

## Application Form

### Lead author, PI

- **Name:** Nataliia Kussul
- **Agency/institution:** Space Research Institute NAS Ukraine and SSA Ukraine
- **Mailing address:** Glushkov Ave 40, 4/1, 03680, Kyiv 187, Ukraine
- **Email:** nataliia.kussul@gmail.com
- **Phone:** +38 093 326 42 33

### Collaborators

- **Co-PI:** Esteban Julian Copati (ecopati@bc.org.ar), Buenos Aires Grains Exchange, Argentina
- **Collaborator:** Nilanchal Patel (nilanchal.patel@gmail.com), Birla Institute of Technology, Mesra, India
- **Collaborator:** Michael Ryabokon (r@noosphere.com), NGO Association Noosphere (Noosphere Engineering school)
- **Collaborator:** Sergii Skakun (skakun@umd.edu), Department of Geographical Sciences and College of Information Studies (iSchool), University of Maryland, College Park, MD, USA

## Executive summary

The 2030 Agenda for Sustainable Development, adopted by all United Nations Member States in 2015, makes an emphasis on the Sustainable Development Goals (SDGs) and corresponding indicators that characterize complex interactions of human influence and environment state. Many indicators are based on geospatial information, and can be derived from satellite data integrated along with in-situ measurements and models, including weather forecasts, biophysical estimation and classification (machine learning) models. Therefore, there is a need to develop methods and tools that will allow derivation of SDG indicators as geoinformation products.

Nowadays, methodologies for calculation SDG indicators are based on coarse spatial resolution (300 m) satellite data and global products. However, our previous studies for Ukraine showed that global coarse resolution products are not very accurate at local (country) scale, especially for agriculture potential estimation, due to producing mixtures in pixel values from combining different land cover types in a single pixel.

The present proposal aims at developing automated workflows for calculating SDG Indicators grouped by use of land cover maps as the baseline. The main goal is to adopt, improve and apply already proposed methodologies, which were used for generating global products with coarse spatial resolution data, to higher spatial resolution data (up to 10 m), which will be better suited for regional products and applications. Specifically, the proposal will focus on the following indicators: 2.4.1 “Proportion of agricultural area under productive and sustainable agriculture”<sup>1</sup>, 11.3.1 “Ratio of land consumption rate to population growth rate”<sup>2</sup>, 15.1.1 “Forest area as proportion of total land area”<sup>3</sup>, 15.3.1 “Proportion of land that is degraded over total land area”<sup>4</sup>.

Within the project, we will elaborate informational technology for SDGs indicators 15.3.1, 15.1.1, 2.4.1 and 11.3.1 calculations and will implement it in the AWS cloud environment. It will be based on Open Data Cube technology and will include deep learning algorithms for land cover classification, biophysical modeling, weather modeling, and satellite data analysis. The technology will be scalable and usable for any country. As a case study all these indicators will be calculated for 3 countries (which are represented by collaborators): Ukraine, Argentina and India with area ranging respectively from ~604,000 to 3297 thousands km<sup>2</sup>.

The main innovation of the project is concerned with the improvements of existing workflows for SDG indicators calculation by the use of high spatial resolution data and filling gaps between existing global products and national ones.

In particular, we propose to use a previously developed by us neural network approach for high resolution land cover and crop type mapping at country level. The proposed approach outperforms global products in terms of accuracy and spatial resolution. To assess agricultural area under productive and sustainable agriculture, we propose to combine this approach with biophysical model WOFOST as an alternative source of information which allows us to increase the temporal resolution of crop state indicators.

---

<sup>1</sup> <http://www.fao.org/sustainable-development-goals/indicators/241/en/>

<sup>2</sup> <https://unstats.un.org/sdgs/metadata/files/Metadata-11-03-01.pdf>

<sup>3</sup> <http://www.fao.org/sustainable-development-goals/indicators/1511/en/>

<sup>4</sup> <https://knowledge.unccd.int/topics/sustainable-development-goals-sdgs/sdg-indicator-1531>

This proposal is in line with, and will contribute to, the following activities of the GEO Work Plan 2017-2019<sup>5</sup>: GEO WP 1 “GEOSS Architecture and Evolution” (information technologies in Earth Observation domain) and WP 5 “Demonstrations Projects” (demo of use of EO data for SDG Indicators calculation). In particular, according to WP5 the project will demonstrate use of GEOSS infrastructure for providing data to end-users (like national authorities) via common protocols.

Cloud computing technologies are essential for solving this task at country level and in a scalable way — in case of Ukraine a 10 m LCLU map (132874 x 88823) consists of about  $6 \cdot 10^9$  pixels. Creation of such a product for Ukraine requires processing of 7 Tb of Sentinel-1 and Sentinel-2 data annually. Cloud based technology is even more crucial for the Argentina (15 Tb) and India (18 Tb) use-cases due to significantly bigger areas. For these estimations Sentinel-2 data are considered with cloudiness < 20% only. The core technology, which will be tested for 3 pilot countries, will be scalable and applicable for any other regions all over the Globe.

Proper data preprocessing also requires significant computational resources. For SAR data processing workflow includes orbit correction, coregistration, border noise removal, thermal noise removal, radiometric calibration, orthorectification, filtering etc. Optical data require atmospheric correction with cloud and shadows masking. Computational time depends on instance type but in general SAR and optical scene could be processed in 10 and 30 minutes respectively. These actions could be performed in parallel on multiple computational instances.

For efficient satellite data collection, processing and indexing for targeted countries an Open Data Cube<sup>6</sup> is planned to be deployed over the cloud infrastructure. Data Cube will be an useful instrument for users in Ukraine, Argentina and India that will give access to processed satellite products for environmental monitoring purposes.

Biophysical WOFOST model (its Python PCSE implementation) itself execution does not require a lot of computational resources. Nevertheless, one of the required inputs into this model is daily weather data that could be obtained from weather models like WRF<sup>7</sup>. As it is shown in <sup>8</sup> the typical instance for WRF model benchmarking is c4.8xlarge. The benchmark case is 2.5km resolution grid covering the Continental U.S. (7,663,941.7 km<sup>2</sup>). On c4.8xlarge model may be run in 9 hours for a single day.

Project deliverables will be publicly available and will be several-fold:

- Deep learning models based on open-source packages and libraries
- High resolution LU/LC maps for Ukraine, Argentina and India (2000 and 2015 as reference years and ongoing)
- High resolution maps for land productivity for Ukraine, Argentina and India
- SDGs indicators 2.4.1, 15.3.1, 15.1.1, 11.3.1 for Ukraine, Argentina and India
- Open source Python code for SDG indicators calculation

All the code of workflows will be written in Python. These solutions will be used in Ukraine by authorities responsible for implementation of the national reporting (Ministry of Ecology of Ukraine) and by Kyiv city Administration for smart city solutions. Partners from Argentina will be also involved into results use and dissemination within HARVEST

<sup>5</sup> <https://www.earthobservations.org/activity.php?id=120>

<sup>6</sup> <https://www.opendatacube.org>

<sup>7</sup> <https://www.mmm.ucar.edu/weather-research-and-forecasting-model>

<sup>8</sup> <https://github.com/aws-samples/aws-hpc-workshops/blob/master/README-WRF.rst#tutorials>

program. Project results will be used within GEOGLAM initiative. Our partners are interested to be involved as end-users of the solutions developed within this project.

## Project plan

### Methodology for SDGs indicators assessment

According to methodological note<sup>9</sup> 3 Tiers of SDG Indicators are defined:

- **Tier 1:** Indicator is conceptually clear, has an internationally established methodology and standards are available, and data are regularly produced by countries for at least 50 percent of countries and of the population in every region where the indicator is relevant.
- **Tier 2:** Indicator is conceptually clear, has an internationally established methodology and standards are available, but data are not regularly produced by countries.
- **Tier 3:** No internationally established methodology or standards are yet available for the indicator, but methodology/standards are being (or will be) developed or tested.

Within this project we will implement workflows for calculating 4 SDG indicators that highly dependent on satellite data and land cover mapping technologies:

- 2.4.1 “Proportion of agricultural area under productive and sustainable agriculture” (Tier 3)
- 11.3.1 “Ratio of land consumption rate to population growth rate”<sup>10</sup> (Tier 2)
- 15.1.1 “Forest area as proportion of total land area” (Tier 1)
- 15.3.1 “Proportion of land that is degraded over total land area” (Tier 2)

During the project we are going to develop and implement the technology to update them to Tier 1.

The next section describes the state-of-the-art methodology for the indicators identification.

### STATE OF THE ART for SDG indicators calculation

#### Methodology for estimation indicator 15.3.1 “Proportion of land that is degraded over total land area”

##### Level:

**Tier 3 -> Tier 2** (Initial Proposed Tier is 3, Updated Tier Classification is 2 after 6th Inter-agency and Expert Group on SDG Indicators, Finalized methodology development with metadata is in progress)

##### Principle:

---

<sup>9</sup>

[https://unstats.un.org/sdgs/files/Tier%20Classification%20of%20SDG%20Indicators\\_11%20May%202018\\_web.pdf](https://unstats.un.org/sdgs/files/Tier%20Classification%20of%20SDG%20Indicators_11%20May%202018_web.pdf)

<sup>10</sup> For cities with at least 100 thousands of inhabitants as it is organized in Urban Atlas product over EU - <https://land.copernicus.eu/local/urban-atlas>

The SDG indicator 15.3.1 is a proportion of degraded land to the total area of the country and based on the analysis of available data or developed at the national level products (namely, Trends in Land Cover, Land Productivity and Carbon Stocks). This indicator is based on statistical principal “One Out, All Out” on evaluation of changes in the sub-indicators<sup>11</sup>. This principle means that we have three types of changes in the sub-indicators, which are depicted as positive or improving, negative or declining and sustainable or unchanging. This methodology unfortunately can’t be used in full mode because it is very hard to obtain Carbon Stocks maps for every year at national level. The newest Carbon Stocks map is dated by 2016 year and already outdated as well as the newest open access land cover map dated by 2015.

If one of the sub-indicators has negative changes for some area, then this area has negative productivity. According to existent methodology as negative changes considered following transitions: decrease of carbon stock level over the period of time, decline in land productivity or negative land cover changes (i.e. forest -> grassland, forest -> cropland, any green area -> urban) etc.

**Data:**

Dataset type	Source	Specs	Comments
Land Cover	ESA CCI land cover dataset	300 m resolution, 1992-2015, 22 classes	Rescaled from 22 classes to 6 main (Forest, Grassland, Cropland, Artificial, Wetland and Water, Bare land). Basic period 2000-2015 (UNCCD)
Land Productivity Dynamics	JRC Productivity Dynamics Dataset	1 km resolution, 5 classes	Based on time series of SPOT-Vegetation data collected during 1999-2013. 5 qualitative classes of land productivity trends are available (Declining productivity, Early signs of decline, Stable, but stressed, Stable, not stressed, Increasing productivity)
Soil Organic Carbon stocks	ISRIC’s SoilGrids250m <sup>12</sup>	250m	Topsoil SOC values (0-30 cm)

Actually, the approach implemented by JRC takes into account total green vegetation and non-suitable for estimation of cropland productivity (the most productive regions often partially covered by forests).

Furthermore, coarse resolution of this global product produces mixtures in pixel values from combining different land cover types in single pixel.

<sup>11</sup> <https://unstats.un.org/sdgs/metadata/files/Metadata-15-03-01.pdf>

<sup>12</sup> <https://www.isric.org/>

## **Methodology for estimation indicator 15.1.1 “Forest area as proportion of total land area”**

### **Level - Tier 1**

The SDG indicator “15.1.1” is a proportion of forest areas to total land area. According to existing methodology FAO has been collecting and analyzing data on forest area since 1946<sup>13</sup>. It was done at intervals of 5-10 years as part of the Global Forest Resources Assessment (FRA)<sup>14</sup>. FRA 2015 contains information for 234 countries and territories on more than 100 variables related to the extent of forests, their conditions, uses and values for five points in time: 1990, 2000, 2005, 2010 and 2015. Assessment of forest area is carried out at infrequent intervals in many countries. Access to remote sensing imagery has improved in recent years.

Another globally available product is Global Forest Change produced by University of Maryland with use of time-series analysis of Landsat images in characterizing global forest extent and change from 2000 through 2017 at 30 m resolution<sup>15</sup>.

So, the main required information for indicators is time series of land cover maps over the territory of interest.

## **Methodology for estimation indicator 11.3.1 “Ratio of land consumption rate to population growth rate”**

### **Level - Tier 2**

The SDG indicator 11.3.1 can be calculated using detailed land cover maps built on moderate and high resolution images for two years at least to estimate ratio of land consumption rate and requires open access statistics about city population. Population statistics over the city provided by statistical authorities is used for population growth rate assessment. In case when national data are not available, it is possible also to use open data such as JRC Global Human Settlement layer (<https://ghsl.jrc.ec.europa.eu/datasets.php>).

This indicator can be used in every country in the world with use of global coarse resolution LC. Information can be obtained at two levels. First level is city level indicator, which measure sustainability in terms of city and population growing. The second level is country based, which aggregate city’s based indicators within the country and indicate urban area and population growth rate for this country.

## **Methodology for estimation indicator 2.4.1 “Proportion of agricultural area under productive and sustainable agriculture”**

### **Level - Tier 3**

**This workflow include the definition of land productivity in particular over cropland.** The methodology implemented by JRC takes into account total green vegetation and non-suitable for estimation of cropland productivity (the most productive regions often covered with forests). Furthermore, coarse resolution of this global product produces mixtures in pixel values from combining different land cover types in single pixel.

<sup>13</sup> <https://unstats.un.org/sdgs/metadata/files/Metadata-15-01-01.pdf>

<sup>14</sup> <http://www.fao.org/forest-resources-assessment/en/>

<sup>15</sup> <https://earthenginepartners.appspot.com/science-2013-global-forest>

The SDG indicator 2.4.1 could be calculated using the same methodology as proposed for indicator 15.3.1 calculation. This indicator is a proportion of agricultural area that has a positive productivity trend value to the total agricultural area by the rule “One Out, All Out”. For this indicator, the same sub-indicators as for indicator 15.3.1 are used, but the area of interest is not the whole area of the country, but rather the agricultural land (cropland) subsetted with use of LC map. For this indicator, the use of high spatial resolution satellite images is particularly important, since mixed pixels greatly affect the value of sub-indicator changes.

### **Data**

Dealing with land cover mapping and SDG calculation workflows at the country level required long term series of satellite data.

Right now, methodology for all indicators identification is based on low resolution data (mostly 300 m or lower) that has a mixture of different land use classes over one pixel.

Our main innovation is use of time series of high resolution data (10 m) from Sentinel-1 and Sentinel-2, which will be directly acquired from Amazon S3.

In particular, Sentinel-1 provides approximately 3 Tb of images for the territory of Ukraine, 5 Tb of images for the territory of Argentina and 6 Tb for the territory of India for one vegetation year. And Sentinel-2 provides approximately 4 Tb of images for the territory of Ukraine, 10 Tb of images for the territory of Argentina and 12 Tb of images for the territory of India with cloudiness less than 20% for one vegetation year.

- Storing and transferring big amounts of satellite data couldn't be solved in efficient way without cloud storage, in particular AWS S3.

### **Our innovations in SDGs indicator assessment workflow**

We propose a universal workflow (Fig. 1) for calculating of all the above mentioned indicators based on satellite data for precise LC/LU classification and using modeling data (weather forecast and biophysical crop growth models) for crop state assessment.

We propose to use our own neural network approach for high resolution land cover mapping at country level (Fig. 2). Proposed approach<sup>16</sup> outperforms global products in terms of accuracy and spatial resolution. For instance, for Kyiv region of Ukraine the overall accuracy is improved by more than 10% compared to ESA's Climate Change Initiative Land Cover dataset; the kappa coefficient for ESA's Climate Change Initiative Land Cover dataset is 0.75, while the kappa coefficient for our map is 0.9. Concerning ESA CCI land cover we have identified significant overestimation of cropland areas (Fig 3.)

---

<sup>16</sup> Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., & Skakun, S. (2017). Exploring Google earth engine platform for Big Data Processing: Classification of multi-temporal satellite imagery for crop mapping. *frontiers in Earth Science*, 5, 17.



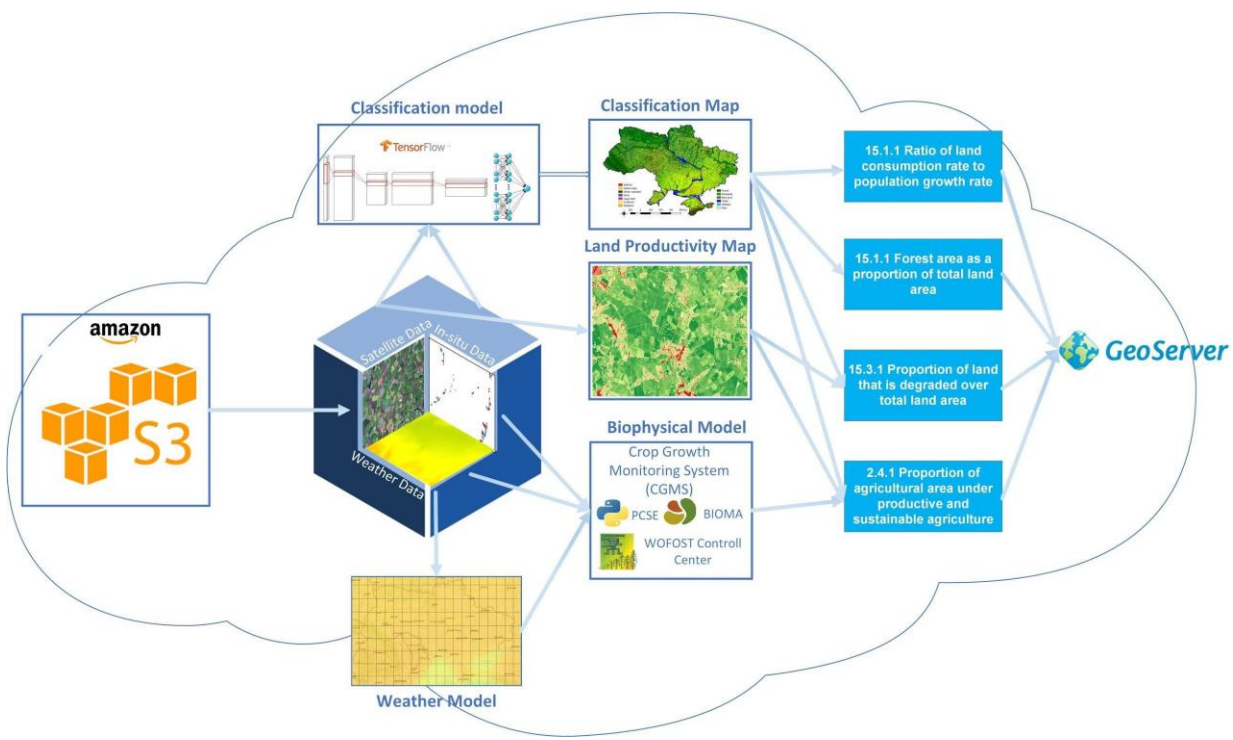


Figure 1. Workflow for calculating Sustainable Development Goals indicators 11.3.1, 15.1.1, 15.3.1 and 2.4.1 within cloud Data Cube technology

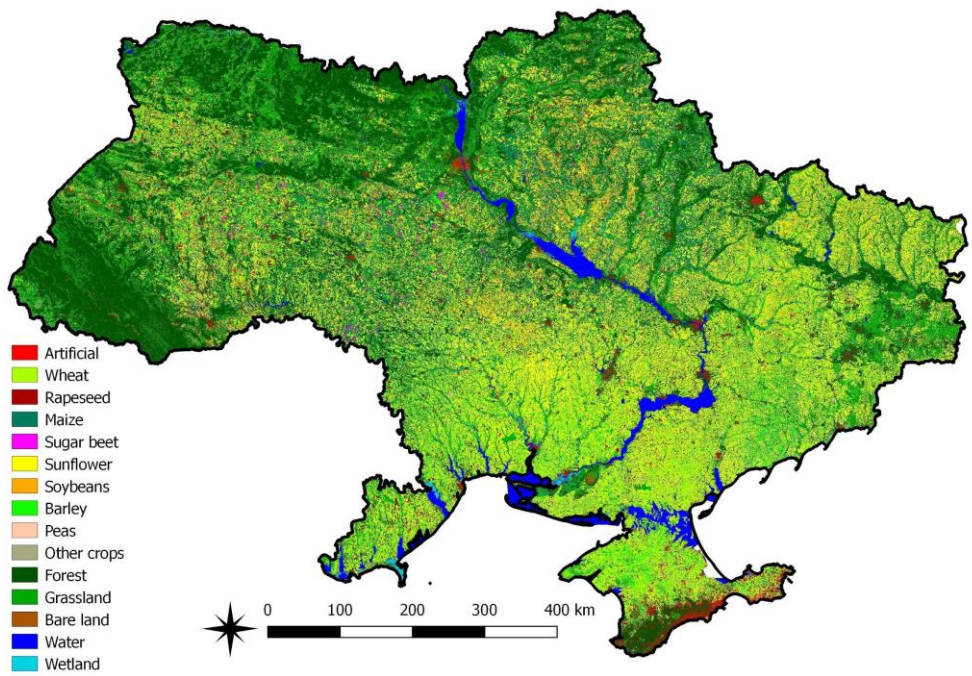


Figure 2. Land cover for Ukraine, 2018, 10 m



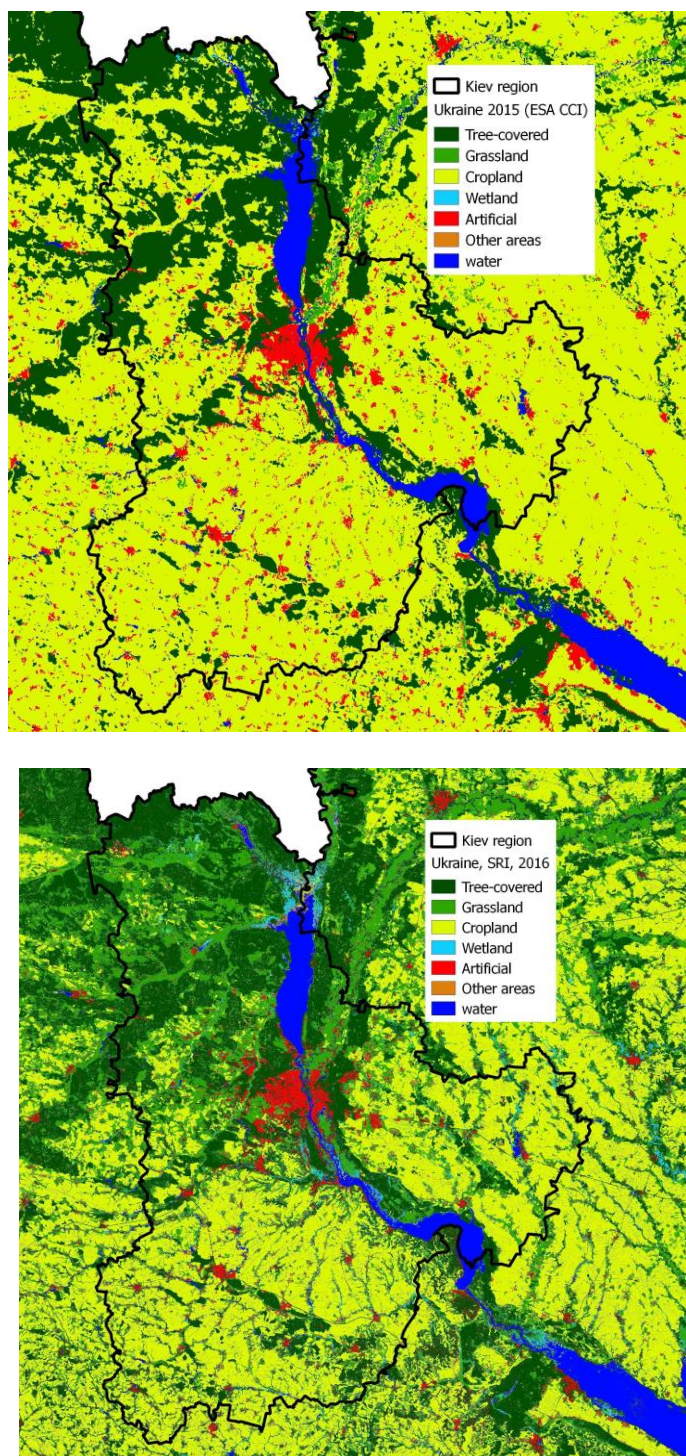


Figure 3. Comparison of ESA CCI land cover (top) with SRI land cover (bottom) over Kyiv region

We will improve land productivity estimation workflow in following ways: (i) high resolution satellite data (sparse time resolution could be a problem for regions with big amount of cloudy days); (ii) Leaf Area Index from biophysical WOFOST model launched

over the uniform grid that incorporates modeled weather data (WRF model, weather data generators etc.).

**Workflow summary** (covered in details in cloud computing technology section) – Fig. 1:

- Satellite data will be taken from Amazon S3 storage and preprocessed with use of Open Data Cube software deployed at Amazon EC2;
- Preprocessed imagery will be used by classification model for producing land cover maps at Amazon EC2;
- Weather model will be deployed at Amazon EC2 and treated as source of information for gridded version of crop growth model;
- Time series of vegetation indices will be extracted from high resolution satellite imagery and cumulative trend analysis will be performed for land productivity assessment (crop specific) at Amazon EC2;
- Fusion of satellite based crop vegetation indices estimates and crop growth modeling allows to estimate crop specific land productivity for main crop types;
- Results are published on GeoServer<sup>17</sup>

Indicators 11.3.1 and 15.1.1 rely only on LC/LU, while 15.3.1 and 2.4.1 additionally on land productivity data and crop state assessment.

## **Our methodology of land cover mapping with machine learning approach**

Proposed methodology is based on a deep learning approach for land cover mapping, in particular an ensemble of neural networks<sup>18</sup>. A committee of neural networks is used for providing crop classification and land cover maps for the territory of interest (starting from whole Ukraine use-case) using high resolution Sentinel-1 and Sentinel-2 imagery and appropriate in-situ data. Time series of satellite data for vegetation period allow achieve better accuracy of land cover classification and could help us more precise discriminate crops and other land cover types. But utilizing time series of high resolution satellite imagery and it's preprocessing is time consuming task. Another issue is training of deep learning models on time series of satellite data.

To address aforementioned big geospatial data challenges two main powerful cloud platforms are available at the moment: Amazon and Google. We have previous experience with utilizing powerful cloud platforms for land cover classification such as Amazon Web Service (AWS)<sup>19</sup> and freely available Google Earth Engine (GEE)<sup>20</sup>.

<sup>17</sup> <http://geoserver.org/>

<sup>18</sup> N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 5, pp. 778-782, 2017.

<sup>19</sup> Shelestov, A., Lavreniuk, M., Kolotii, A., Vasiliev, V., Shumilo, L., & Kussul, N. Cloud approach to automated crop classification using Sentinel-1 imagery. Proc. of the 2017 conference on Big Data from Space (BiDS'17). 122-125.

<sup>20</sup> Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., & Skakun, S. (2017). Exploring Google earth engine platform for Big Data Processing: Classification of multi-temporal satellite imagery for crop mapping. *frontiers in Earth Science*, 5, 17.

Cloud platforms allow us to overcome challenges of satellite data download and processing. GEE platform that provides built-in functions and intrinsically-parallel computational access. At the same time, Amazon cloud platform provides an opportunity for exploiting any software for image processing and libraries with advanced classifiers. As well as Amazon does not provide ready to use functionality of satellite data processing and classification, it's utilization requires much more efforts. We have developed an automated workflow for cloud-based crop classification using AWS based on Sentinel-1 and Sentinel-2 imagery. This methodology has been applied for providing 10 m resolution crop classification maps for the Ukraine territory in 2016 - 2018. This improvement provide much better accuracy in indicators estimation. Collaborators will be enable to collect and share in-situ data for Argentina and India that are essential for crop classification mapping for these countries. Also, collaborators from the experienced organization that are working on study area (Argentina and India) will assist with the providing interpretation and final product validation.

## **Our methodology for crop productivity assessment**

### **Satellite based approach**

Satellite based approach for estimation of indicator 2.4.1 that we propose is based on assessment of Vegetation indexes change trends from year to year and corresponding slope definition (ascending, descending or stable). Currently we use NDVI index, but due to value saturation during vegetation period, we are going to use more complex indices such as EVI and LAI. To deal with cloudiness of optical data fusion of Landsat-8 (30 m) and Sentinel-2 (10 m) data is necessary with coregistration of products from different satellites and values scaling. This solution is intended to replace global land productivity products within workflow of 15.3.1 indicator calculation.

### **WOFOST based approach**

The second approach that we propose for indicator 2.4.1 estimation is based on more complicated calculations. It based on LAI extracted from crop growth model and satellite data. The idea is compare forecasted LAI under certain conditions without land degradation influence (ideal case) and close to real LAI obtained from satellite data in the same conditions but with influence of land degradation.

Ideal case LAI is estimated with use of WOFOST model (gridded CGMS system) that uses as input agrometeorology data, crop type parameters, soil type information and standard agromanagement rules. As model output for specific location over the regular country-wide grid it's possible to obtain LAI for several crop types that correspond to meteorology and soil type.

Intersection of crop type map with WOFOST model output allows us to obtain crop specific LAI time series with high temporal resolution (daily) and 10 meters spatial resolution (all fields of the same crop within one WOFOST model grid cell have the same LAI value). Grid cell size depends on weather and soil data resolution.

To estimate satellite based LAI we plan to use harmonized Sentinel-2 and Landsat-8 data. This approach provides the possibility to built harmonized LAI time series with higher temporal resolution and up-to 10 meters spatial resolution. Satellite LAI product shows not only crop condition related to meteorology, crop type and soil type but also

takes into account (in indirect way) land degradation as factor of land productivity decreasing and proper land management as factor for land productivity increasing.

Dealing with time series of modeled LAI and satellite based LAI estimates we'll build two land productivity maps. For both maps LAI time series will be processed in similar way - for each pixel land productivity index equal to area under LAI development curve during vegetation period. For identification of land state (degradation or sustainable) it's enough to subtract two productivity index maps and cluster result into three groups by pixel value. Land should be considered as degraded if subtraction result is less than zero, as sustainable if result is equal to zero and as productive if result is higher than zero.

This method' benefits include:

- possibility for land degradation estimation on one year time series (in case of enough amount of cloud-free images)
- We don't pretend on calculation of land productivity that has quite a fuzzy interpretation, but instead operate qualitative land productivity index that express land productivity
- In proposed methodology there is no need to calculate vegetation indices trend that is the main problem of state-of-the art approach. This problem expressed by necessity of vegetation index curve normalization with respect to crop type for correct trend estimation.
- The outcome of this approach could also be used within Water-Food-Energy NEXUS for Essential Variables (EV's) estimation. These EV's provide possibility for environmental monitoring and analysis.

## **Cloud computing justification**

Amazon Cloud Services provide unique opportunity to process Big amounts of data in quick way (in our case total annual satellite data coverage for these countries is up to 15 Tb of data from Sentinel missions only). In some cases, actually no time is spent on data download.

Classification approach that is the main component of SDG indicators workflows is based on time series of satellite data thus it requires a lot of storage and computational resources for satellite data processing.

As a part for more complicated workflow for SDG indicator 2.4.1 calculation crop growth modelling over the regular grid will be used. Crop growth models require daily weather data with high enough spatial resolution (this is even more crucial for precipitation data). Weather modelling with use of WRF model will be implemented over Amazon EC2 instance.

### **Data Access**

Registry of Open Data on AWS include numerous datasets. Among them we plan to use Sentinel-1 (ASF S3 bucket), Sentinel-2 (requested pay bucket - s3://sentinel-inventory/sentinel-s2-11c for L1C products and s3://sentinel-s2-12a for L2A level products) and Landsat-8 (s3://landsat-pds/c1/L8/). These datasets are stored by Amazon at scalable storage S3 and could be obtained via high-speed network connections.

In particular, SAR data (Sentinel-1) on Amazon computing facilities are available from Alaska Satellite Facility <https://www.asf.alaska.edu/>. These resources provide data portal [<https://vertex.daac.asf.alaska.edu/>] for data search and download with use of web-interface or API.

### Satellite data preprocessing

For land cover mapping that is a baseline product for SDG indicators estimation time series of processed satellite data are required. Most of satellite data preprocessing stages require a lot of computational resources.

For SAR data processing workflow includes numerous stages (orbit correction, border noise removal, thermal noise removal, radiometric calibration, orthorectification, filtering etc.). Computational time depends on instance type but in general 1 scene could be processed in 10 minutes.

Optical data require atmospheric correction with cloud and shadows masking. In general, about 30 minutes is required for processing of one Sentinel-2 scene with use of sen2cor software.

To extract maps for big territories we need coregistration for time series of satellite data (between images from different sensors and between images in one time serie). We will do this operation with GDAL function `gdal_merge.py`<sup>21</sup>, which is very time consuming.

The problems with processing of Big amount of satellite data could be solved in efficient way with scalable computational infrastructure (like Amazon EC2). This could be ensured by parallel data processing on numerous computational instances (paradigm of “*Data parallelism*”).

### Weather data modelling

As the source of data for enhanced workflow for SDG 2.4.1 WRF model<sup>22</sup> is planned to be used. WRF is a numerical weather prediction system designed for both atmospheric research and operational forecasting applications. WRF model doesn't require in-situ data for launch, however publicly available in-situ weather data (for instance from Global Surface Summary of the Day from NOAA<sup>23</sup>) will be used for model output validation.

Due to complicated modeling process WRF model requires a lot of computational resources and cluster for running WRF model. As it is shown in <sup>24</sup> the typical instance for WRF model benchmarking is c4.8xlarge (with 36 vCPUs and 60 Gb RAM). The benchmark case is 2.5km resolution grid covering the Continental U.S. (7,663,941.7 km<sup>2</sup>). On c4.8xlarge model may be run in 9 hours from cold start for single day.

In case of Ukraine we earlier had an experience in WRF model launch over GRID infrastructure. Forecasts were constructed on a grid of size 200x200 with the mesh size equal to 10x10 km. Detailed background is available via link:

- Kravchenko, A. N., Kussul, N. N., Lupian, E. A., Savorsky, V. P., Hluchy, L., & Shelestov, A. Y. (2008). Water resource quality monitoring using heterogeneous

<sup>21</sup> <https://gdal.org/>

<sup>22</sup> <https://www.mmm.ucar.edu/weather-research-and-forecasting-model>

<sup>23</sup> <https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00516>

<sup>24</sup> <https://github.com/aws-samples/aws-hpc-workshops/blob/master/README-WRF.rst#tutorials>



data and high-performance computations. *Cybernetics and Systems Analysis*, 44(4), 616-624.

### Data cube operations

As a part of cloud-based solution we plan to implement data processing with Open Data Cube (ODC). The Committee on Earth Observation Satellites (CEOS) founded ODC initiative to provide a data architecture solution that has value to global users and to increase the impact of EO satellite data.

We plan to use ODC for range of applications including land, water, cloud, and time series analysis. Applications for mosaic creation, Spectral Index calculation, Water Mapping, Land Classification and Land Change are available after system deployment and we plan to use them as input for land cover mapping. Earlier we deployed ODC on local computational resources.

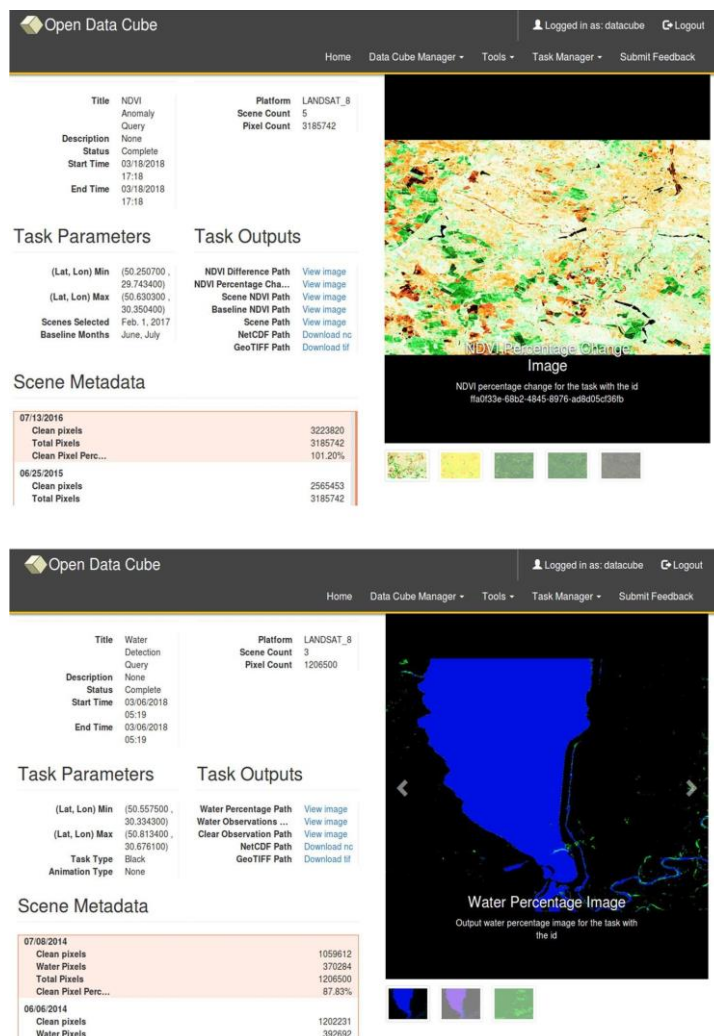


Figure 3. ODC sample deployment

## Deliverables





Scalable deep learning land cover mapping approach implementation (Argentina and India use-cases)													
Implementation of workflow for SDG indicator 2.4.1 (simplified version), 11.3.1, 15.1.1 and 15.3.1 for 3 use-cases													
Deployment of WRF model over Amazon Cloud Infrastructure													
Implementation of complex workflow for SDG indicator 2.4.1 calculation													

### Project team

Role	Focal Point	Institution	Tasks description
PI	Nataliia Kussul	Space Research Institute NASU-SSAU, Ukraine	Coordinator, scientific leader, cloud technology development
CO-PI	Esteban Julian Copati <sup>25 26</sup>	Buenos Aires Grains Exchange, Argentina	In-situ data collection in Argentina, local expert from Argentina, results validation for Argentina
Collaborator	Nilanchal Patel <sup>27 28</sup>	Birla Institute of Technology, Mesra, India	In-situ data collection in India, local expert from India, results validation for India
Collaborator	Michael Ryabokon <sup>29</sup>	NGO Association Noosphere (Noosphere Engineering school)	In-situ data collection in Ukraine, results validation for Ukraine
Collaborator	Sergii Skakun <sup>30 31 32</sup>	University of Maryland, College Park, MD, USA	Scientific advisor in data processing and harmonization,

<sup>25</sup> <https://ar.linkedin.com/in/esteban-j-copati-4271882a>

<sup>26</sup> <https://nasaharvest.org/partner/esteban-j-copati>

<sup>27</sup> [https://www.researchgate.net/profile/Nilanchal\\_Patel](https://www.researchgate.net/profile/Nilanchal_Patel)

<sup>28</sup> <https://in.linkedin.com/in/nilanchal-patel-7401b975>

<sup>29</sup> <https://www.linkedin.com/in/mryabokon/>

<sup>30</sup> <http://lcluc.umd.edu/people/sergii-skakun>

<sup>31</sup> [https://www.researchgate.net/profile/Sergii\\_Skakun](https://www.researchgate.net/profile/Sergii_Skakun)

			crop yield forecasting
--	--	--	------------------------

The main power of the project is connection of high-performance cloud computations with modern land cover and crop mapping techniques, **SRI** has strong expertise in both domains. Within the project **SRI** will perform all main tasks of LAI and crop grows modeling, satellite data processing, classification methods development and technologies deployment (Open Data Cube, CGMS and WRF) with their validation for the territory of Ukraine. After technology will be successfully tested for the territory of Ukraine it will be extended for the territory of India and Argentina.

**CO-PI from Argentina (Esteban J. Copati)** is a member of GeoGlam (Oct-2015), NASAHarvest (Sept-2017) and AmeriGeoss (Sept-2018). He has a remarkable experience in collection of georeferenced in-situ data that is crucial for successful crop type / land cover mapping and validation. Additionally he has wide experience in crop conditions assessment and collaboration in multinational projects. He'll be responsible for collection of in-situ data and product validation for the territory of Argentina as well as for communication with local end users and authorities.

**Our local collaborator from Ukraine (NGO Noosphere)** will participate in in-situ data collection and preparation for the territory of Ukraine for land cover and crop mapping. They also will be involved into product validation based on data from publicly available sources (statistics, online surveys etc.), project results dissemination and popularization over the internet. Noosphere will prepare projects results usage manuals for end-users. Noosphere will organize webinars for end-users (in particular India and Argentina authorities) to make them capable for usage of developed technologies. In particular, training on satellite products interpretation for agriculture domain, basic cloud technologies exploitation, SDG indicators interpretation for decision making will be organized.

**Our collaborator from USA (Sergii Skakun)** is a member of NASAHarvest program. He has a remarkable expertise of satellite data usage and will develop a procedure for Sentinel-2 and Landsat-8 data coregistration and values harmonization for vegetation development assessment from NDVI time series with use of both satellites jointly.

**Our collaborator from India (Nilanchal Patel)** is a Prof. of Birla Institute of Technology. Ukrainian part of project team will provide to him necessary knowledge so he'll be capable to provide these knowledge to his students. This is really important for capacity building in India. Also he will be responsible for collection of in-situ data and product validation for the territory of India as well as for communication with local end users and authorities.