Satellite-based analysis of classification algorithms applied to the riparian zone of the Malaya Kokshaga river

L V Tarasova and L N Smirnova
Institute of Forest and Nature Management, Volga State University of Technology, 424000, Yoshkar-Ola, Mari El Republic, Russia.

E-mail: tarasoval@volgatech.net https://orcid.org/0000-0001-7768-4218

Abstract. The paper comparatively analyses the accuracy of land cover classification in the riparian zone of the Malaya Kokshaga river in the Mari El Republic of Russia using Sentinel-2A satellite images with the algorithms of supervised classification: Maximum Likelihood (ML), Decision Tree (DT) and Neural Net (NN) in the ENVI-5.2 software package. Six main classes of land cover were identified based on field studies: coniferous, mixed (deciduous), shrublands, herbaceous, and water. The assessment of the area and the structure of land cover showed that forest covers 76% of the entire territory of the riparian area of the Malaya Kokshaga river. The analysis of the results of thematic mapping shows that the overall classification accuracy obtained by the ML algorithm is 96.09%, by NN - 94.51%, and by DT - 86.54%. The producer's accuracy and user's accuracy for most classes have the maximum value when the ML algorithm is used. For the NN algorithm, the maximum value of producer’s accuracy is observed for the mixed (deciduous) class, while for the DT algorithm – for the coniferous. When classified using all three algorithms the water and bare land classes were mixed, which requires more detailed work when estimating riparian forest ecosystems.

1. Introduction
Water protection forests of river basins are complex ecosystems that serve water-regulating, sanitary-hygienic, protective-accumulative, anti-erosion, and recreational purposes. They protect the river banks from destruction, accumulate sandy alluvium in floodplains, protect steep valley slopes from erosion and landslides, thereby protecting rivers from erosion products and siltation. Riparian forest ecosystems also redirect the surface runoff from higher-lying barren areas to subsurface [1, 2] and fulfill recreation functions. In recent years, there has been a high anthropogenic impact on such ecosystems in different regions of the world, which leads to disturbances and consequent deterioration of water conservation properties [3, 4].

To assess the ongoing disturbances and to monitor riparian forests, the remote sensing data have become a priority source of information about their condition and dynamics [5, 6, 7]. Traditionally, researchers use Landsat satellite images to monitor vegetation cover due to their availability, adequate spatial resolution, as well as a large archive of time seies data [8]. The use of Sentinel high-resolution satellite data provides new opportunities for accuracy improvement of thematic mapping [9]. The data from these satellites of the European Copernicus programme have high spatial and temporal resolution and wide coverage of the territory [10]. In forestry, Sentinel products are used for various purposes such
as forest mapping [11, 12], determining forest types, setting boundaries of forest stands, analyzing vegetation indices, and assessing damage of forest cover [13].

Various algorithms based on traditional methods of classification with training (Parallelepiped, Minimum Distance, Mahalanobis Distance, Spectral Angle Mapper, Binary Encoding) are successfully used for monitoring and analysis of land cover [14]. However, the features of the study area and the quality of the selected remote sensing data affect the accuracy of thematic mapping. Researchers are trying to solve this problem by combining various classification methods [15, 16] and using algorithms based on a Decision Tree [17, 18] and machine learning (Neural Net, Random Forest, Support Vector Machine) [19, 20], as well as comparing classification algorithms [21, 22, 23].

It should be noted that satellite data are not sufficiently used to assess the state and monitoring of water-protected forests in Russia. The accuracy of classification of riparian forests based on medium and high-resolution satellite data arouses practical interest. The purpose of this study is to compare the accuracy of classification algorithms Maximum Likelihood (ML), Decision Tree (DT), Neural Net (NN) on the example of the land cover of the riparian area of the Malaya Kokshaga river in the Mari El Republic of Russian Federation using Sentinel-2A satellite data. These algorithms were selected for the analysis based on the results of the literature review and their widespread use in the classification of vegetation [24, 25, 26].

In order to achieve this goal, the following objectives have been addressed:
- the dominant types of land cover in the study area based on high-resolution images and forest inventory data have been identified;
- field studies have been carried out with the outlining of sample plots on the territory of water protection forests;
- thematic maps of the riparian zone of the Malaya Kokshaga River have been obtained from the Sentinel-2A image using three classification methods such as ML, NN, DT and spectral indices;
- the assessment and analysis the thematic mapping accuracy has been carried out by all three classification methods.

2. Materials and methods

2.1. Study location

The focus of our study is the land cover of the riparian zone of the Malaya Kokshaga river including the water conservation zone which has a width of 200m as stipulated in the Water Code of the Russian Federation (figure 1). Within the boundaries of this water conservation zone, there is a riparian protective strip where economic or other types of activities are restricted.

The Malaya Kokshaga river is the longest river of the Mari El Republic (part of its basin is located in the Kirov region) and a left tributary of the Volga river. The river originates from the village Maly Kuglanur of the Orshanka district, flows southwest along the eastern edge of the Russian Plain and feeds the Volga river. The length of the river is 194 km, its total basin area is 5,160 km² (the basin area in the republic is 4,760 km²). The main tributaries are the Oshla, Bolshaya Oshla (right tributaries) and Maly Kundysh (left tributary) rivers. According to the State Water Register of Russia, it belongs to the Upper Volga Basin.

The climate of the basin is temperate continental. The average annual precipitation is about 540 mm. Solid precipitation prevails from November to March. The snow cover lasts for about 155 days. The river is fed mainly by snow (maximum water runoff in May). The Malaya Kokshaga has an Eastern European type of water regime with high spring flood, stable winter low water and intermittent summer-autumn low water. The annual range of water level fluctuations is 5.8 m. The vegetation cover in the study area is represented by pine shrubby green moss and lichen forests, as well as floodplain forests.
Figure 1. Water protection area of the Malaya Kokshaga river on a Sentinel-2A image.

The dominant tree species in the study area are pine (*Pinus sylvestris*) and birch (*Betula pendula*). Non-dominant tree species are spruce (*Picea abies*), linden (*Tilia cordata*), oak (*Quercus robur*), aspen (*Pululus tremula*). Willow (*Salix*) and alder (*Alnus glutinosa*) are also common along the rivers [27]. In the basin of the Malaya Kokshaga river, observed on the Sentinel, there are Kokshaisk, Kuyar, experimental, suburban, and Orshanka forest enterprises of the Mari El Republic.

2.2. Research methodology
The flowchart presents step-by-step actions to classify and evaluate the accuracy mapping of three methods of supervised classification of the riparian zone of the Malaya Kokshaga river using satellite data (figure 2).
The Sentinel-2A satellite scene (RT_T38VPH_20201029T081051_4328TIF) obtained on October 29, 2020 was selected for the classification [28]. Only the second (blue, 490 nm), third (green, 560 nm), fourth (red, 665 nm) and eighth (near-infrared, 842 nm) spectral bands with a spatial resolution of 10 m per pixel were used in the study. The satellite image underwent atmospheric, radiometric and geometric corrections using the ESA SNAP software package [29].

A 200 m wide water conservation zone of the Malaya Kokshaga river was highlighted in the image on both sides of the river course in the Arcmap 10.3 software package according to the Rosreestr public cadastral map. To do this, the PKK6_Zones.lyr file was uploaded to Arcmap 10.3 from the Rosreestr website [30], which was converted into several bitmaps. Further, these bitmaps were converted into shapefiles (*.shp) and merged into a single file. The area under study also comprises the Malaya Kokshaga river and urbanized areas including roads along its course.

**Figure 2.** Algorithm for classification and evaluation the accuracy of thematic maps.
In 2020-2021, fieldwork was carried out to establish the sample plots and examine them. They were established on the base of circular relascopic plots. At the preparatory stage of sample plot selection, we carefully analysed the existing maps of forest stands, forest inventory, and forest plans of the Mari El Republic. High-resolution satellite images (Yandex, Google Earth Internet resources, SAS Planet) were used as supplementary sources. Based on a detailed analysis of the available data, the study identified six dominant thematic classes of land cover (table 1) and selected the most represented areas for these classes. When selecting sample plots, the main condition was their representativeness in all classes of land cover and even distribution over the study area.

The geographical coordinates of each sample plot were recorded onsite using the GARMIN eTrex 20 GPS receiver in order to identify them on satellite images. The data on the area, species composition and the main dendrometry indicators of the dominant species of the stand (average height, diameter, age) were recorded and linked to the adjacent plots. As a result of fieldwork, 50 sample plots were established and studied: 30 sample plots were used for classification and 20 - for accuracy estimation.

### Table 1. Dominant classes in the area under study.

| Type no | Land cover types       | No of ROI (total) | No of pixels |
|---------|------------------------|-------------------|--------------|
| 1       | Coniferous             | 52                | 522          |
| 2       | Mixed (deciduous)      | 65                | 707          |
| 3       | Shrublands             | 64                | 293          |
| 4       | Herbaceous             | 51                | 344          |
| 5       | Bare land              | 71                | 382          |
| 6       | Water                  | 163               | 1,312        |

After establishing a region of interest (ROI) for each type of land cover in ENVI 5.2, their spectral separability was evaluated on the Sentinel-2A satellite images using the Jeffries-Matusita (JM) method (a measure of inter-cluster distance), which is calculated by comparing each pair of sample plots [31, 32]. The values of statistical separability of the selected ROI can vary from 0 to 2, and the index value above 1.4 indicates suitable spectral separability of the studied types of land cover [33]. The results of the spectral separability estimation are presented in table 2. The analysis of pairwise separability using the JM method indicates high separability for the considered ROIs, which have the values of statistical separability above 1.4.

### Table 2. Separability of the studied types of land cover in accordance with the Jeffries–Matusita method.

|        | Coniferous | Mixed (deciduous) | Shrublands | Herbaceous | Bare land | Water |
|--------|------------|-------------------|------------|------------|-----------|-------|
| Coniferous | 1.98       | 1.97              | 1.99       | 1.99       | 1.99      |       |
| Mixed (deciduous) |           | 1.72              | 2.00       | 1.92       | 1.92      |       |
| Shrublands            |           | 1.73              |            | 1.69       | 1.98      |       |
| Herbaceous            |           |                   |            |            | 1.94      | 1.99  |
| Bare land             |           |                   |            |            |           | 1.97  |

We carried out a supervised classification of the area under study based on the algorithms of the DT, ML, and NN in the ENVI-5.2 software package. When carrying out the classification by the DT method, various indicators characterizing the land cover can be used. For this purpose, we have identified and analyzed seven spectral indices: \( \text{VI}_{\text{GREEN}} \) (Vegetation Index Green), \( \text{DVI} \) (Difference Vegetation Index), \( \text{NDVI} \) (Normalized Difference Vegetation Index), \( \text{NDWI} \) (Normalized Difference Water Index), \( \text{OSAVI} \) (Optimized Soil Adjusted Vegetation Index), \( \text{MSAVI2} \) (Modified Soil Adjusted Vegetation Index 2), \( \text{EVI} \) (Enhanced Vegetation Index). The analysis of the results shows that land cover classes under
study have close and sometimes overlapping ranges of index values. Therefore, only four indexes were selected for the further supervised classification by the DT method: NDVI, NDWI, VIgreen and DVI (figure 3).

![Decision tree](image)

**Figure 3.** Decision tree used to classify water protection forests.

In order to carry out classification according to the ML and NN methods, sample plots were established using ROI for each land cover type (table 1). Classification parameters typical for these methods were used in the study.

The ENVI 5.2 NN uses a standard error backpropagation algorithm (an iterative gradient algorithm used to minimize the standard deviation of the current output of the multilayer perception and the desired output). Neurons are divided into groups with a common input signal, i.e. layers. All elements of the external input signal are fed to each neuron of the first layer (Hidden Layer). All outputs of the neurons of the nth layer are fed to each neuron of the n+1 layer. Neurons perform a weighted summation of input signal elements with the addition of neuron displacement. Nonlinear priming is performed on the summation result – the activation function. The value of this function is the output of the neuron. When classifying by NN, we used the default parameters in the ENVI 5.2 programme: Activation - Logistic, Training Threshold Contribution - 0.9, Training Rate - 0.2, Training Momentum - 0.9, Training RMS Exit Criterion – 0.1, Number of Hidden Layers - 1, Number of Training Iteration – 1000.

Evaluation of the accuracy of the obtained thematic maps was carried out in the Accuracy Estimation utility module of the ENVI-5.2 software package. Sets of pixels (ROI) with the main legend types were used as sample plots. They were allocated according to the data of the forest management of the Ministry of Forestry of the Mari El Republic and field data. The resulting thematic maps are shown in figure 4. As a result, we obtained a Confusion Matrix with the main indicators (statistics) used when assessing the accuracy of thematic maps [34]: general classification accuracy, Kappa coefficient; producer’s accuracy; user’s accuracy, commission error, and omission error.
3. Results and discussions

The data for assessing the accuracy of thematic maps obtained by three classification methods from the Sentinel-2A image are presented in table 3. Analysis of the results of thematic mapping shows that the overall classification accuracy carried out using ML 96.09% was higher than using DT and NN. At the same time, the Kappa coefficient was 0.94 for ML, 0.83 for DT and 0.93 for NN. The PA indicator for coniferous, shrublands, herbaceous, water body reaches the maximum value using ML. When using NN for mixed (deciduous) type, the PA indicator amounts to 98.34%. For the bare land type, PA has the same value of 94.05% when estimated by the ML and NN methods. The UA or coniferous, mixed (deciduous), herbaceous, bary land, and water shows the maximum values when estimated by ML, and for shrublands class - by DT.

When classification is carried out using all three methods, we observed mixing of shrublands class with mixed (deciduous), herbaceous, and bare lands. The water class is mixing with the bare land. This can be explained by the fact that the river course of the Malaya Kokshaga often has narrowings, as a result, the pixels of the water class have the same spectral characteristics as the pixels of the bare land, which can be attributed to shadowed areas. Various approaches can be used to avoid these errors in the classification of water protection forests. For example, using high spatial resolution images, shadow masks, raster with a mask of water class.

![Thematic maps of the riparian zone of the Malaya Kokshaga river, obtained by various classification methods: a) DT, b) ML, and c) NN.](image)

**Figure 4.** Thematic maps of the riparian zone of the Malaya Kokshaga river, obtained by various classification methods: a) DT, b) ML, and c) NN.
Table 3. Accuracy estimation of thematic maps.

| Land cover types | Producer's Accuracy (PA, %) | User's Accuracy (UA, %) |
|------------------|-----------------------------|-------------------------|
|                  | ML  | DT  | NN  | ML  | DT  | NN   |
| Coniferous       | 99.39 | 96.79 | 96.95 | 99.79 | 84.56 | 99.58 |
| Mixed (deciduous) | 97.23 | 91.65 | 98.34 | 92.72 | 83.29 | 91.38 |
| Shrublands       | 87.83 | 54.61 | 82.54 | 76.15 | 76.15 | 74.64 |
| Herbaceous       | 92.53 | 79.53 | 85.41 | 97.74 | 97.14 | 96.00 |
| Bare land        | 94.05 | 66.49 | 94.05 | 96.23 | 70.56 | 95.40 |
| Water            | 96.72 | 94.72 | 95.25 | 99.62 | 92.66 | 99.39 |
| Overall accuracy |     |     |     | ML-96.09 %, DT-86.54%; NN-94.51% |     |     |
| Kappa            |     |     |     | ML - 0.94 ; DT-0.83; NN-0.93 |     |     |

Thus, the ML algorithm proved the highest accuracy. Therefore, further classification of the land cover of the water protection area of the Malaya Kokshaga river was carried out by this method. The results of the assessment of the area and the structure of the land cover after classification are presented in table 4. The largest area under study is covered with mixed (deciduous) (41.24%) and shrublands (25.44%) classes, which is typical for the vegetation cover of the riparian zone. The smallest class represented in the water protection zone is herbaceous (4.81%).

Table 4. Land cover classes in the water protection area.

| Land cover classes | No of pixels | Area, ha | % of the total area |
|--------------------|--------------|----------|---------------------|
| Coniferous         | 60,130       | 5,411.70 | 9.69                |
| Mixed (deciduous)  | 255,911      | 23,031.99| 41.24               |
| Shrublands         | 157,852      | 14,206.68| 25.44               |
| Herbaceous         | 29,877       | 2,688.93 | 4.81                |
| Bare land          | 33,405       | 3,006.45 | 5.38                |
| Water              | 83,359       | 7,502.317| 13.43               |
| Total              | 620,534      | 55,848.06| 100                 |

4. Conclusion

As a research result, we have developed the classification of water protection forests along the Malaya Kokshaga river using three algorithms: ML, NN and DT. Comparison of the results of thematic mapping of the riparian land cover according to Sentinel-2A data showed that the ML classification method proved to be the most accurate (96.09%). Despite the fact that in this study, the accuracy of classification by the DT method demonstrates a lower accuracy (86.54%) as compared with the ML and NN methods, this algorithm also has great potential for mapping water protection forests. The use of more suitable parameters (vegetation indices, soil indicators, topographic data, etc.) can improve the accuracy of classification carried out using DT. The NN method also provides rather accurate thematic maps. The change of standard classification parameters such as Activation, Training Threshold Contribution, Training Rate, Training Momentum, Training RMS Exit Criteria, Number of Hidden Layers, Number of Training Iteration, Min Output Activation Threshold can improve the accuracy of classification carried out using this method. The vegetation cover classified as coniferous, mixed (deciduous) and shrubland reaches 76% of the entire territory of the riparian zone of the Malaya Kokshaga river. Such distribution of classes indicates the current sustainability of the ecosystem under study. Further studies of the riparian zone of water-protected forests using satellite images can be aimed at identifying the dynamics and assessing the disturbance of its cover, as well as determining the main factors which influence these changes (illegal logging, fires, pests).
Acknowledgments
This research is funded by GIS and Remote Sensing for Sustainable Forestry and Ecology/SUFOGIS project of the Erasmus+. The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

References
[1] Kousky C and Walls M 2014 Ecol. Econ. 104 119-128
[2] Hasselquist E M et al. 2021 Forest Ecol. Manag. 493 119254
[3] Vorobiev O N and Kurbanov E A 2021 Curr. Probl. Remote Sens. Earth Space 18 214-225
[4] King D J, Harley G L, Maxwell J T, Heeter K J, Vandermyde B J and Cosgriff R J 2021 Forest Ecol. Manag. 496 119454
[5] Ongong A, Krhoda G and Mukhovi S 2020 Int J Agric Environ and Bioresearch 5 68-85
[6] El-Tantawi A M, Bao A, Chang C and Liu Y 2019 Environ. Monit. Assess. 191 480
[7] Pontes P R M, Cavalcante R B L, Sahoo P K, Silva Júnior R O, Silva M S, Dall’Agnol R and Siqueira J O 2019 J. Environ. Manage. 235 pp 489-99
[8] Kogoa B K, Kumara L and Koechb R 2017 RSASE 14 pp 75–83
[9] Kurbanov E A, Vorob’ev O N, Men’shikov S A and Smirnova L 2018 Curr. Probl. Remote Sens. Earth Space 15 pp 154–66
[10] Phiri D, Simwanda M, Salekin S, Nyirenda V R, Murayama Y and Ranagalage M 2020 Remote Sens. 12 2291
[11] Szostak M, Hawylo P and Piela D 2018 Eur. J. Remote Sens. 5 142–49
[12] Ottosen T B, Petch G, Hanson M and Skjøth C A 2020 Int. J. Appl. Earth Obs. 84 101947
[13] Bogdanov A P and Aleshko R A 2017 Problemy ekologicheskogo monitoringa i modelirovaniya ekosistem 28 98–110
[14] Obeidat M, Awawdeh M and Lababneh A 2019 Arab J Geosci 12 685
[15] Mleczko M and Mróz M 2018 Remote Sens. 10 78
[16] Popov M, Michaelides S, Stankevich S, Kozlova A, Piestova I, Lubskiy M, Titarenko O, Svideniuk M, Andreiev A and Ivanov S 2021 RSASE 22 100477
[17] Mendez-Estrella R, Romo-Leon J R, Castellanos A E, Gandarilla-Aizpuro F J and Hartfield K 2016 Remote Sens. 8 664
[18] Garcia A S, Vilela V M N, Rizzo R, West P, Gerber J S, Engstrom P M and Ballester M V R 2019 RSASE 15 100233
[19] Zhang Y, Ling F, Wang X, Foody G M, Boyd D S, Li X, Du Y and Atkinson P M 2021 Remote Sens. Environ. 261 112470
[20] Feng Q, Yang J, Zhu D, Liu J, Guo H, Bayartungalag B and Li B 2019 Remote Sens. 11 1006
[21] Ngo K D, Lechner A M and Vu T T 2020 RSASE 17 100272
[22] Rana V K and Suryanarayana T M V 2020 RSASE 19 100351
[23] Xie Z, Chen Y, Lu D, Li G and Chen E 2019 Remote Sens. 11 164
[24] Sun J, Yang J, Zhang C, Yun W and Qu J 2013 Math. Comput. Model. 58 573-81
[25] Pouliot D, Latifovic R, Fernandes R and Olthof I 2009 Remote Sens. Environ. 113 1749-59
[26] Sothe C et al. 2020 Gisci Remote Sens 57 369-94
[27] Kurbanov E, Vorobyev O, Leznin S, Polevshikova Y, Demisheva E 2017 Int. J. Wildland Fire 26 772-782
[28] Copernicus Open Access Hub URL: https://scihub.copernicus.eu/dhus/#/home
[29] Sentinel-2 Toolbox URL: https://step.esa.int/main/toolboxes/sentinel-2-toolbox/
[30] ArcGIS REST Services Directory (Rosreestr) URL: https://pkk.rosreestr.ru/arcgis/rest/services/PKK6/Zones/MapServer
[31] Huang H, Roy D P, Boschetti L, Zhang H K, Yan L, Kumar S S, Gomez-Dans J and Li J 2016 Remote Sens. 8 873
[32] Kurbanov E A, Vorobeiv O N, Menshikov S A and Ali M S 2019 Proc. Int. Conf. Earth observation advancements in a changing world (Florence, Italy) vol 1 p 33–36
[33] Chuvieco E 2020 Fundamentals of Satellite Remote Sensing: An Environmental Approach (Florida: CRC Press) p 468
[34] Vorob'ev O N, Kurbanov E A, Demisheva E N, Men'shikov S A and Smirnova L N 2019 Bulletin of the Volga State Technological University Ser: Forest Ecology Nature management 41 5-20