EARTH OBSERVATION DATA AND SPATIAL DATA SETS ANALYSIS

Pavel Simek, Jan Jarolimek, Eva Kanska, Michal Stoces, Jiri Vanek, Jan Pavlik and Alexandr Vasilenko

Department of Information Technologies, Faculty of Economics and Management, Czech University of Life Sciences Prague

simek@pef.czu.cz, jarolimek@pef.czu.cz, kanska@pef.czu.cz, stoces@pef.czu.cz, vanek@pef.czu.cz, pavlikjan@pef.czu.cz, vasilenko@pef.czu.cz

ABSTRACT

Aerial or satellite imagery allows for non-destructive remote sensing and monitoring in agriculture and related fields. The advantage of data from Sentinel missions is their availability and regular acquisition (approximately every 3 days in the Czech Republic). For effective remote sensing, there are often some limitations resulting from the properties of the acquired data – large data transmission (especially in the case of mobile use), resolution, cloudiness, sunlight reflections, etc. The research was focused on data sets from SENTINEL-2 Level-2A and higher, which are already preprocessed, especially in cartographic projection (UTM / WGS1984), creating tiles 100 x 100 km with atmospheric corrections with subsequent possibility to calculate vegetation indices, especially NDVI, GNDVI, LAI, EVI, RENDVI and MSI. Due to the size of the data, the usable datasets (area of interest and without clouds) are divided into a 5x5 km tile network grid, calculated vegetation indices and saved in datastorage. The main goal is to propose a solution for the calculation of indexes from smaller datasets, to design a prototype and to subsequently verify the solution on a pilot.

Keywords*:* Spatial data, Earth observation, remote sensing, Sentinel

1. INTRODUCTION

Aerial or satellite imagery enables non-destructive remote sensing and monitoring in agriculture and related fields, such as monitoring the condition of agricultural and forest land, water in the landscape, rural development, and offering operational use for modeling agri-technical interventions (Fülöp et al., 2015). Remote sensing of optical and radar data can help map out crop types and estimate biophysical parameters, especially with the availability of an unprecedented amount of free SENTINEL satellite data within the Copernicus program (Veloso et al., 2017). Especially the SENTINEL-2 multispectral imaging is of great importance to agriculture (Wang et al., 2018).

The use of SENTINEL and Landsat satellites can be divided into use in crop and livestock production. However, these two directions are very closely linked. Images can be used for modeling and calculating spectral indexes, and they can just as well be used for visualization in the form of color synthesis, whether in the visible spectrum or in other wavelength ranges. Longer mapping of land use / cover can also be automated, for example in the field of monitoring forested areas (Szostak et al., 2019). Monitoring of vegetation cover in the winter months is also of great importance (Denize et al., 2019). Before processing, satellite images must be pre-processed using atmospheric corrections, resampling, and spatial cropping (Rodriguez-Ramirez et al., 2019). Also, many projects that tried to integrate various data of various formats were implemented recently (Řezník et al., 2015).

A suitably chosen color synthesis may indicate plant stress or soil erosion by a different color shade than is normal for healthy and prosperous growth. The most recently usable indices in plant production include, in particular, the Normalized Difference Vegetation Index - NDVI, Enhanced Vegetation Index - EVI (Halabuk et al., 2015), but also the Red Edge Normalized Difference Vegetation Index - RENDVI (Deng et al., 2018), Green Normalized Difference Vegetation Index - GNDVI, Moisture Stress Index – MSI or Leaf Area Index - LAI.

The original images of SENTINEL missions may be too large or even unusable to process and calculate the indexes (e.g. due to clouds). The main goal is to propose a solution for the calculation of indexes from smaller datasets, to design a prototype and subsequently to verify the solution on a pilot.

2. Methodology

The basic scientific method of analysis and synthesis was used to tackle the research problem. Sentinel satellites are made up of several types of satellites, usually discharged in pairs to provide faster data recovery. Currently, the satellites of these missions are two SENTINEL-1 satellites designed for radar mapping of the Earth's surface and the detection of information available through their sensors, two SENTINEL-2 satellites that provide image data across multiple spectral bands, two SENTINEL-3 satellites and one SENTINEL-5P satellite.

The research was focused mainly on data from the SENTINEL-2 mission. To ensure potential use of SENTINEL-2 data in time series, data must be geometrically registered and radiometrically corrected. Both procedures are the basic operation in Level-1 data level processing. Level 2 includes atmospheric corrections. Level-3 processing creates space-time synthesis from “Bottom of Atmosphere” images of SENTINEL-2 Level-2A. One important step is resampling images into a cartographic projection (UTM / WGS1984), creating 100 x 100 km tiles, and calculating cloud masks, land surface and water surfaces (Level-1C).

A SENTINEL-2 product refers to a directory folder that contains a collection of information. It includes: a manifest.safe file which holds the general product information in XML, a preview image in JPEG2000 format, subfolders for measurement datasets including image data (granules/tiles) in GML-JPEG2000 format, subfolders for datastrip level information, a subfolder with auxiliary data (e.g. International Earth Rotation & Reference Systems (IERS) bulletin) and HTML previews.

The Level-2 product is also in SAFE format, which groups together several types of file: metadata file (XML file), preview image (JPEG2000 with GML geo-location), tiles files with BOA reflectances image data file (GML / JPEG2000) for each tile, datastrip files, auxiliary data and ancillary data (Ground Image Processing Parameters (GIPPs)) (ESA).

Default dataset name from the SENTINEL-2 mission looks like:

```
MMM_MSIL1C_YYYYMMDDHHMMSS_Nxxyy_ROOO_Txxxxx_<Product Discriminator>.SAFE
```

where:

- MMM: is the mission ID (S2A/S2B)
- MSIL1C: denotes the Level-1C product level
- YYYYMMDDHHMMSS: the data sensing start time
- Nxxyy: is the Processing Baseline number (e.g. N0204)
- ROOO: is the Relative Orbit number (R001 - R143)
- Txxxxx: is the Tile Number
- SAFE: Product Format (Standard Archive Format for Europe) (ESA)

The processing of remote sensing image data and the interpretation of image information involves several sequential digital image analysis procedures. Lillesand et al. (2008) describes several basic steps in digital image processing; for the purposes of this research, the following are the most important:

- Image Preprocessing - targets image correction for the most accurate representation of the original scene. Includes correction of geometric, radiometric and atmospheric distortion and image noise. Preprocessing procedures depend on the characteristics of the sensor used to capture the image.
- Image Enhancement - represents methods of effective image representation for its further processing or visual interpretation. This creates a new image with increased information content. It involves manipulation of the color component of the image or contrast, or, for example, the use of spectral bands to highlight vegetation in the form of vegetation indices, etc.

For the pilot Image Enhancement software for downloading selected images of SENTINEL-2 Level-2 mission and the subsequent calculation of indexes were written in C ++. For the processing itself a high performance computer with two processors AMD EPYC 7601 32 core, 2.2 GHz and 1.4 TB RAM, OS Windows 10 was used.

3. RESULTS

Sentinel mission data are effectively used in the agrarian sector. Each data set represents the image data of a surface sub-area of the Earth. However, it is necessary to reckon with a large production of data, especially during their continuous renewal. The data are available through the Sentinel organizations' web presentations. The basic idea is slightly different, given by local customs and view of open data, but the data are in accessible form that allows automatic download and subsequent machine processing. All datasets are available via https://scihub.copernicus.eu/.

The current state of calculations within existing datasets assumes work with selected sectors recorded by the SENTINEL-2 scan tool. The tile, which is then available at the Science Open Hub, has the dimensions of 100 x 100 km, which requires considerable computing. The way to avoid this is to divide the tile into sub-units (making a grid of smaller squares). Due to the optimum ratio of computational complexity and size of the displayed area, the tiles should have a size of 5 x 5 km.

Each tile has even partial addressing - the columns are marked with letters and lines with numbers as a grid. Thus, the grid will have the origin A1 (the first column marked as A and the first row marked as 1) and the end T20 (the last 20th column marked as T and the last 20th row marked as 20). For addressing was used ASCII characters. This makes it possible to use the current API, which is based on string search in the name of data sets. Before that, however, it is still necessary to perform a cloud analysis to ascertain whether the image is usable or not, and the source data is recalculated to a pixel resolution of 10 m. Subsequently, the tiles that are meaningful for the calculation of vegetation indices are selected, e.g. without clouds, correct location (in the Czech Republic), and so on.

Figure 1. The procedure of calculating indexes (starting from Open Access data download)

Accessing the Copernicus repository at national level is often accomplished with the same software as the ESA Open Access Hub. When using a national repository, it is not possible to influence selections and work with data other than displaying currently published datasets. These can be displayed based on the selected satellite, selected tool (sensor), time stamp and polygon (location of location within data sets).

Among the most widely used spectral indices in crop production are: NDVI, GNDVI, LAI, EVI, RENDVI and MSI. For index calculations, new datasets are uploaded to the repository that respect the following naming rules:

Original (default) dataset name:

`S2B_MSIL2A_20190330T101029_N0211_R022_T33UVR_20190330T144328`

New dataset name:

`T1_S2B_MSIL2A_20190330T101029_N0211_R022_T33UVR_20190330T144328`

where the identification of the grid tile of the dataset was added - here the square T1 - the twentieth column and the first row

After calculating the index, its name is reflected in the dataset naming, e.g.

`NDVI_T1_S2B_MSIL2A_20190330T101029_N0211_R022_T33UVR_20190330T144328`

As can be seen from the example, the name reflects that it is NDVI from T1 tile (5x5 km) inside the 33UVR tile (100x100 km).

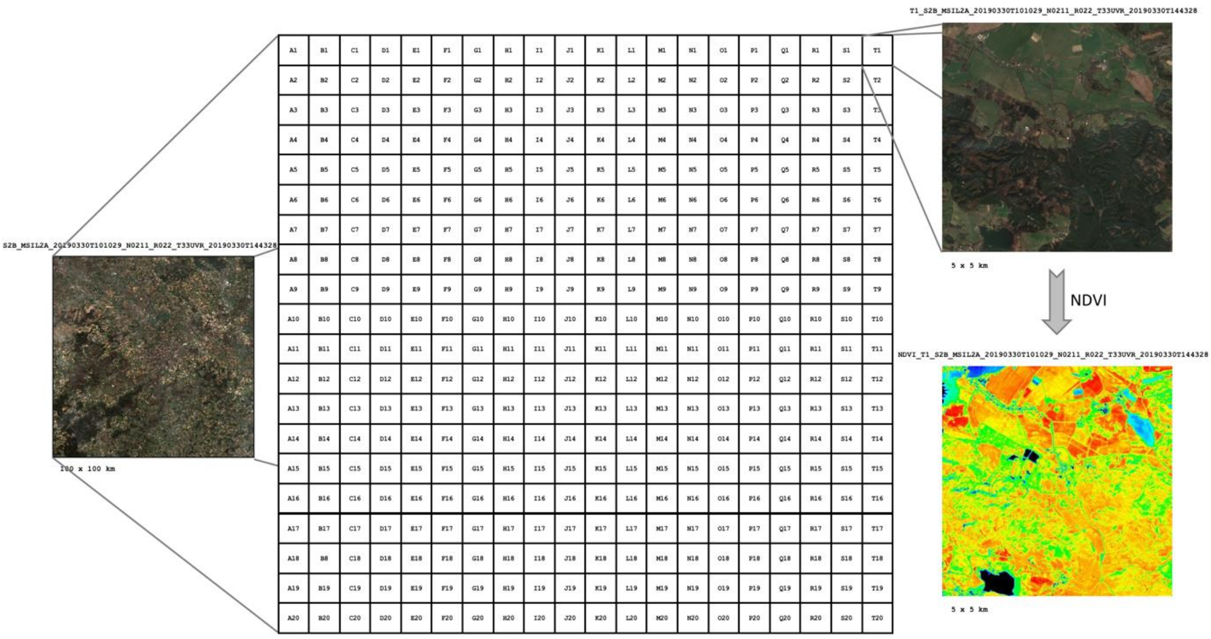

Figure 2. Layout of 100x100 km frame divided into 5x5 km tiles

4. DISCUSSION

The advantage of data from SENTINEL-2 is their availability and regular acquisition (approximately every 3 days in the Czech Republic). To work with remote sensing data, there are often some limitations that result from the properties of the acquired data – large data transmission (especially in the case of mobile use), resolution, cloudiness, sunlight reflections, etc. For the analysis of cloudiness, it is necessary to create a cloudiness mask from SENTINEL-2 data. This is already provided at level 1C, and is subject to validation (Baetens et al., 2019) but even then, cloud classification can be difficult (Lie et al., 2019).

Figure 3. Example of cloud mask cirrus, source: ESA, 2019.

Another source for the calculation of vegetation indices can be data sets from the Landsat-8 program, or it is possible to combine both data sets, which can be beneficial, for example, in longer term forest monitoring (Arekhi et al., 2019) or landscape coverage (Carrasco et al., 2019).

The big limit is the addition of indices requiring heavy calculations that cannot be realized shortly after the release of the new dataset. These are mainly indices that require resampling. Thanks to the current cloud computing capabilities, this limit can be successfully overcome using?? applications for processing remote sensing images significantly reduce processing time (Defourny et al., 2019).

Images for processing and indices computing can also be captured with UAV. In this case, the images are at a much higher resolution, but at a much higher price. However, for certain crops, SENTINEL-2 images produce some distortion (Khaliq et al., 2019).

5. Conclusions

The advantage of data sets from the SENTINEL-2 mission of the Copernicus program is their availability, regularity of new images and their preprocessing into cartographic projection (UTM / WGS1984), creation of tiles 100 x 100 km, calculation of cloud mask, earth surface and water surfaces and atmospheric corrections (Level-2A). The disadvantage can be its spatial resolution, size of files and especially cloudiness.

The research was focused on the possibility of automated SENTINEL-2 data usage, with subsequent possibility of calculation of vegetation indices and other indices, especially NDVI, GNDVI, LAI, EVI, RENDVI and MSI. Due to the size of the data, the usable datasets (the area of interest) are divided into a 5x5 km grid network labeled A1 - T20 (400 tiles from each 100x100 km grid square). Before the calculation of the vegetative indices, the area to be examined will be selected and the cloud will be analyzed (whether the calculation from the given tile makes sense). Subsequently, from much smaller tiles, the desired vegetation indexes are calculated (smaller tiles are faster and easier to process). The whole project is in the phase of the proposed prototype and preparations are currently underway to test a functional prototype with a focus on data and computing complexity.

Acknowledgements

This article was created with the support of the Internal Grant Agency (IGA) of FEM CULS in Prague, no. 20181010 – „Identification of limitations in processing of remote sensing data and in their efficient use“.

References

Arekhi, M., Goksel, C., Sanli, F. B. and Senel, G. (2019) 'Comparative Evaluation of the Spectral and Spatial Consistency of SENTINEL-2 and Landsat-8 OLI Data for Igneada Longos Forest', ISPRS International Journal of Geo-information, vol. 8, issue 2, article no. 56. doi: 10.3390/ijgi8020056.

Baetens, L., Desjardins, C. and Hagolle, O. (2019) 'Validation of Copernicus SENTINEL-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure', Remote Sensing, vol. 11, issue 4, article no. 433, EISSN 2072-4292. doi: 10.3390/rs11040433.

Carrasco, L., O'Neil, A. W., Morton, R. D. and Rowland, C. S. (2019) 'Evaluating Combinations of Temporally Aggregated Sentinel-1, SENTINEL-2 and Landsat 8 for Land Cover Mapping with Google Earth Engine', Remote Sensing, vol. 11, issue 3, article no. 288. doi: 10.3390/rs11030288.

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L,. Rabaute, T., Savinaud, M., Udroiu, C., Valero, S., Begue, A., Dejoux, J. F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., Lebourgeois, V., Miao, Z., Newby, T., Nyamugama, A., Salh, N., Shelestov, A., Simonneaux, V., Traore, P. S., Traore, S. S. and Koetz, B. (2019) 'Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world', Remote Sensing of Environment, vol. 221, pp. 551-568. doi: 10.1016/j.rse.2018.11.007.

Denize, J., Hubert-Moy, L., Betbeder, J., Corgne, S., Baudry, J. and Pottier, E. (2019) 'Evaluation of Using Sentinel-1 and-2 Time-Series to Identify Winter Land Use in Agricultural Landscapes', Remote Sensing, vol. 11, issue 1, article no. 37. doi: 10.3390/rs11010037.

Deng, L., Mao, ZH., Li, XJ., Hu, ZW., Duan, FZ and Yan, YN. (2018) 'UAV-based multispectral remote sensing for precision agriculture: A comparison between different cameras', ISPRS Journal of Photogrammetr and Remote Sensing, vol. 146, pp. 124-136. doi: 10.1016/j.isprsjprs.2018.09.008.

European Space Agency. 'Data Formats' [on-line] Available at https://sentinels.copernicus.eu/web/sentinel/user-guides/SENTINEL-2-msi/data-formats.

European Space Agency. 'Naming Convention' [on-line] Available at https://sentinels.copernicus.eu/web/sentinel/user-guides/SENTINEL-2-msi/naming-convention.

Fülöp, G., Sárközi, E. and Szigeti, I. S. (2015) 'Applying earth observation data for detection and estimation of woody biomass volumes of agricultural areas accessible for energy production', Proceedings of the 7th International Conference on Information and Communication Technologies in Agriculture, Food and Environment, HAICTA 2015. CEUR Workshop Proceedings, vol. 1498, pp 64-67. ISSN 16130073.

Halabuk, A., Mojses, M., Halabuk, M and David, S. (2015) 'Towards Detection of Cutting in Hay Meadows by Using of NDVI and EVI Time Series', Remote Sensing, vol. 7, issue 5, pp. 6107-6132. doi: 10.3390/rs70506107.

Lillesand, T. M., Kiefer, R. W. andChipman, J. W. (2008) 'Remote sensing and image interpretation', 6th issue, Hoboken, NJ: John Wiley and Sons. ISBN-10: 0470052457, ISBN-13: 978-0470052457.

Liu, C-C., Zhang, Y-C., Chen, P-Y., Lai, C-C., Chen, Y-H,, Cheng, J-H. and Ko, M-H. (2019) 'Clouds Classification from SENTINEL-2 Imagery with Deep Residual Learning and Semantic Image Segmentation', Remote Sensing, vol. 11, issue 2, article no. 119. doi: 10.3390/rs11020119.

Khaliq, A., Comba, L., Biglia, A., Aimonino, D. R., Chiaberge, M. and Gay, P. (2019) 'Comparison of Satellite and UAV-Based Multispectral Imagery for Vineyard Variability Assessment', Remote Sensing, vol. 11, issue 4, article no 436. EISSN 2072-4292. doi: 10.3390/rs11040436.

Reznik, T., Lukas, V., Charvat, K., Horakova, S and Charvat jr., K. (2015) 'Towards Farm-Oriented Open Data in Europe: the Scope and Pilots of the European Project "FOODIE"', Agris on-line Papers in Economics and Informatics, vol. 8, issue 1, pp. 51-58. ISSN 1804-1930.

Rodriguez-Ramirez, R., Sanchez, M. G., Rivera-Caicedo, J. P., Fajardo-Delgado, D and Avila-George, H. (2019) 'Automating an Image Processing Chain of the SENTINEL-2 Satellite', Proceedings of the Trends and Applications in Software Engineering, Books series: Advances in Intelligent Systems and Computing, vol. 865, pp. 216-334. doi: 10.1007/978-3-030-01171-0_20.

Szostak, M., Knapik, K., Likus-Cieslik, J., Wezyk, P and Pietrzykowski, M. (2019) 'Monitoring the spatial range of forested areas in the reclaimed sites using SENTINEL-2 images', Sylwan, vol. 163, issue 1, pp 55-61. ISSN 0039-7660.

Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F. and Ceschia, E. (2017) 'Understanding the temporal behavior of crops using Sentinel-1 and SENTINEL-2-like data for agricultural applications', Remote Sensing of Environment, vol. 199, pp 415-426. ISSN 00344257. DOI 10.1016/j.rse.2017.07.015.

Wang, B., Jia, K., Liang, SL., Xie, XH., Wei, XQ., Zhao, X., Yao, YJ and Zhang, XT. (2018) 'Assessment of SENTINEL-2 MSI Spectral Band Reflectances for Estimating Fractional Vegetation Cover', Remote Sensing, vol. 10, issue 12, article no. 1927. doi: 10.3390/rs10121927.