield and estimation of parameters a_0 and a_1 in the following equations.

$$T_i = a_0 + a_1 * i, \text{ where } i \text{ is the year}$$

(3) Deriving the NDVI values for each DOY by spatially averaging NDVI maps over the region. In general, it is advisable to make an averaging by a cropland map (e.g. from the ESA GlobCover map) or crop mask.

(4) Running the LOOCV procedure to find the DOY value that provides minimum RMSE value. LOOCV stands for a leave-one-out cross-validation procedure within which the model is routinely calibrated based on data for all years except the one (testing data). Therefore, for each administrative region and for each DOY for which NDVI values are available, we have n predicted values of the crop yield for testing data (e.g. $n=10$ when calibrating models for years 2000-2009). RMSE is the root mean square error which is equal to

$$RMSE = \sqrt{\frac{1}{n} \sum_i (P_i - O_i)^2}, \text{ where } P_i \text{ and } O_i \text{ are predicted and observed winter wheat yields, respectively.}$$

(5) Estimating parameters b_0 and b_1 in the linear regression model: $dY_i = Y_i - T_i = f(NDVI_i) = b_0 + b_1 * NDVI_{ij}$,

4. Explanations and recommendations

Abstract on the recommended practice

As population on the Earth increases and climate changes likely to impact global crop production, food security becomes a topic of great importance. This problem is addressed within the Global Agricultural Monitoring System of Systems (GLAM) that aims to integrate multiple satellite datasets and in situ observations to provide services for monitoring crop production, agro-meteorological parameters, and water resources. The two main components of crop production monitoring are crop yield forecasting and crop area estimation. Therefore, providing accurate crop yield forecasts several months in advance of the harvest is crucial at global, national and region scales.

This recommended practice focuses on using NDVI values derived from MODIS to build a linear regression model for crop yield forecasting.

Requirements

The following are required for the recommended practice:

- Data requirements: NDVI values from MOD13 product for 2000-present; vector file with administrative regions, and yield statistics as references data.
- For processing NDVI values QGIS is required. Also, a QGIS zonal statistics plugin is required. For regression model calibration it is required to use Matlab or R statistical package (freely available).
- Skills requirements: basic to intermediate knowledge of image processing and working in statistical packages.
- Hardware requirements: for processing a patch of MODIS images, it is advisable to have minimum 2 Gb of RAM.

Applications

This recommended practice targets timely crop yield forecasting to assess potential losses due to droughts.

Advantages and limitations

Advantages:

- Little data inputs requirements.
- Easy to implement.

Accuracy: up to 90% (depending on the crop and area).

Limitations:

- Lack of robustness and generalization ability.
- Data-driven, i.e. performance strongly depends on available datasets and quality of reference data.

Bibliography

1. Kogan, F., Kussul, N., Adamenko, T., Skakun, S., Kravchenko, O., Kryvobok, O., Shelestov, A., Kolotii, A., Kussul, O. & Lavrenyuk, A., (2013) **“Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and**

biophysical models", *International Journal of Applied Earth Observation and Geoinformation*, vol. 23, pp. 192-203. (<http://dx.doi.org/10.1016/j.iaq.2013.01.002>)

2. Kogan F., Kussul N., Adamenko T., Skakun S., Kravchenko O., Kryvobok O., Shelestov A., Kolotii A., Kussul O. & Lavrenyuk A. (2013) "**Winter wheat yield forecasting: A comparative analysis of results of regression and biophysical models**", *Journal of Automation and Information Sciences*, vol. 45., issue 6, pp. 68-81.

3. Shelestov, A.Yu., Kravchenko, A.N., Skakun, S.V., Voloshin, S.V., Kussul, N.N., (2013). "**Geospatial information system for agricultural monitoring**", *Cybernetics and Systems Analysis*, Volume 49, Issue 1, pp 124-132. (<http://dx.doi.org/10.1007/s10559-013-9492-5>)

5. Step-by-step procedures using QGIS

Here we will only focus on an example of deriving average NDVI values by administrative regions.

Data access

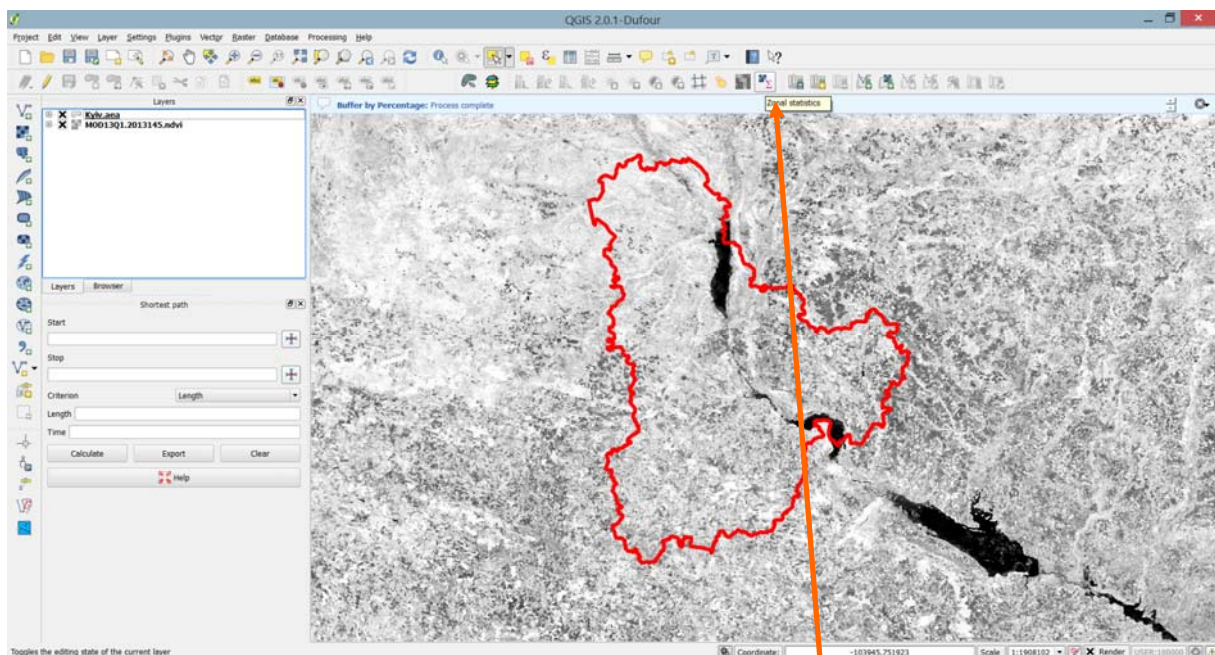
MOD13 can be freely downloaded from the LPDAAC USGS web site at <https://lpdaac.usgs.gov/>.

Data preparation / Pre-processing

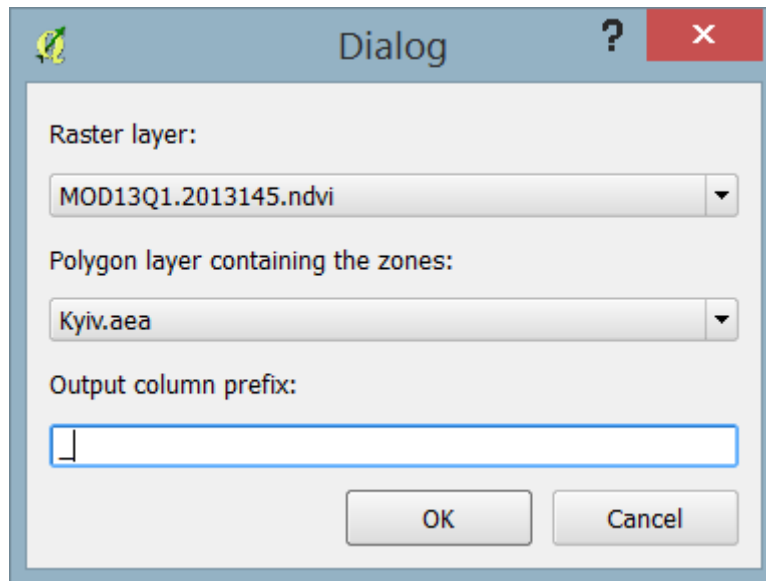
MOD13 products should be downloaded from the LPDAAC USGS. Data are provided in hdf format. The NDVI values should be extracted from the hdf file, and reprojected to the some projection, for example UTM.

Processing steps

Here we show how average raster values (NDVI) by a vector polygon (administrative region). Figure below shows an NDVI raster overlaid with administrative region.

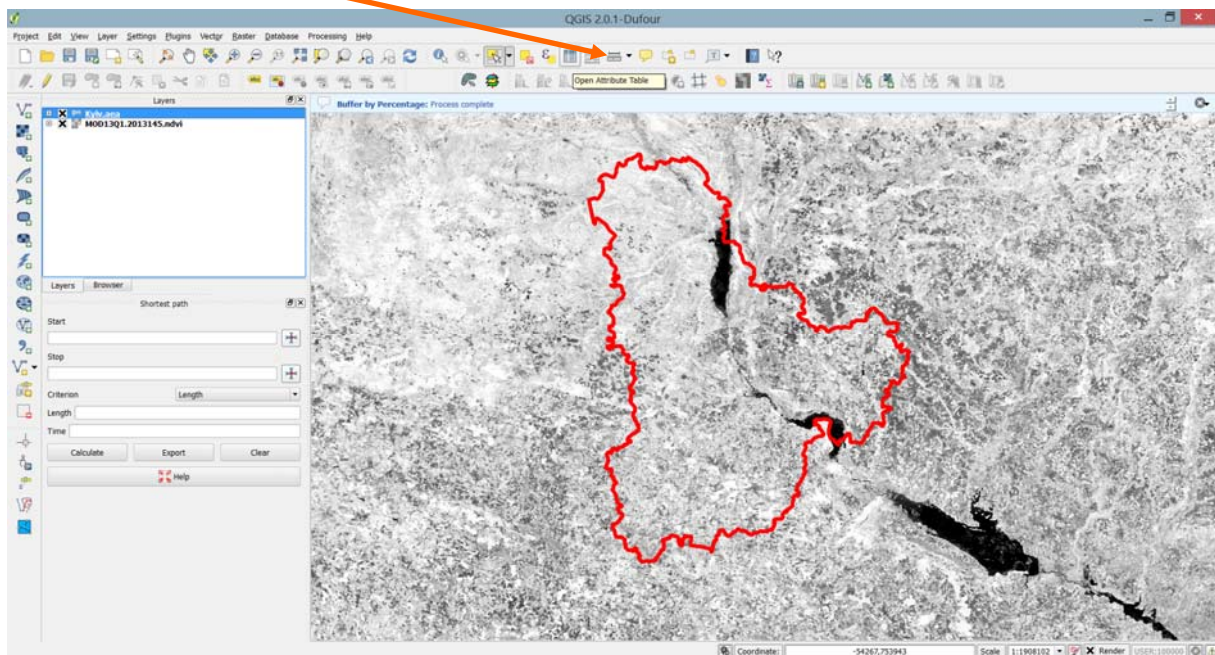


To estimate NDVI values over the region we will use a **Zonal statistics** package. After clicking the following window will appear:



Here you need to specify the raster layer, the vector layer, and a prefix which will be added to the columns of the attribute tables of the vector layer. Then click the OK button.

The statistics will be added to the vector layer. To view it, one needs to open vector layer attribute table.



Here the Attribute table:

Attribute table - Kyiv.aea :: Features total: 1, filtered: 1, selected: 0

CODE	ID	SPATU	NAME	CENTER	V_COUNTIES	POP	AREA	HUMUS	statcount	statsum	statme	_count	_sum	_mean
0181101P	9	3200000000	Ee7anuea	Ee7a	25	1911.60000	28.90000	4.41367078...	463165.000...	3262632228...	7044.21	463165.0...	3262632228.0...	7044.21...

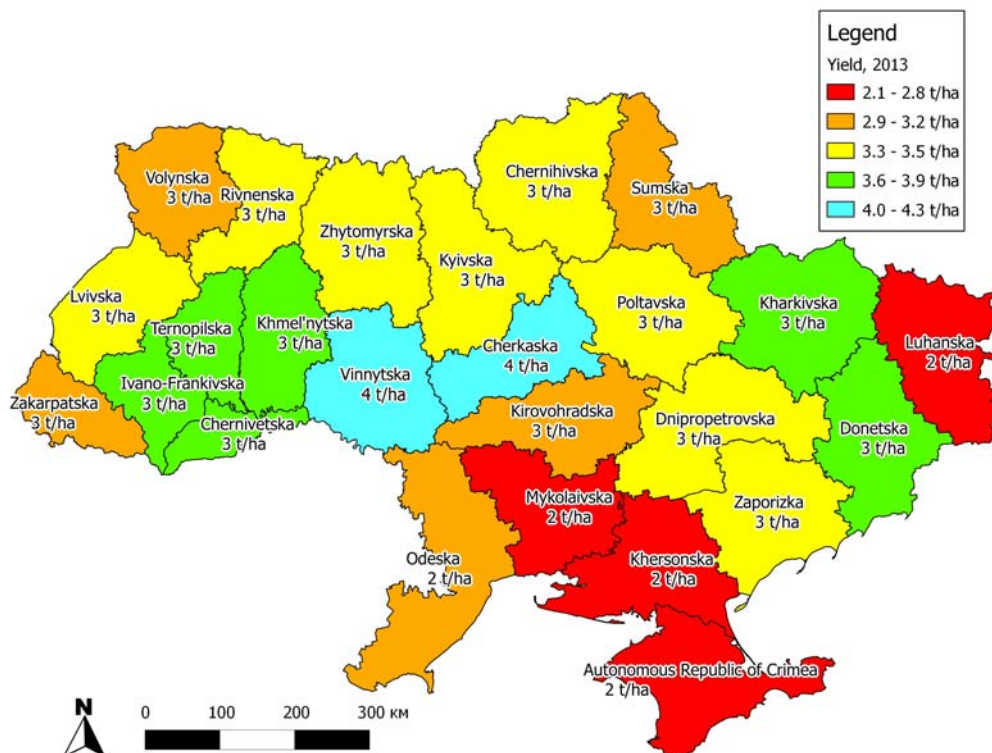
You should see 3 fields that were added to the polygon attributes: **_count** which counts the number of pixels in the vector layer; **_sum** is the total sum of raster pixel values within the vector; **_mean** is average value. In our case, the mean value is 7044 which corresponds to the NDVI value 0.7044 (MOD13 NDVI values are scaled by 10000).

Assessment of the resulting product

Models accuracy needs to be assessed in the corresponding statistical package.

6. Example of the resulting product (e.g. map)

The example of the resulting map is prediction of winter wheat for the Ukraine in 2013. The map was created on 23/05/2013, about 1,5 months before harvest.



7. Real life example (optional)

The example of the resulting product presents the real example of application of MODIS data for winter wheat yield forecasting for 2013.