

# Enhancement of Land Cover Classification by Geospatial Data Cube Optimization

Artem Andreiev<sup>1</sup>, Anna Kozlova<sup>1</sup>, Leonid Artiushyn<sup>2</sup> and Peter Sedlacek<sup>3</sup>

<sup>1</sup> Scientific Centre for Aerospace Research of the Earth, Institute of Geological Sciences, National Academy of Sciences of Ukraine, Olesia Honchara str., 55-b, Kyiv, 01054, Ukraine

<sup>2</sup> State Research Institute of Aviation, Hryhoriia Andriuschenka Str., 6-V, Kyiv, 01135, Ukraine

<sup>3</sup> University of Zilina, Univerzitna, 8215/1, Zilina, 01026, Slovakia

## Abstract

This paper presents the optimization technique to reduce the geospatial data cube size and enhance the land cover classification. The technique is based on training sample separability. Accordingly, the Separability Index of the Training Sample (SITS) was developed and used as the object function for the optimization. In order to test the effectiveness of the optimization technique, the experiment was conducted. It implied the land cover classification of the highly heterogeneous natural landscapes in the case of the Shatsky National Natural Park, where the prevailing landscape is wetlands. After the optimization of the input geospatial data cube, classification enhancement was evidenced by increasing indicators such as overall accuracy by 0.04 from 0.9 to 0.94 and the kappa coefficient by 0.06 from 0.86 to 0.92. In addition, the data cube size was reduced by 5.55 times from 222 to 40 layers

## Keywords

Remote sensing, land cover classification, supervised classification, training sample separability, geospatial data cube, data optimization

## 1. Introduction

Land cover classification is a critical process in remote sensing, providing spatially explicit information at different scales for numerous environmental applications [1]. Such information is widely applied to issues that require practical geospatial solutions like land cover change detection [2], environmental monitoring [3,4], fossil fuel exploration [5], and landmine detection [6]. Land classification techniques, likewise, play a crucial role in the integration of Earth observation data into comprehensive interdisciplinary issues on sustainable development goals achieving [7, 8], in particular, combat climate change and its impacts [9], reverse land degradation [10] and halt biodiversity loss [11], protect water-related ecosystem for safety water supply [12], and provide support for food security and sustainable agriculture [13, 14].

Today, most classification methods are mainly divided into supervised and unsupervised [15]. However, in remote sensing, the supervised classification methods are the most appropriate for the majority of the thematic tasks because applying these methods can establish the characteristics of the output classes, unlike unsupervised ones. A training sample set is used to set the characteristics of classes in supervised classification methods. Such a set contains the signatures of features of each class.

The input data for classification is heterogeneous geospatial data, which can be represented in the form of raster layers. To combine such layers into a single array, it is customary to form a geospatial data cube [16]. From ordinary datasets, data cubes differ by integrating different data types into a coherent and interoperable structure [17, 18]. After the cube's formation, the training sample's signatures must be determined in each layer. Hence, each layer is a feature of the training sample.

---

CMIS-2024: Seventh International Workshop on Computer Modeling and Intelligent Systems, May 3, 2024, Zaporizhzhia, Ukraine

✉ artem.a.andreev@gmail.com (A. Andreiev); ak.koann@gmail.com (A. Kozlova); artleonid2017@gmail.com (L. Artiushyn); peter.sedlacek@fri.uniza.sk (P.Sedlacek)

🆔 0000-0002-6485-449X (A. Andreiev); 0000-0001-5336-237X (A. Kozlova); 0000-0002-7488-7244 (L. Artiushyn); 0000-0002-7481-6905 (P.Sedlacek)



© 2024 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

A geospatial data cube for specific classification tasks can comprise numerous layers. Including multitemporal data, a data cube aims to distinguish dynamic objects that change significantly during a specific period or vary much from each other at different stages of their development, e.g. vegetation cover, wetlands in particular [12]. Different physical aspects of multisource data, e.g. optical and radar, highlight diverse object traits and variations in land cover types [19]. Multiple ancillary data, e.g. data on geomorphology, hydrology, or phenology, help to differentiate land cover types due to their context [20].

However, the redundancy of the geospatial data cube causes two significant problems [21]. Firstly, the processing of such a data cube has high computational complexity. Secondly, since the signatures of the training sample are defined in each layer of the geospatial data cube, the separability of the training sample can be reduced if the layers are either incorrectly created or irrelevant to the selected thematic task. In turn, the low separability of the training sample leads to a decrease in classification accuracy [22].

In light of the above, optimization of the geospatial data cube is seen as a solution to the mentioned problems [23, 24]. Among approaches to reduce input data, the Principal Component Analysis (PCA) [25] and the Minimum Noise Fraction (MNF) [26] are the most widely used. There are also similar methods, for example, Noise-Adjusted Principal Components (NAPC) [27], Independent Component Analysis (ICA) [28], Non-Negative Matrix Factorization (NMF) [29] and Spatio-Spectral Decomposition (SSD) [30]. However, a common disadvantage of the considered approaches is that they do not consider the training sample's structure (in particular separability) and the selected classifier's specificity.

The presented study aims to enhance land cover classification by selecting the cube layers, the training sample separability of which will be the highest among other options. For this purpose, the optimization technique of the geospatial data cube was developed. It has two goals: enhancement of land cover classification and reduction of geospatial data cube size.

Hence, in the relevant sections of this article, the separability assessment of the training sample, the optimization technique of the geospatial data cube, and the experiment conducted to demonstrate the effectiveness of the developed technique are described.

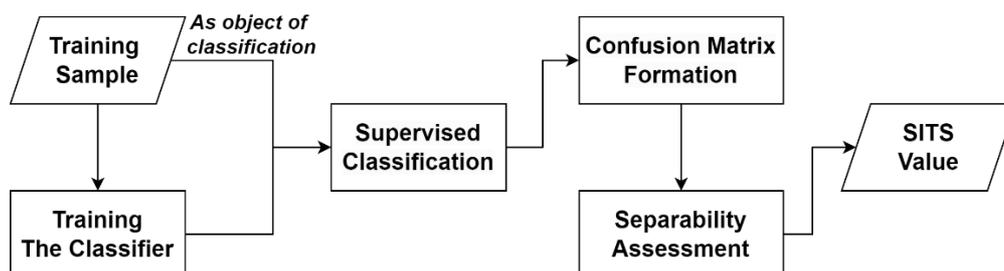
## 2. Methods

This section presents the optimization technique of a geospatial data cube. Since this optimization is based on training sample separability, the objection function is the developed separability index of the training sample (SITS). Thus, the training sample separability assessment is also presented below as an algorithm for SITS calculation.

### 2.1. Assessment of the training sample separability

Separability is one of the training sample characteristics that affect classification accuracy. This characteristic shows the extent to which signatures representing different classes do not overlap. A low degree of separability is inherent in a high level of training sample mixing. In turn, this leads to a significant number of misclassified objects in the classification. Thus, the training sample separability is directly proportional to the classification accuracy.

The algorithm depicted in the flowchart(Figure 1) describes the separability assessment of the training sample.



**Figure 1:** Algorithm of the separability assessment of the training sample

The first step implies classifier training by the training sample. Importantly, the supervised classification method must be the same as the one selected to classify the geospatial data cube further. Moreover, due to the proposed, separability depends on its structure (i.e., the set of layers) and the selected supervised classification method.

In the second step, the classifier is used to classify each signature from the training sample set.

The third step is the formation of the confusion matrix [31] for the classification obtained in the previous step.

The fourth and final step is calculating the SITS. This index quantifies the separability of training samples by measuring the ratio of correctly classified training samples to the total number of training samples. In other words, SITS equals the overall accuracy [31] based on the confusion matrix obtained in the previous step. The calculation of the SITS is shown in the following formula:

$$SITS = \frac{\sum_{i=1}^K x_{ii}}{N}, \quad (1)$$

where  $K$  is the number of classes,  $N$  is the total number of training sample signatures,  $x_{ii}$  is the number of class  $i$  signatures classified as class  $i$  (i.e. diagonal elements of the obtained confusion matrix that correspond to correctly classified signatures).

The values of the considered index range from 0 to 1. In this case, the value 0 shows that the training sample is entirely mixed (minimum separability), and the value 1 corresponds to the training sample, which is entirely separable (maximum separability).

## 2.2. Optimization technique

This technique is an optimization procedure that aims to reduce the number of layers of the geospatial data cube and increase the separability of the training sample, the signatures of which are defined in each layer of this cube. The objective function implies using the SITS. Thus, geospatial data cube optimization can be described as a search of the minor number of cube layers for which the training sample has the highest SITS value among all other sets of cube layers.

The flowchart of the technique algorithm is shown in Figure 2.

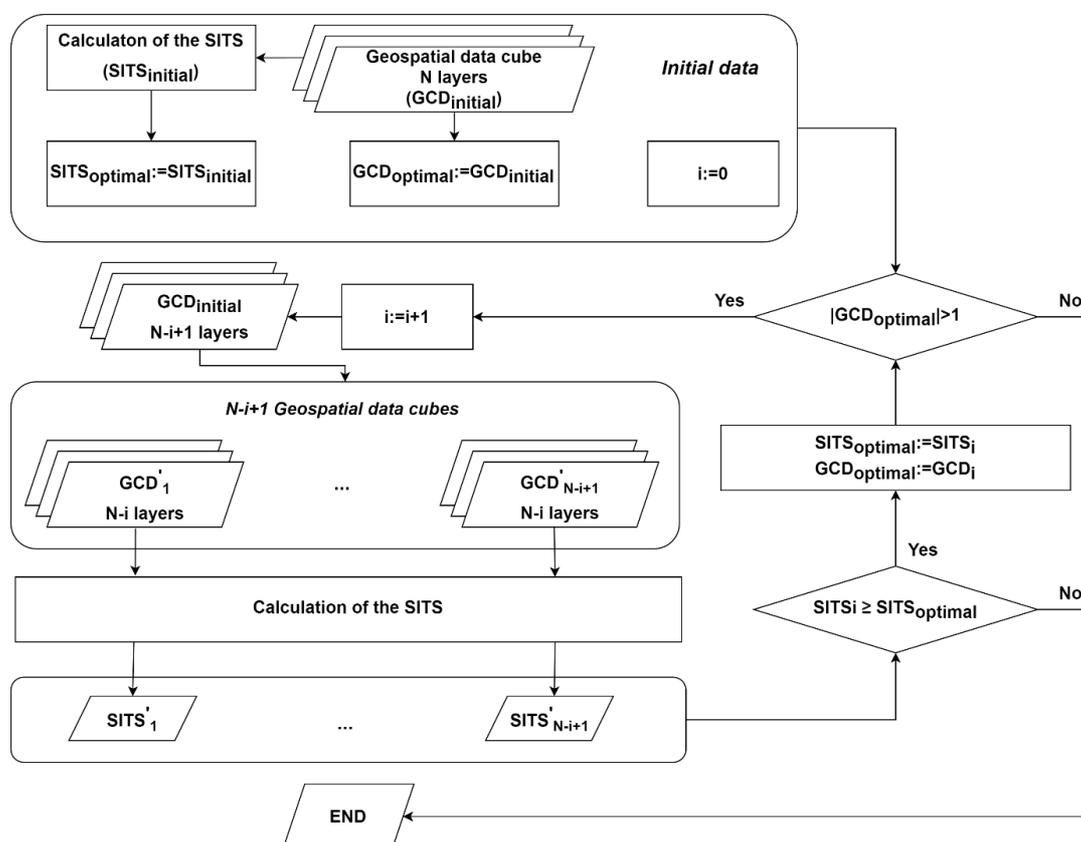


Figure 2: Algorithm of the optimization technique

The initial data and their characteristics will be introduced below for a detailed description of the technique algorithm.

Let the initial geospatial data cube have the following form:

$$GCD_{initial} = \{L_1, L_2, \dots, L_N\},$$

where  $L_i$  is the layer  $i$  of the geospatial data cube, and  $N$  is the total number of layers included in the initial geospatial data cube.

Then, as shown in Figure 2, the SITS value of the initial training sample, which has the signatures defined in each layer of the initial geospatial data cube, is first calculated. Let this value be  $SITS_{initial}$ .

Let us assign the values obtained above to the variables corresponding to the optimal set of layers of the geospatial data cube and the corresponding value of the SITS. Thus, we have:

$$GCD_{optimal} := GCD_{initial},$$

$$SITS_{optimal} := SITS_{initial}.$$

Next, an iterative procedure follows, in which the following steps are performed at each iteration.

**Step 1.** At the iteration  $i$ , the current geospatial data cube  $GCD_{optimal}$ , consisting of  $N-(i-1)$  layers, is decomposed into  $N-(i-1)$  cubes. Each newly formed cube is obtained by discarding one of the layers from the current cube. Then, each of the newly created geospatial data cubes will have the following form:

$$GCD'_1 = GCD - \{L_1\} = \{L_2, L_3, \dots, L_{N-(i-1)}\},$$

$$GCD'_2 = GCD - \{L_2\} = \{L_1, L_3, \dots, L_{N-(i-1)}\},$$

...

$$GCD'_{N-(i-1)} = GCD - \{L_{N-(i-1)}\} = \{L_1, \dots, L_{N-(i-1)-1}\}.$$

Since the number of layers decreases by one at each iteration, the obtained cubes will contain  $N-i$  layers. Therefore, the following is valid:

$$|GCD'_1| = |GCD'_2| = \dots = |GCD'_{N-(i-1)}| = N - i.$$

Thus, the generated cubes can be written in the form of the following set:

$$C = \{GCD'_1, GCD'_2, \dots, GCD'_{N-(i-1)}\}.$$

**Step 2.** For each newly formed cube, the SITS value is calculated for the training sample, the signatures of which are defined in each cube layer. Then, the value of SITS for a particular cube  $GCD'_t$  will be denoted as  $SITS'_t$ . Thus, a set containing the value of the SITS for each newly formed cube will be obtained:

$$S = \{SITS'_1, SITS'_2, \dots, SITS'_{N-(i-1)}\}.$$

**Step 3.** Among the obtained cubes, the one with the highest value of the SITS is selected. Such a cube will be denoted as  $GCD_i$ . The selected cube can be expressed as follows:

$$GCD_i = \{GCD'_t \in C | SITS'_t = \max\{S\}\}.$$

$SITS_i$  denotes the value of the SITS of the geospatial data cube  $GCD_i$ .

**Step 4.** The values of the variables  $SITS_i$  and  $SITS_{optimal}$  are compared, and two options are considered:

1) if  $SITS_i < SITS_{optimal}$ , then the execution of the optimization algorithm will be interrupted, and further steps will be ignored. The optimal geospatial data cube will be the one obtained in the previous iteration, namely  $GCD_{optimal}$ . Accordingly, the SITS of the training sample of the corresponding cube has the value  $SITS_{optimal}$ .

2) if  $SITS_i \geq SITS_{optimal}$ , then the variable  $SITS_{optimal}$  is assigned the value of the variable  $SITS_i$ , i.e.:

$$SITS_i := SITS_{optimal}.$$

**Step 5.** At this step, as at the previous one, two options are considered:

1) if the number of layers of the obtained cube  $GCD_{optimal}$  is 1, i.e.:

$$|GCD_{optimal}| = 1,$$

then the execution of the optimization algorithm will be interrupted. The optimal geospatial data cube will be the one obtained at the current iteration, namely  $GCD_{optimal}$ . Accordingly, the SITS value of the training sample of the corresponding cube is  $SITS_{optimal}$ .

2) if the number of layers of the obtained geospatial data cube  $GCD_{optimal}$  is greater than 1, i.e.:

$$|GCD_{optimal}| > 1,$$

then a new iteration will start, and the actions specified in **step 1** will be performed again.

The result of the described optimization procedure is the geospatial data cube with the set of layers that achieves the highest value of the SITS among all other considered sets. This geospatial data cube will be used for further land cover classification.

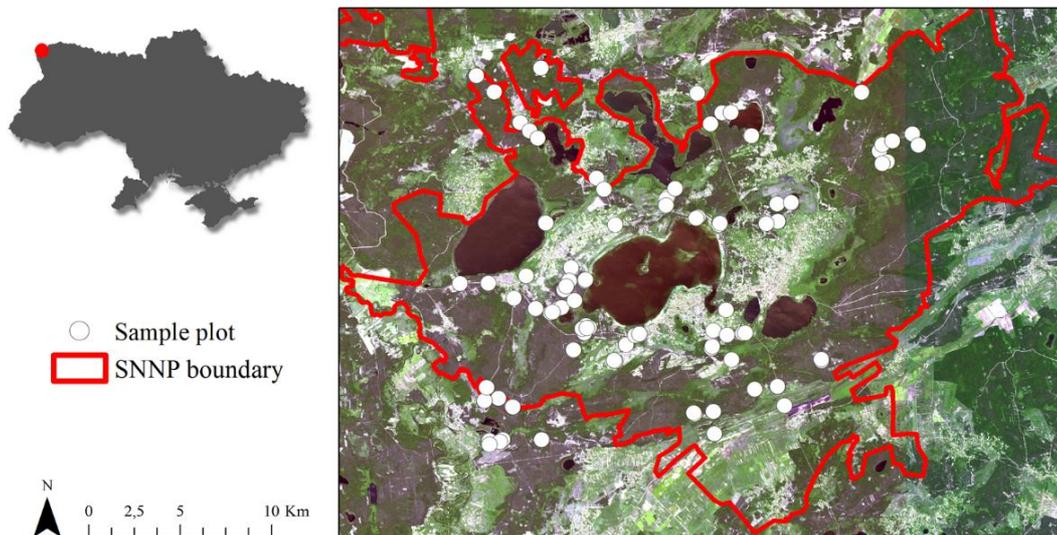
### 3. Experiment

The experiment was conducted to test the effectiveness of the developed technique. It consisted of carrying out a classification of the selected study area. For this purpose, an initial geospatial data cube with an excessive number of layers was formed. Then, the optimization technique was applied, resulting in an optimized cube. Finally, two classifications were obtained – before and after the optimization.

#### 3.1. Study area

Since 2007, the Ukrainian network of test sites has provided validation and calibration of various remote sensing techniques and satellite-based products, including land cover classification [32]. The proposed technique was tested at the site within the Shatsk National Natural Park (SNNP). It is situated in the northwest of Ukraine, within Volyn' oblast, between  $51^{\circ} 28' 25''$ N and  $23^{\circ} 49' 29''$ E. The SNNP encompasses highly heterogeneous natural landscapes, like forests, peat bogs, transitional mires, meadows, and lakes.

The site comprises more than 100 georeferenced sample plots and gives comprehensive ground truth information about the representative landscapes of the West Polissia region (Figure 3).



**Figure 3:** Location of the study area and sample plots within the Shatsk National Natural Park. The background is the true-colored composite of the Sentinel-2 Multispectral Instrument (MSI) image acquired on 1 June 2018

#### 3.2. Training sample

Six broad land cover classes that characterize the study area were defined: artificial surfaces, tree-covered areas, grassland, agricultural areas, water bodies, and wetlands.

The given classes varied considerably both in spatial extent and heterogeneity. The smallest class of artificial surfaces included diverse features of built-up areas and transport units. While the biggest ones, like tree-covered areas and wetlands, included various sub-types that could still

be quite homogenous due to the big extent. Water bodies represent the most homogeneous class. Therefore, the number of training pixels of each class also varied disproportionately. The overall number of all training pixels accounted for 6474. Table 1 shows labels, descriptions, and training pixel amounts for the land cover classes assigned for the experiment.

**Table 1**  
**The classification scheme used in the experiment**

#	Land Cover Class	Description	Training pixels
1	Artificial surfaces	Urban public and industrial built-up areas, transport units, and construction sites	319
2	Tree-covered areas	Broadleaved, coniferous, mixed and swamped forests, orchards, roadside tree lines	2313
3	Grasslands	Natural herbaceous vegetation, permanent grasslands of natural origin, pastures	634
4	Agricultural areas	Arable land, permanent crops, fallow lands, heterogeneous agricultural areas, open soils	887
5	Water bodies	Lakes, rivers and streams of natural origin, including man-made reservoirs and canals	634
6	Wetlands	Non-forested areas of peat bogs, transitional mires, eutrophic marshes, and reed beds	1687

### 3.3. Initial geospatial data cube

The experimental classification focuses on wetlands, the prevailing landscape of the test site and one of the most important for conservation within the SNNP. Evident differences in the seasonal development of wetlands and other vegetative land cover classes help distinguish them and require the application of multitemporal data [12].

The primary data source for forming the geospatial data cube was Sentinel-2 satellite imagery [33]. The images were selected for 4 dates (04.07.2018, 05.12.2018, 06.01.2018 and 10.14.2018) with minimal or no cloudiness.

Each Sentinel-2 image contains 13 spectral bands. At the preprocessing stage, atmospheric correction was performed for each image to eliminate the influence of the atmosphere and calculate the pixel values corresponding to the surface reflectance (bottom of atmosphere). During this procedure, 3 bands (B1, B9 and B10) that consider the effects of aerosols and water vapour on reflectance were removed. The spectral bands of the Sentinel-2 image have different spatial resolutions, namely 10, 20 and 60 meters.

Next, a complete set of normalized difference indices was calculated for each image. The following combinatorial formula is used to calculate such an index:

$$Index = \frac{b_i - b_j}{b_i + b_j}, i \neq j,$$

where  $b_i$  is the spectral band  $i$ .

This set could be presented in the following form:

$$NDI = \left\{ \frac{b_i - b_j}{b_i + b_j} \mid i, j \in \{1, 2, \dots, 10\}, i \neq j \right\}.$$

The cardinality of this set (i.e. number of normalized difference indices of one image) is calculated by the formula below:

$$|NDI| = C_n^m = \frac{n!}{m!(n-m)!} = \frac{10!}{2!(10-2)!} = \frac{8! * 9 * 10}{1 * 2 * 8!} = \frac{90}{2} = 45,$$

where variable  $n$  corresponds to the number of image bands, and  $m$  is the number of arguments in the index calculation formula.

As a result, 45 spectral indices were obtained for each image.

Another component of the input cube was the geomorphological data obtained from the ALOS PALSAR DEM [34]. In particular, this data contains the height above sea level and the slope. Their spatial resolution is 12.5 m.

All the above-described data must be spatially regularized to form the input geospatial data cube. It involves bringing all layers to the same spatial resolution, map projection, and size. With this in mind, all layers were scaled to a spatial resolution of 10 m, transformed to Universal Transverse Mercator Projection, Zone 34N (EPSG:32634) and resized so that all layers lie within the study area.

So, the input geospatial data cube contained 222 raster layers, namely 40 spectral bands of 4 different time Sentinel-2 satellite images, 180 corresponding spectral indices, and 2 raster layers of geomorphological parameters. The layers of the input cube are described in Table 2.

**Table 2**  
**Layers of the initial geospatial data cube**

Date (DD/MM/YYYY)	Spectral bands	Spectral indices	Geomorphological data
07.04.2018	10	45	
12.05.2018	10	45	
01.06.2018	10	45	2
14.10.2018	10	45	

Hence, signatures of the initial training sample were assigned in each layer of the initial cube. To assess the separability of the training sample, the SITS was calculated using Formula 1:

$$SITS_{Initial} = \frac{\sum_{i=1}^K x_{ii}}{N} = \frac{6469}{6474} \approx 0.9992.$$

As seen above, this training sample had 5 misclassified signatures.

### 3.4. Optimized geospatial data cube

After applying the developed optimization technique, the geospatial data cube size was reduced from 222 to 40 layers. The selected layers are listed below in Table 3.

**Table 3**  
**Layer of the optimized geospatial data cube**

Date (DD/MM/YYYY)	Spectral bands	Spectral indices	Geomorphological data
07.04.2018	0/10	4/45	
12.05.2018	4/10	0/45	
01.06.2018	2/10	17/45	1/2
14.10.2018	1/10	11/45	

Along with the cube optimization, the training sample signatures were reassigned according to the selected cube layers. The separability of that training sample was assessed by the SITS calculated below:

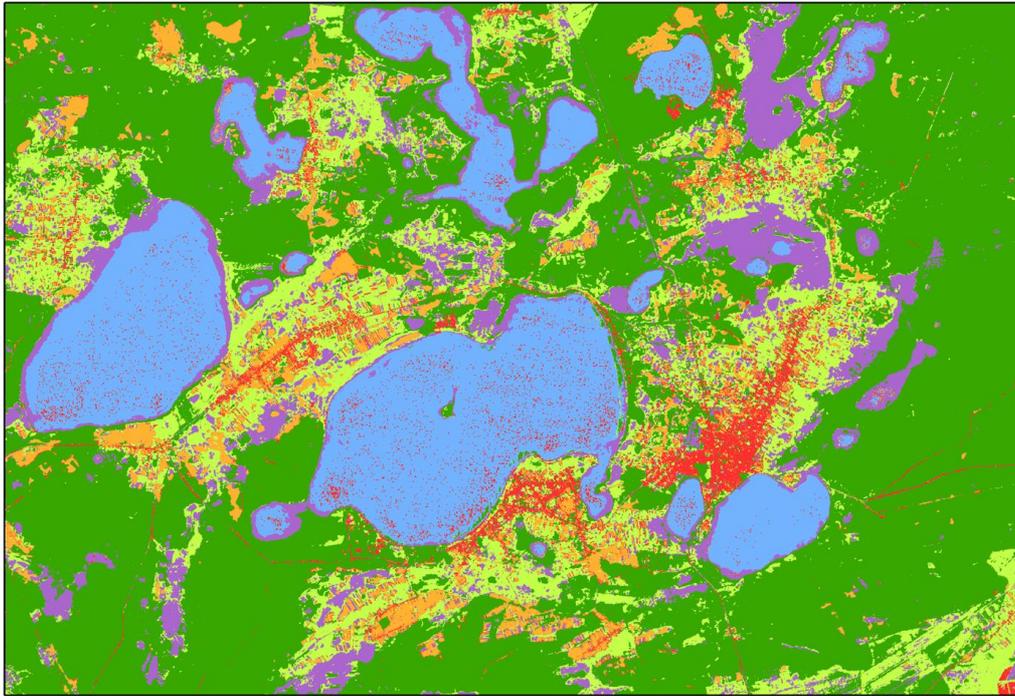
$$SITS_{Optimal} = \frac{\sum_{i=1}^K x_{ii}}{N} = \frac{6474}{6474} = 1.$$

This training sample had no misclassified signatures, so the optimized training sample was entirely separated.

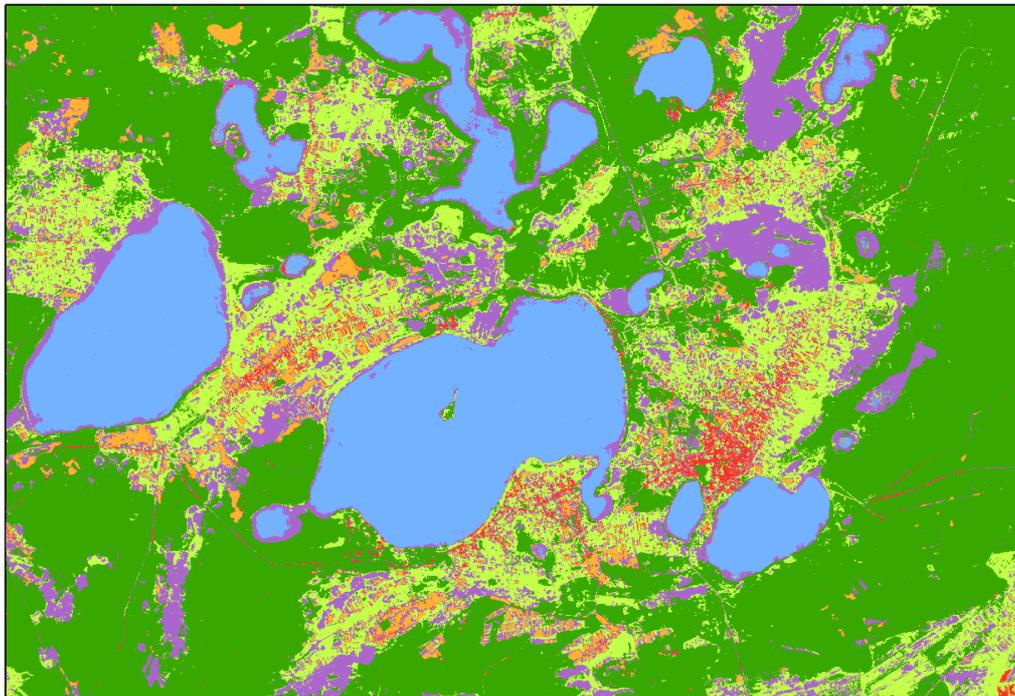
### 3.5. Land cover classifications

In order to test the effectiveness of the developed technique, the classifications using the initial cube and the optimized one were compared.

Firstly, the classification was obtained using the initial cube and the appropriate training sample. Then, the classification was obtained using the optimized cube and the appropriate training sample. These classifications are depicted in Figure 4.



a)



b)



**Figure 4:** Land cover maps of the study area were obtained using a) the initial cube and b) the optimized cube

The classifications above were obtained using Mahalanobis distance [35] as a supervised classification method. Exactly for this method, the separability assessment of the training sample was carried out for both initial and final geospatial data cubes.

### 3.6. Accuracy assessment

Classification accuracy assessment involved independent verification of initial and final land cover maps using proportionate stratified random samplings. This sampling technique produces sample set sizes directly related to the size of the classes and is widely used in assessing the classification accuracy of classes disproportionate in their extent. As the required sample size for a class, 0.01% of the total classified pixels of this class were analyzed. Thus, test sample sets were equal to 300 pixels for each land cover map. Satellite images (QuickBird) of high spatial resolution, available for 2018 in the Google Earth Pro app, were used for verification as reference data.

Table 4 shows the confusion matrix of the initial land cover map. In addition to the two dimensions ("Reference" and "Prediction"), this matrix shows metrics such as producer accuracy (PA) and user accuracy (UA) for each class [31].

**Table 4**  
**Confusion matrix of the initial land cover map**

		Reference						Total (pixels)	UA
		1	2	3	4	5	6		
Prediction	1.Artificial surfaces	14	0	1	2	7	0	24	0,58
	2.Tree-covered areas	0	134	0	0	1	0	135	0,99
	3.Grasslands	0	3	29	5	0	1	38	0,76
	4.Other lands	0	0	0	15	0	0	15	1
	5.Water bodies	0	0	0	0	53	0	53	1
	6.Wetlands	1	0	6	2	1	25	35	0,71
	Total (pixels)	15	137	36	24	62	26	300	
PA	0,93	0,98	0,81	0,63	0,85	0,96			

The accuracy of the obtained land cover classification was assessed by indicators of overall accuracy and the kappa coefficient [31]. The overall accuracy value was calculated by the following formula:

$$OA = \frac{\sum_{i=1}^K x_{ii}}{N} = \frac{14 + 134 + 29 + 15 + 53 + 25}{300} = \frac{270}{300} = 0.9,$$

where  $K$  is the number of classes,  $N$  is the total number of test samples,  $x_{ii}$  is diagonal element  $i$  of the confusion matrix (i.e. number of correctly classified samples of class  $i$ ). The value of the kappa coefficient was obtained following the calculations below:

$$Kappa = \frac{N * \sum_{i=1}^K x_{ii} - \sum_{i=1}^K (\sum_{j=1}^K x_{ij} * \sum_{j=1}^K x_{ji})}{N^2 - \sum_{i=1}^K (\sum_{j=1}^K x_{ij} * \sum_{j=1}^K x_{ji})} = \frac{300 * 270 - 24779}{300^2 - 24779} = \frac{56221}{65221} \approx 0.86.$$

Table 5 shows the confusion matrix of the final land cover map.

**Table 5**  
**Confusion matrix of the final land cover map**

		Reference						Total (pixels)	UA
		1	2	3	4	5	6		
Prediction	1.Artificial surfaces	8	0	0	1	1	0	10	0,8
	2.Tree-covered areas	0	128	0	0	0	0	128	1
	3.Grasslands	0	6	40	1	0	0	47	0,85
	4.Other lands	1	0	0	14	0	0	15	0,93
	5.Water bodies	0	0	0	0	56	0	56	1
	6.Wetlands	1	1	1	5	0	36	44	0,82
	Total (pixels)	10	135	41	21	57	36	300	
	PA	0,8	0,95	0,98	0,67	0,98	1		

The same indicators were selected for the final land cover classification as for the initial one. Thus, the value of overall accuracy value is stated below:

$$OA = \frac{\sum_{i=1}^K x_{ii}}{N} = \frac{8+128+40+14+56+36}{300} = \frac{282}{300} = 0.94.$$

Then, the following calculations are for the kappa coefficient value:

$$Kappa = \frac{N * \sum_{i=1}^K x_{ii} - \sum_{i=1}^K (\sum_{j=1}^K x_{ij} * \sum_{j=1}^K x_{ji})}{N^2 - \sum_{i=1}^K (\sum_{j=1}^K x_{ij} * \sum_{j=1}^K x_{ji})} = \frac{300 * 282 - 24779}{300^2 - 24779} = \frac{59821}{65221} \approx 0.92.$$

## 4. Discussion

The aims of the developed optimization technique are stated as layers reduction of the geospatial data cube and enhancement of the classification. Thus, the experiment result should be considered in terms of these two aspects.

Firstly, the size of the optimized cube was 40 layers, whereas the initial one contained 222 layers. Therefore, the number of layers was reduced by 5.55 times.

Secondly, classification enhancement was evidenced by increasing indicators such as overall accuracy and the kappa coefficient. Namely, the overall accuracy increased by 0.04 from 0.9 to 0.94, and the kappa coefficient increased by 0.06 from 0.86 to 0.92. Since the classification focused on the wetlands, the accuracy of this class should be considered individually. Thereby, both user and producer accuracy of wetlands class were significantly increased, viz. by 0.11 from 0.71 to 0.82 and by 0.4 from 0.96 to 1, respectively.

## 5. Conclusion

This article presents an optimization technique to reduce geospatial data cube size and enhance land cover classification. The technique is based on the separability of the training sample, which is defined in each layer of the geospatial data cube. To assess the separability, the appropriate index (i.e. SITS) was developed and used as an object function in the technique frame. The

algorithm of the optimization technique implies stepwise band discarding to define the optimal set of the geospatial data cube layers. Such a set has the highest value of SITS among other options.

The conducted experiment implied techniques application to the land cover classification of the highly heterogeneous natural landscapes in the case of the Shatsky National Natural Park. This classification covered six land cover classes where wetlands are prevailing. The technique's effectiveness was approved by geospatial data cube reduction and classification accuracy enhancement, evidenced by the increase in such indicators as overall accuracy and kappa coefficient.

Further research should be aimed at technique application in other study areas and thematic tasks. Also, the separability assessment of the training sample could be extended by additional criteria. For example, the kappa coefficient could substitute overall accuracy as the basis of the developed separability index.

## References

- [1] A.M. Melesse, Q. Weng, P.S. Thenkabail, G.B. Senay, Remote sensing sensors and applications in environmental resources mapping and modelling, *Sensors*, vol. 7, no. 12, pp. 3209–3241 (2007). doi: 10.3390/s7123209
- [2] A.H. Chughtai, H.U. Abbasi, İ. R. Karaş, A review on change detection method and accuracy assessment for land use land cover, *Remote Sensing Applications: Society and Environment*, vol. 22, p. 100482 (2021). doi: 10.1016/j.rsase.2021.100482
- [3] E. Zaitseva, S. Stankevich, A. Kozlova, I. Piestova, V. Levashenko, P. Rusnak, Assessment of the risk of disturbance impact on primeval and managed forests based on earth observation data using the example of Slovak Eastern Carpathians, *IEEE Access*, vol. 9, pp. 162847–162856 (2021). doi: 10.1109/access.2021.3134375
- [4] A. Kozlova, S. Stankevich, M. Svideniuk, A. Andreiev, Quantitative Assessment of Forest Disturbance with C-Band SAR Data for Decision Making Support in Forest Management, *Lecture notes on data engineering and communications technologies*, pp. 548–562 (2021). doi: 10.1007/978-3-030-82014-5\_37
- [5] M.A. Popov, Topolnytskyi M.V., O.V. Titarenko, S.A. Stankevich, A.A. Andreiev, Forecasting gas and oil potential of subsoil plots via co-analysis of satellite, geological, geophysical and geochemical information by means of subjective logic, *WSEAS Transactions on Computer Research*, vol. 8, pp. 90–101 (2020). doi: 10.37394/232018.2020.8.11
- [6] M.A. Popov, S.A. Stankevich, S.P. Mosov, O.V. Titarenko, S.S. Dugin, S.I. Golubov, A.A. Andreiev, Method for Minefields Mapping by Imagery from Unmanned Aerial Vehicle, *Advances in Military Technology*, vol. 17, no. 2, pp. 211–229 (2022). doi: 10.3849/aimt.01722
- [7] A. Andries, S. Morse, R.J. Murphy, J.M. Lynch, E. Woolliams, J. Fonweban, Translation of Earth observation data into sustainable development indicators: An analytical framework, *Sustainable Development*, vol. 27, no. 3, pp. 366–376 (2018). doi: 10.1002/sd.1908
- [8] G. Scott, A. Rajabifard, Sustainable development and geospatial information: a strategic framework for integrating a global policy agenda into national geospatial capabilities, *Geospatial Information Science*, vol. 20, no. 2, pp. 59–76 (2017). doi: 10.1080/10095020.2017.1325594
- [9] M. Popov, S. Stankevich, Y. Kostyuchenko, A. Kozlova, Analysis of Local Climate Variations Using Correlation between Satellite Measurements of Methane Emission and Temperature Trends within Physiographic Regions of Ukraine, *International Journal of Mathematical, Engineering and Management Sciences*, vol. 4, no. 2, pp. 276–288 (2019). doi: 10.33889/ijmems.2019.4.2-023
- [10] O. Dubovyk, The role of Remote Sensing in land degradation assessments: opportunities and challenges, *European Journal of Remote Sensing*, vol. 50, no. 1, pp. 601–613 (2017). doi: 10.1080/22797254.2017.1378926
- [11] E. Agrillo, F. Filipponi, A. Pezzarossa, L. Casella, D. Smiraglia, A. Orasi, F. Attorre, A. Taramelli, Earth Observation and Biodiversity Big Data for forest habitat types classification and mapping, *Remote Sensing*, vol. 13, no. 7, p. 1231 (2021). doi: 10.3390/rs13071231
- [12] I. Dronova, P. Gong, L. Wang, L. Zhong, Mapping dynamic cover types in a large seasonally flooded wetland using extended principal component analysis and object-based

- classification, *Remote Sensing of Environment*, vol. 158, pp. 193–206 (2015). doi: 10.1016/j.rse.2014.10.027
- [13] P. Defourny, S. Bontemps, N. Bellemans, C. Cara, G. Dedieu, E. Guzzonato, O. Hagolle, J. Inglada, L. Nicola, T. Rabaute, M. Savinaud, C. Udroui, S. Valero, A. Bégué, J. Dejoux, A. Harti, J. Ezzahar, N. Kussul, K. Labbassi, V. Lebourgeois, M. Zhang, T. Newby, A. Nyamugama, N. Salh, A. Shelestov, V. Simonneaux, P. Traoré, S. Traoré, B. Koetz, 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 (2019). doi: 10.1016/j.rse.2018.11.007
- [14] S. Fritz, I. McCallum, L. You, A. Bun, E. Moltchanova, M. Duerauer, F. Albrecht, C. Schill, C. Perger, P. Havlík, A. Mosnier, P. Thornton, U. Wood-Sichra, M. Herrero, I. Becker-Reshef, C. Justice, M. Hansen, P. Gong, S. Aziz, M. Obersteiner, Mapping global cropland and field size, *Global Change Biology*, vol. 21, no. 5, pp. 1980–1992 (2015). doi: 10.1111/gcb.12838
- [15] R. Sathya, A. Abraham, Comparison of supervised and unsupervised learning algorithms for pattern classification, *International Journal of Advanced Research in Artificial Intelligence*, vol. 2, no. 2 (2013). doi: 10.14569/ijarai.2013.020206
- [16] D. Montero, G. Kraemer, A. Anghilea, C.A. Camacho, G. Brandt, G. Camps-Valls, F. Cremer, I. Flik, F. Gans, S. Habershon, C. Ji, T. Kattenborn, L. Martínez-Ferrer, F. Martinuzzi, M. Reinhardt, M. Söchting, K. Teber, M. Mahecha, Data Cubes for Earth System research: Challenges ahead, *EarthArXiv (California Digital Library)* (2023). doi: 10.31223/x58m2v
- [17] M. Schramm, E. Pebesma, M. Milenković, L. Foresta, J. Dries, A. Jacob, W. Wagner, M. Mohr, M. Neteler, M. Kadunc, T. Miksa, P. Kempeneers, J. Verbesselt, B. Gößwein, C. Navacchi, S. Lippens, J. Reiche, The OpeNEO API—Harmonizing the use of earth observation cloud services using virtual Data Cube functionalities, *Remote Sensing*, vol. 13, no. 6, p. 1125 (2021). doi: 10.3390/rs13061125
- [18] L.M. Estupiñán-Suárez, F. Gans, A. Brenning, V.H. Gutierrez-Velez, M.C. Londoño, D.E. Pabon-Moreno, G. Poveda, M. Reichstein, B. Reu, C. Sierra, U. Weber, M.D. Mahecha, A Regional Earth System Data Lab for Understanding Ecosystem Dynamics: An Example from Tropical South America, *Frontiers in Earth Science*, vol. 9 (2021). doi: 10.3389/feart.2021.613395
- [19] T. Hermosilla, M.A. Wulder, J.C. White, N.C. Coops, Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes, *Remote Sensing of Environment*, vol. 268, p. 112780 (2022). doi: 10.1016/j.rse.2021.112780
- [20] J.R.M. Flórez, I. Lizarazo, Land cover classification at three different levels of detail from optical and radar Sentinel SAR data: a case study in Cundinamarca (Colombia), *Dyna-colombia*, vol. 87, no. 215, pp. 136–145 (2020). doi: 10.15446/dyna.v87n215.84915
- [21] H. Li, J. Cui, X. Zhang, Y. Han, L. Cao, Dimensionality reduction and classification of hyperspectral remote sensing image feature extraction, *Remote Sensing*, vol. 14, no. 18, p. 4579 (2022). doi: 10.3390/rs14184579
- [22] I. Piestova, A. Kozlova, A. Andreiev, J. Rabcan Local Quality Improvement of Multispectral Imagery Classification with Radiometric-spatial Feedback. *Computer Modeling and Intelligent Systems* 2864:158–168 (2021). <https://doi.org/10.32782/cmisis/2864-14>
- [23] F. Luo, L. Zhang, B. Du, L. Zhang, Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5336–5353 (2020).
- [24] H. Huang, G. Shi, H. He, Y. Duan, F. Luo, Dimensionality reduction of hyperspectral imagery based on spatial-spectral manifold learning, *IEEE transactions on cybernetics*, vol. 50, no. 6, pp. 2604–2616 (2019).
- [25] N. Salem, S. Hussein, Data dimensional reduction and principal components analysis. *Procedia Computer Science*, 163, 292–299 (2019). <https://doi.org/10.1016/j.procs.2019.12.111>
- [26] A. Green, M. Berman, P. Switzer, M. Craig, A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1), 65–74 (1988). <https://doi.org/10.1109/36.3001>

- [27] C. Chang, Q. Du, Interference and noise-adjusted principal components analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 37(5), 2387–2396 (1999). <https://doi.org/10.1109/36.789637>
- [28] A. Hyvärinen, J. Karhunen, E. Oja, *Independent Component Analysis*. John Wiley & Sons. (2001).
- [29] D. D. Lee, H. S. Seung, Algorithms for Non-negative Matrix Factorization. In *Advances in Neural Information Processing Systems 13 NIPS 2000*, 556-562 (2001).
- [30] X. Kong, Y. Zhao, J. Chan, J. Xue, Hyperspectral image restoration via Spatial-Spectral residual total variation regularized Low-Rank tensor decomposition. *Remote Sensing*, 14(3), 511 (2022). <https://doi.org/10.3390/rs14030511>
- [31] M.O. Popov, Methodology of accuracy assessment of classification of objects on space images, *Journal of Automation and Information Sciences*, vol. 39, pp. 1–10 (2007). doi: 10.1615/JAutomat Inf Scien.v39.i1.50
- [32] V.I. Lyalko, M.A. Popov, S.A. Stankevich, J.I. Zelyk, S.V. Cherny, Calibration/Validation Test Sites in Ukraine: current state and directions of further research and development, *Ukrainian Metrological Journal*, vol. 2, pp. 15-26 (2014).
- [33] I. Shurmer, F. Marchese, J.-M. Morales-Santiago, P. P. Emanuelli, Sentinels Optical Communications Payload (OCP) Operations: From Test to In-Flight Experience, Paper Presentation, 2018 SpaceOps Conference, p. 24. doi: 10.2514/6.2018-2654
- [34] J. S. Stewart, Combining satellite data with ancillary data to produce a refined land-use/land-cover map, pp. 1-10 (1998). doi: 10.3133/wri974203
- [35] P. C. Mahalanobis, On the Generalized Distance in Statistics, *Proceedings of the National Institute of Sciences of India*, vol. 2, no. 1, pp. 49–55 (1936).