Environment. Technology. Resources. Rezekne, Latvia
Proceedings of the 14th International Scientific and Practical Conference. Volume 1, 209-213

Wetland Change Detection Using Sentinel-2 in the Part of Latvia

Andris Skromulis
Rezekne Academy of Technologies
Institute of Engineering
Rezekne, Latvia
andris.skromulis@rta.lv

Juris Breidaks
Institute of Electronics and Computer Science
Riga, Latvia
juris.breidaks@edi.lv

Mārtiņš Puķītis
Institute of Electronics and Computer Science
Riga, Latvia
martins.pukitis@edi.lv

***Abstract*. In the article, the possible impact of changes on wetland were analysed by the semi-supervised classification method of statistical analysis. The Sentinel-2 raw data between two different seasons are combined together. The data preparation is shortly described in the article. Data is clustered with unsupervised method. The article describes a supervised method – how data credibility and classification can be estimated if its reference is poor quality.**

Keywords: Wetlands, raised bogs, Sentinel-2, Semi-supervised classification, K-means, credibility.

I. Introduction

During the last fifteen years, interest in peatlands has significantly increased. Peatlands – a type of wetland – are among the most valuable ecosystems on Earth, providing a wide range of important ecosystem services including global biodiversity preservation [1, 2], mitigating water supply [3], [4], recreation [5], flood risk minimization [3], and climate change mitigation [6], [7], [8].

In Latvia, the most important wetland areas comprise more than 12400 rivers and 2256 lakes larger than one hectare with artificial water reservoirs occupying around 3.7% of the territory of Latvia [8]. A natural peatland is a wetland ecosystem in which organic matter production exceeds its decomposition. Under conditions of almost permanent water saturation and a lack of oxygen, dead plants and mosses accumulate as peat [9, 10]. Relatively intact swamps occupy 4.9%, while peat deposits (these consist of swamps and some types of wet forests on peat soils) occupy 10.4% of Latvia's territory [11]. Latvia is rich in peat resources, reaching 1.5 billion tonnes in the peatlands, and there are significant mineral deposits in the country [12].

As peat mines are abandoned, the areas, that were previously used for peat extraction, can start to regenerate naturally over time [13]. This regeneration can result in changes to the land cover, such as the growth of new vegetation, the development of wetlands and the expansion of peat bogs. These changes can have important ecological benefits, such as the restoration of important habitats for plants and animals, as well as providing important ecosystem services, such as carbon storage and water filtration.

However, it is also important to identify potential land cover changes that may occur during the regeneration process. For example, the expansion of new vegetation may lead to changes in the hydrology of the area, which may affect the local water balance and the availability of water resources for other uses. Additionally, changes in land cover can also impact the surrounding landscape, potentially leading to changes in the amount of sunlight that reaches the ground, changes in soil moisture, and changes in nutrient cycling [14].

Therefore, identifying areas of potential land cover change is important for understanding the potential impacts of bog habitat regeneration on the surrounding landscape, as well as for informing land management and restoration practices to ensure that ecological and social objectives are met.

In this paper, semi-supervised classification method – a technique used in remote sensing to classify land cover based on satellite imagery – is used. This approach involves combining both labelled and unlabelled data to train a classification model.

Semi-supervised classification is a technique used in remote sensing to classify land cover based on satellite imagery. Sentinel 2 is an Earth observation satellite

Print ISSN 1691-5402
Online ISSN 2256-070X
https://doi.org/10.17770/etr2023vol1.7305

developed and operated by the European Space Agency (ESA). Its imagery has 13 bands in the visible, near infrared and shortwave infrared part of the spectrum. It has a spatial resolution of 10 m, 20 m, and 60 m depending on the spectral band [16]. This paper presents a methodology for part of Sentinel-2 tile 34VEJ, which is part of the Latvia (Fig. 1).

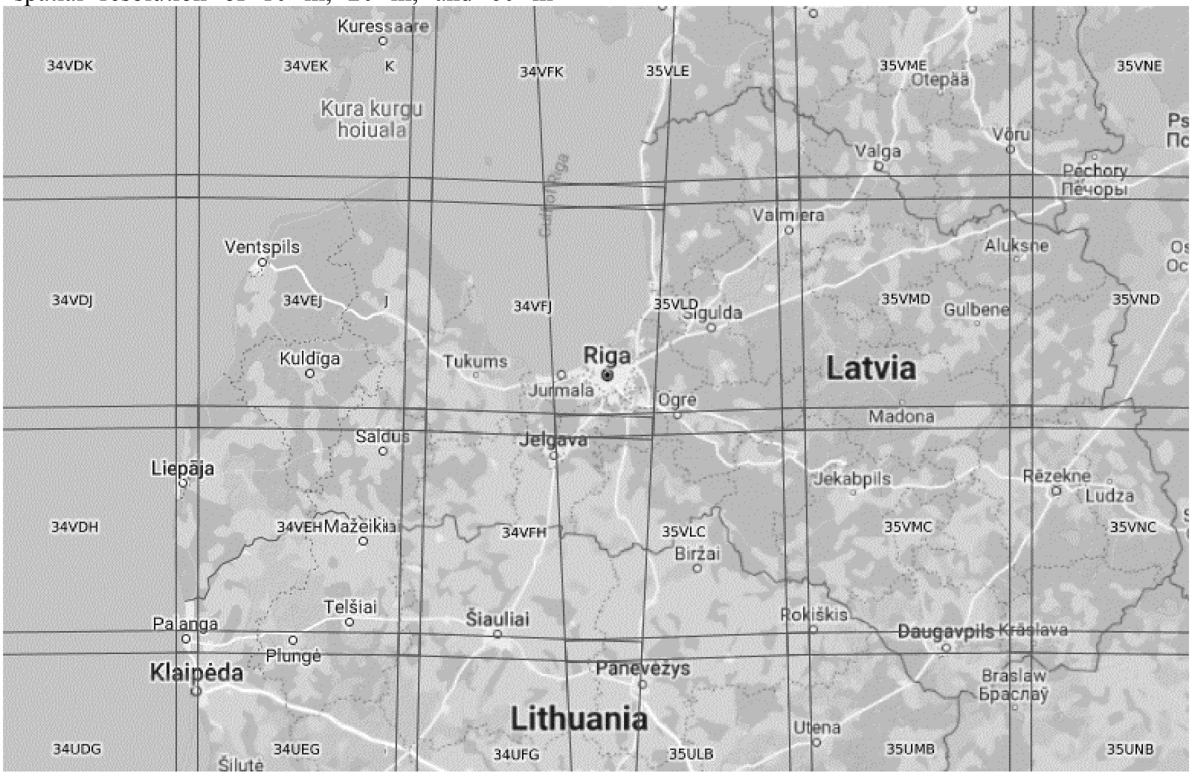

Fig. 1. Sentinel -2 UTM Tiling Grid for Latvia. [17]

II. Materials and Methods

A. Semi – supervised classification

Semi-supervised learning is a technique of machine learning that utilizes both labelled and unlabelled data to train a model.

In our case it combines the benefits of both unsupervised and supervised learning by using the unlabelled data to guide the clustering process and the labelled data to improve the generalization of the clusters. Semi-supervised learning has been shown to be effective in many real-world scenarios where labelled data are scarce, expensive, or difficult to obtain. It is a powerful tool that has been applied to various domains such as computer vision, natural language processing., etc.

Learning problems of this type are challenging as neither supervised nor unsupervised learning algorithms are able to make effective use of the mixtures of labelled and unlabelled data. Various approaches for data classification with partial labelling has been explored in [18], [19], [20], [21], [22], [23], [24], [25]. Our classification is based on semi-supervised algorithm, basis of which is described in [18], [25].

B. Algorithm

Our objective was to classify Sentinel-2 tile 34VEJ into 12 categories (Artificial objects, Agriculture, Forest, Bog habitats, Swamp forests (91D0*), Intact raised bogs (7110*) and Degraded raised bogs with potential or natural regeneration (7120), Transitional bogs and raised bogs (7140), Peat extraction sites, Licensed peat sites, Abandoned peat sites, Water, Other (unclassified)). To achieve the objective, tile 34VEJ of Sentinel-2 level 1C images were used. For this tile, the boundary mask of Latvia was used.

Tile was split in 16 overlapping fragments. We have also outdated information consisting of 2018 Copernicus Land Cover (CLC) data on man-made features, forests, as well as information from the Peat Association on developed bogs (abandoned peatlands) and licensed peatlands, as well as the Nature Conservation Management System “Ozols” data on bog habitats in 2021. Our algorithm is robust and even imprecise reference is still effective.

Initial classification of overlapping fragments can produce conflicting labelling for some areas. This was solved by providing a measure of confidence for classification - how certain we are that the category we have assigned is the correct one. We have named this measure credibility, each pixel in an image has a credibility value ranging from (theoretically) 0 (no confidence) to 1 (perfect confidence). In case of labelling conflict, classification with highest confidence was assigned.

First step is to divide image into clusters (unsupervised) based on spectral similarity. Any clustering algorithm could have been used. For this publication, K-means was picked due to its speed.

Let n be the number of categories, R_i,i ∈ [1,n] reference of the category i within the image, m - number of clusters and C_j,j∈[1,m] set of pixels in cluster j. We will use notation ∥ S ∥ for number of pixels in set S. If a cluster overlaps with reference set of exactly one category, it is obvious choice to assign that category to all pixels in such cluster. Equally obvious is credibility value of 1 for all pixels in cluster.

If a cluster overlaps more than one set of reference, the category it overlaps the most should be assigned. However, such approach had issues with smaller reference sets: if ∥ R_k ∥ ≪ ∥ R_j ∥ ∀j≠k , then category k will be ill-represented in final classification. Therefore, adding weights was required. On the other hand, too large weights caused over-representation of small categories, so we introduced max weight limit w_M. Final weight for category i is

$$w_i = min\left(\frac{\|R\|}{\|R_i\|}, w_M\right). \quad (1)$$

Cluster i is assigned to category

$$arg\ max_j \left(w_i \frac{\|C_i \cap R_j\|}{\|C_i\|}\right). \quad (2)$$

Credibility is assigned ignoring weights

$$C_{ik} = \frac{\|C_i \cap R_k\|}{\sum_j \|C_i \cap R_j\|}\ . \quad (3)$$

Finally, we are left with clusters, that do not overlap any reference at all. We use pixels from already assigned clusters to assign temporary category labels to all unassigned pixels based on spectral similarity. Any supervised algorithm can be used for this step, we used KNN [26]. Afterwards, we treat those temporary labels as reference set for clusters that did not overlap actual reference:

$$arg\ max_j \left(\frac{\|C_i \cap L_j\|}{\|C_i\|}\right), \quad (4)$$

where L_j is temporary labels of category j. Credibility is assigned analogically:

$$C_{ik} = \frac{\|C_i \cap L_k\|}{\sum_j \|C_i \cap L_j\|}\ . \quad (5)$$

III. Results and discussion

The images of tile 34VEJ with the least cloud cover are selected (to avoid any effect on ground cover). Then images of two different seasons are selected. For the combination of the part of tile 34VEJ images is selected image for the June 14, 2021 and image for the May 12, 2021.

The tiles are separated in 16 overlapping fragments, divided into the following parts along each axis: [0, 1/3], [2/9, 5/9], [4/9, 7/9], [2/3, 1].

Prepare a mask with populated areas from the available data of the dataset. Score the number of points in each pellet fragment after applying the mask. The reference of this pellet is made from 2018 Copernicus Land Cover (CLC) data on Artificial objects, forests, as well as peat association information on developed bogs (Abandoned peat sites) and licensed peat sites, as well as Nature Conservation Management System "Ozols" for habitats: Active raised bogs (7110) and degraded raised bogs in which natural regeneration is possible or in progress (7120), Transitional bogs and bogs (7140), and the water layer created after clustering.

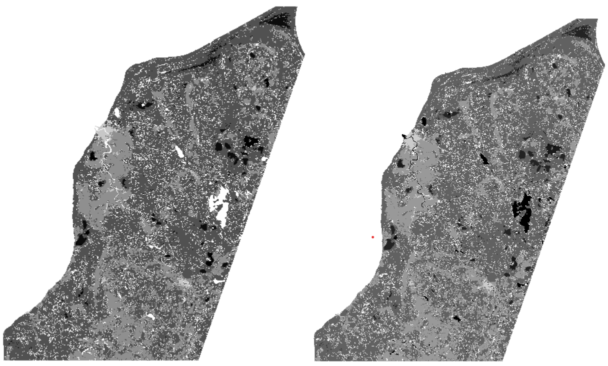

Fig. 2. Prepared reference before clustering and prepared reference after clustering.

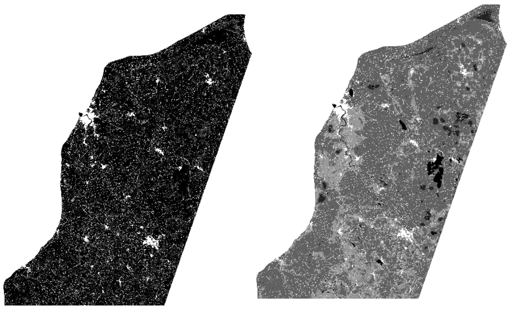

Fig. 3. First for updated credibility and second for labels.

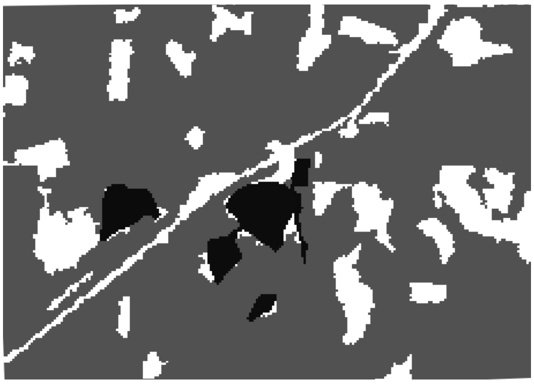

Fig. 4. Reference for field with center coordinates 57,46215° 21,93763° (EPSG:4326).

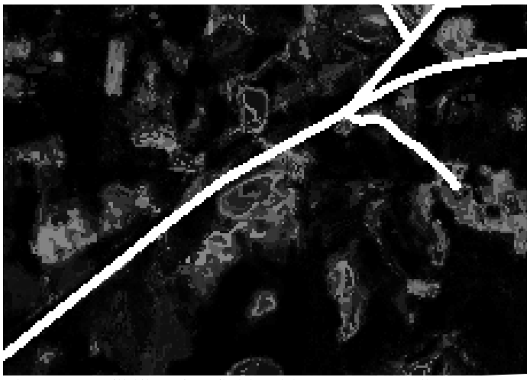

Fig. 5. Credibility for field with center coordinates 57,46215° 21,93763° (EPSG:4326).

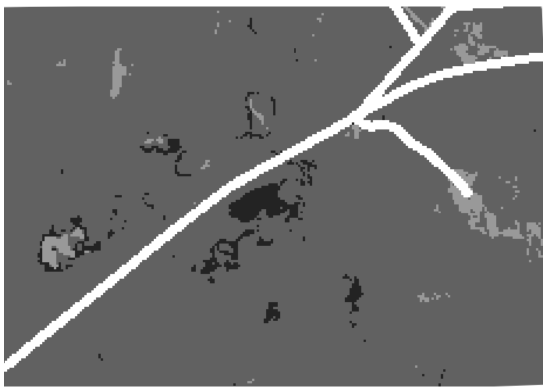

Fig. 6. Final labels for field with center coordinates 57,46215° 21,93763° (EPSG:4326).

IV. CONCLUSIONS

Analysing the data of the obtained results, it has been found that spectrally similar layers can be determined with the help of this algorithm. If Credibility is darker, then it is better. The obtained results provide important information about changes in the bog biotope.The obtained results provide important information about changes in the bog biotope.

Since the granule is divided into several overlapping fragments, a situation may arise that reference data for a class is missing in some fragments, thus it is necessary to improve the algorithm that takes into account clustering information from neighbouring areas.

Using this approach to satellite data, spectrally similar land cover as wetlands can be determined, but it is important that the reference data contains enough information about this class so that this method can be applied.

If, when preparing reference data, information about forests and meadows is used, then with the help of this method it is possible to restore information for these classes.

Further studies are related to the changes of the reference groups during the time when the previous classification result is the reference of the classifier of the next time period.

One potential conclusion from comparing semi-supervised learning with supervised and unsupervised learning algorithms is that semi-supervised learning can provide benefits over the other two approaches in scenarios where labelled data are limited or expensive to obtain. By leveraging the unlabelled data, semi-supervised learning can improve the accuracy of the model while reducing the need for a large amount of labelled data. This can be particularly useful in real-world applications where labelled data are scarce or difficult to obtain. However, the specific benefits of semi-supervised learning may vary depending on the specific problem domain and dataset, and it is important to evaluate the performance of different learning approaches on a case-by-case basis.

REFERENCES

[1] W. Wichtmann, C. Schröder, H. Joosten (Eds.), Paludiculture – Productive use of wet peatlands climate protection − biodiversity − regional economic benefits, Schweizerbart Science Publishers, Germany, Stuttgart, pp. 200-203

[2] Parish, F., Sirin, A. A., Charman, D., Joosten, H., Minaeva, T. Y., & Silvius, M. (2008). Assessment on peatlands, biodiversity and climate change.

[3] Ramsar Convention on Wetlands. (2018). Global wetland outlook: state of the world's wetlands and their services to people. Gland.

[4] Urák, I., Hartel, T., Gallé, R., & Balog, A. (2017). Worldwide peatland degradations and the related carbon dioxide emissions: the importance of policy regulations. *Environmental Science & Policy*, *69*, 57-64.

[5] Bonn, A., Allott, T., Evans, M., Joosten, H., & Stoneman, R. (Eds.). (2016). *Peatland restoration and ecosystem services: science, policy and practice*. Cambridge University Press.

[6] Smith, P., Clark, H., Dong, H., Elsiddig, E. A., Haberl, H., Harper, R., ... & Tubiello, F. (2014). Agriculture, forestry and other land use (AFOLU).

[7] Zhou, Y., Davidson, T. A., Yao, X., Zhang, Y., Jeppesen, E., de Souza, J. G., ... & Qin, B. (2018). How autochthonous dissolved

organic matter responds to eutrophication and climate warming: Evidence from a cross-continental data analysis and experiments. *Earth-Science Reviews*, *185*, 928-937.

[8] Morris, P. J., & Waddington, J. M. (2016). Climate change impacts on northern peatland carbon dynamics. Ecosystems, 19(2), 185-197

[9] Opermanis, O. (1998). Wetlands and the Ramsar Convention in Latvia. Published by Latvian Ornithological Society.

[10] Dinesen, L. & Hahn, P. (2019). Draft Ramsar Technical Report on peatland restoration and rewetting methodologies in Northern bogs. STRP22 Doc.7.2. https://www.ramsar.org/document/strp22-doc72-draft-ramsartechnical-report-on-peatland-restoration-and-rewetting

[11] Ķuze, I. (2012), Wetlands for nature and people, LIFE+ project "Restoring the hydrological regime of Ķemeri National Park" LIFE10 NAT/LV/000160 HYDROPLAN, pp 2, Nature Conservation Agency

[12] Pakalne, M., Kalnina, L. (2005): Mire ecosystems in Latvia / Moor-Ökosysteme in Lettland – Stapfia – 0085: 147 - 174.

[13] Šnore, A. (2013). Purvi un kūdra. Kūdras ieguve. Rīga: Nordik, 452

[14] Couwenberg, J., & Joosten, H. (2012). The role of peatlands in mitigating climate change. In High Latitude Peatlands and Global Climate Change (pp. 211-226). Springer, Dordrecht.

[15] Turner, T., Swindles, G., Charman, D. et al. (2016) Solar cycles or random processes? Evaluating solar variability in Holocene climate records. Sci Rep 6, 23961. https://doi.org/10.1038/srep23961

[16] Spatial Resolution, (https://sentinels.copernicus.eu/web /sentinel/user-guides/sentinel-2-msi/resolutions/spatial)

[17] Sentinel-2 UTM Tiling Grid (ESA). (2023). https://eatlas.org.au/data/uuid/f7468d15-12be-4e3f-a246-b2882a324f59

[18] Gabrys, B., Petrakieva, L. (2004). Combining labelled and unlabelled data in the design of pattern classification systems. International journal of approximate reasoning, 35(3):251–273.

[19] Bouchachia, A. (2007). Learning with partly labeled data. Neural Computing and Applications, 16(3):267–293.

[20] Blum, A., Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. Proceedings of the Workshop on Computational Learning Theory, page 92–100.

[21] Goldman, S., Zhou, Y. (2000). Enhancing supervised learning with unlabeled data. International Conference on Machine Learning, page 327–334.

[22] Ghahramani, Z.. Jordan, M.I. (1994). Supervised learning from incomplete data via an em approach. Advances in Neural Information Processing Systems, 6:120–127.

[23] Eick, C., Zeidat, N., Zhao, Z. (2004) Supervised clustering–algorithms and benefits. IEEE International Conference on Tools with Artificial Intelligence, page 774–776.

[24] Forestier. G., Wemmert, C. (2016). Semi-supervised learning using multiple clusterings with limited labeled data. Information Sciences, page 361–362.

[25] Chawla, N. V., Karakoulas, G. (2005). Learning from labeled and unlabeled data: An empirical study across techniques and domains. Journal of Artificial Intelligence Research, 23:331–366.

[26] Ferreira, D. (2019). k-Nearest Neighbors (kNN), MathWorks. Retrieved January 23, 2019, (https://www.mathworks.com/ matlabcentral/fileexchange/67018-k-nearest-neighbors-knn).