Digital technologies for forest monitoring in the Baikal natural territory

Igor V. Bychkov, Gennady M. Ruzhnikov, Roman K. Fedorov, Anastasia K. Popova

Matrosov Institute for System Dynamics and Control Theory, Siberian Branch of Russian Academy of Sciences, Lermontov st. 134, Irkutsk, Russia

Abstract

The paper considers the problem of forest resources monitoring over large areas on the example of the Baikal natural territory. As the main data source, we use Sentinel-2 remote sensing data due to their regularity, broad coverage, multispectral parameters of the resulting image. The Random forest and Support Vector Machines (SVM) machine learning algorithms were used to classify land cover from the Sentinel-2 products. Both methods have shown good results with a fairly high accuracy. The training was carried out with data labeled manually into 12 classes.

Keywords

Machine learning, remote sensing, forest monitoring

1. Introduction

Forest monitoring is an assessment and forecasting system for the forest fund state in space and time for the rational use, protection and reproduction of forests, increasing their ecological functions. Monitoring supports tracking the forest resources dynamics caused by forest management, natural and anthropogenic impacts, compiling predictive and analytical models for their protection and use, sustainable development of forest economics. The effectiveness of forest monitoring directly depends on the completeness and accuracy of observations data of various environment elements.

The Baikal Natural Territory (BNT) covers Lake Baikal, the water protection zone around it, specially protected natural areas and adjacent areas 200 km to the west and northwest from the lake. The area of the BNT is 386 thousand km², there are 31 specially protected natural areas, including 3 reserves, 2 national parks, 6 recreational areas and more than 128 natural monuments.

Among the most important natural resources of the BNT are forest resources, which ensure the sustainability of environment, performing water and soil protection, water regulation functions. The area of BNT lands covered with forest vegetation is about 8350.73 thousand hectares and 92% of these lands are covered by forests, represented by two groups of forest-forming species: coniferous and deciduous trees. BNT forests are negatively affected by fires, forest diseases, insect pests, unfavorable weather conditions, which can lead to the loss of forest biological stability.

Forest monitoring of the BNT has poor efficiency and limited access to in-situ data, which complicates the support of decision-making and the conduct of interdisciplinary research. Official forest inventory information is not always up-to-date, there is no unified system for storing and processing forest monitoring data at the regional level [1].

This determines the relevance of forest monitoring in a digital format, which is essential for sustainable forest management in the BNT and compliance with the requirements of continuous, rational use of forests, their reproduction, and conservation of resource, recreational, ecological potential and biological diversity.

2. Organization of digital monitoring

The advantages of the digital forest monitoring system of the BNT are a large number of participants and their information resources; the participants' emphasis on their strengths and the transfer of non-core activities and services to outsourcing to others; efficiency of updating open digital information resources, attraction of scientific knowledge. Planning and implementation of several services provided by different participants, including third-party ones, reducing the cost of obtaining services helps to increase the complexity and validity of management decisions.

Digital forest monitoring of the BNT is based on an information and analytical environment that provides collection, transmission, search, storage of spatio-temporal data from forest monitoring, the ability to assess, model and forecast the state of forest resources of the BNT [2-3]. Such an environment should contain spatial and thematic data of forest monitoring, including remote sensing data, unified reference books and classifiers. A catalog of services is intended for processing monitoring data: providing data, assessing forest dynamics, machine learning, publishing results in the form of maps and diagrams. The scheme of such digital forest monitoring system is shown in the figure 1.
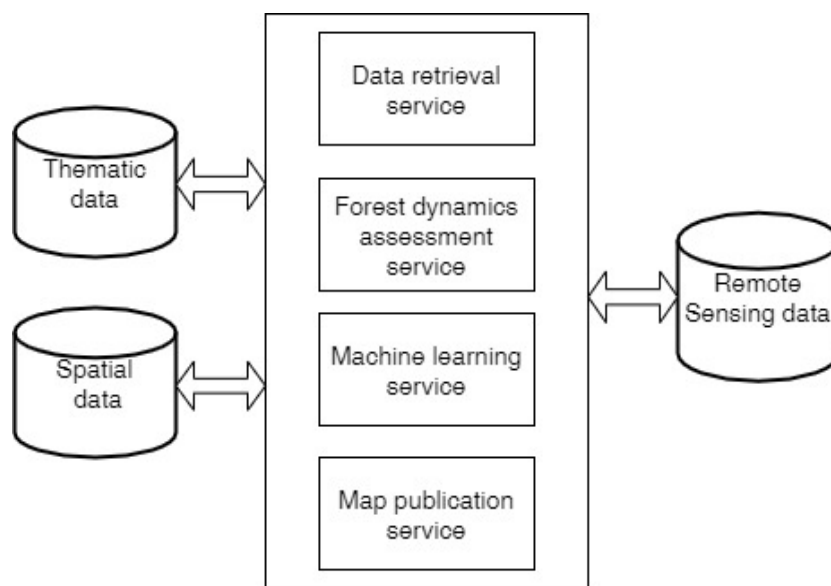
Figure 1: Scheme of a digital forest monitoring system

Remote sensing is an important source of data for digital forest monitoring [4-5]. Traditional forest in-situ data are not always able to provide up-to-date data for a large area, such as BNT, they are often based on a sample of small areas, or contain aggregated information without an accurate spatial description. Remote sensing data creates opportunities to obtain forest data in a more efficient way, provides information about their spatial species distribution with wide temporal coverage and higher refresh rates.

Medium resolution satellite imagery such as Landsat and Sentinel-2 allow map large areas in economical manner. The resolution of 10 m in the main Sentinel-2 bands allows to detect a number of forest parameters quite accurately, making them more preferable than Landsat images with a 30 m resolution. Sentinel-2 satellites are equipped with MultiSpectral Instrument with 13 spectral bands, covering channels from blue to short wave infrared (SWIR) with a resolution of 10 to 60 m. Provides global coverage on average every 5 days.

Intelligent analysis of remote sensing data provides an opportunity to identify changes in the forest fund as a result of anthropogenic impacts and environmental disturbances, fires, damage to forests by pests, diseases, windblows, etc. At the initial stage, it is necessary to classify satellite images for the study area by compiling land cover maps. In this study, we use machine learning methods for automated land cover classification.

3. Results and discussion

Machine learning algorithms are used to process large datasets consisting of multi-temporal images with spectral metrics [6-7]. These methods are effective for classifying complex multidimensional data, providing nonlinear and nonparametric classifications. The most popular machine learning algorithms used in remote sensing research are Random Forest (RF) and Support Vector Machines (SVM).

RF is a nonparametric ensemble machine learning algorithm based on decision trees. It can process various data such as satellite images and numerical data. Each decision tree produces a classification result for samples not selected as training samples. The decision tree chooses some class, and the final class is determined by the highest number of votes.

SVM is a machine learning method developed based on the theory of statistical learning and the principle of minimizing structural risks. Compared with traditional teaching methods, it has high accuracy, fast computation speed, and strong generalizability, which is widely used in image mapping and land classification.

Study area in this research covers south of the lake Baikal. Little cloudy Sentinel-2A MSI granules used in the study were freely acquired on 25 June 2017 and downloaded from the Copernicus Scientific Data Hub as a Level-1C product. Figure 2 present Sentinel-2 image RGB composite for the study area.
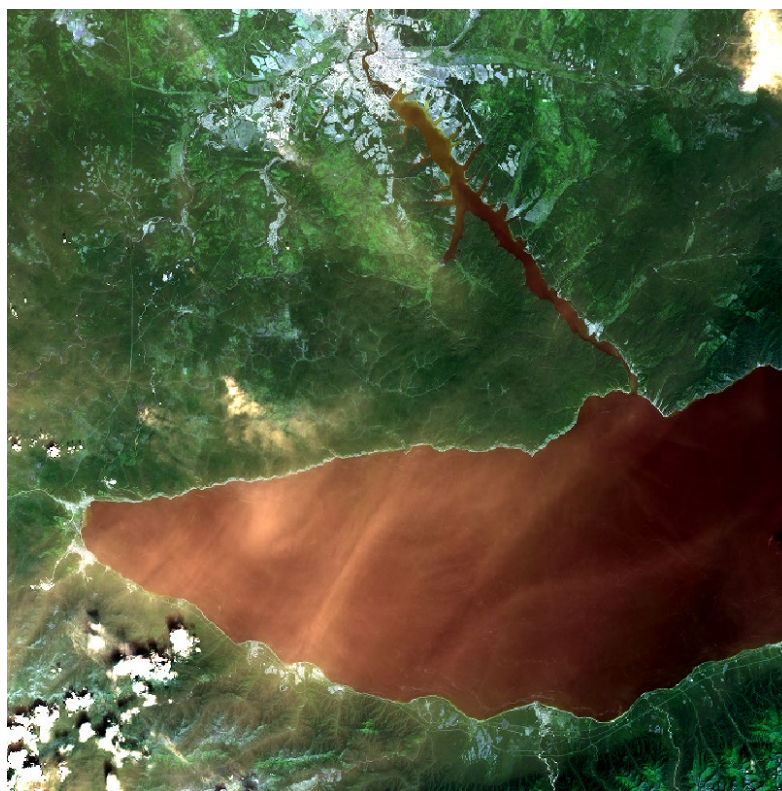
Figure 2: Sentinel-2 image for study area

To monitor the forest resources of the BNT, a part of the study area was previously labeled manually with 12 classes: felling, shrubs, coniferous forest, woodland, deciduous forest, mixed forest, rocks, pastures, arable land, residential area, clouds, water. For this step, polygon-shaped samples were generated based on visual interpretation of the high resolution satellite images and expert knowledge.

All 13 bands of the Sentinel-2 image were used for training. The labeled sample was randomly divided into training and test parts in a proportion 70/30. We used implementations of machine learning algorithms from the "scikit-learn" Python library. The Random forest method uses 200 decision trees, SVM parameters are: kernel = "linear", C = 1.0. The estimation of the accuracy of the

algorithms is given in the table 1. We used macro average values for precision, recall, and f1-score parameters.

Table 1

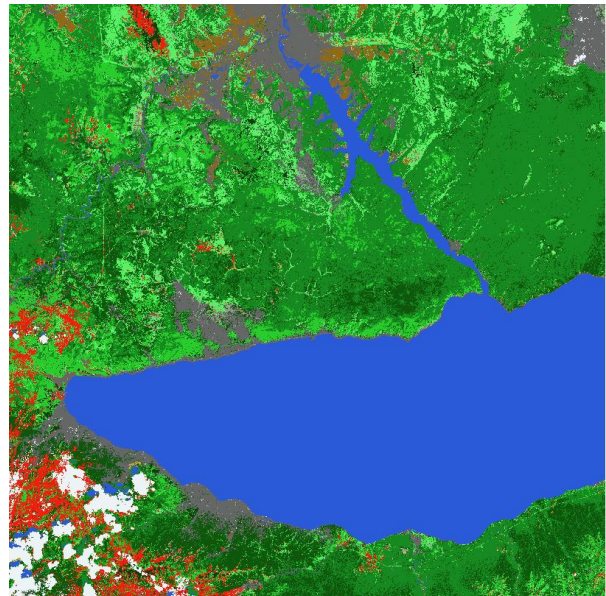Accuracy assessment for the classification algorithms

Algorithm	Precision	Recall	F1-score
RF	98.92%	0.95	0.95
SVM	93.79%	0.89	0.87

The Random forest made the most mistakes in the Lightwood and Deciduous Forest classes. SVM misclassifies Pasture, Woodland, and Deciduous Forest classes. Errors occurred due to the similarity of spectral characteristics in the classes associated with vegetation. In the future, it is necessary to expand the training dataset, filling it with a large number of samples of different forest species.

The classification results are presented in Figures 3 and 4. It is visible that the Random forest method misclassifies "Living area" class (highlighted in gray). The SVM method also misclassifies the "Logging" class (lower left corner of the map, highlighted in red). At the same time, the value of the accuracy of both algorithms in these classes is quite high (96-97%), which can be explained by the insufficient size of the test sample, on which the accuracy was assessed.

Figure 3: Result of the RF   **Figure 4**: Result of the SVM

In general, Random forest showed the best result in the classification of remote sensing data for BNT. To improve the calculations of the SVM algorithm, more complex non-linear kernel in the method parameters can be used.

4. Conclusion

Effective management of forest resources is impossible without full and timely information about their condition. Remote sensing images are well suited for regular monitoring of large areas due to their high repeatability, wide coverage, and easy accessibility. 13 spectral bands of Sentinel-2 satellites images are quite enough to distinguish various tree species.

The work compares the results of two machine learning algorithms - Random forest and SVM - to classify the land cover. The test array with 12 classes on the BNT area training samples were generated and labeled manually. The learning showed high results: 98.92% OAA for Random forest and 93.79% for SVM. The main calculation errors are associated with an insufficient number of test

samples, which does not allow the methods to separate accurately from each other classes with similar spectral characteristics. We plan to expand sample dataset to improve the classification results.

The resulting classification of land cover can be used for BNT forest monitoring. Fast tracking of logging, burnt-out areas, and reforestation will allow to assess the forest resources dynamics and to make management decisions.

5. Acknowledgements

6. References

[1] Anastasia K. Popova, Evgeny A. Cherkasin, Igor N. Vladimirov, Forest Resources of the Baikal Region: Vegetation Dynamics Under Anthropogenic Use, Information Technologies in the Research of Biodiversity, Springer Proceedings in Earth and Environmental Sciences, Springer, Cham (2019) 96-106. doi:10.1007/978-3-030-11720-7_14.

[2] I. V. Bychkov, G. M. Ruzhnikov, R. K. Fedorov and A. K. Popova, Digital platform for forest resources monitoring in the BAIKAL natural territory, Journal of Physics: Conference Series, Volume 1864, 13th Multiconference on Control Problems (MCCP 2020) 6-8 October 2020, Saint Petersburg, Russia, J. Phys.: Conf. Ser. 1864 012111 (2020) doi:10.1088/1742-6596/1864/1/012111.

[3] I. V. Bychkov, G. M. Ruzhnikov, R. K. Fedorov, A. E. Khmelnov and A. K. Popova, Organization of digital monitoring of the Baikal natural territory, IOP Conference Series: Earth and Environmental Science, Volume 629, Environmental transformation and sustainable development in Asian region 8-10 September 2020, Irkutsk, Russian Federation, IOP Conf. Ser.: Earth Environ. Sci. 629 012067 (2021) doi:10.1088/1755-1315/629/1/012067.

[4] J. Lastovicka, P. Svec, D. Paluba, N. Kobliuk, J. Svoboda, R. Hladky, et al., Sentinel-2 data in an evaluation of the impact of the disturbances on forest vegetation, Remote Sens 12:1914 (2020). doi:10.3390/rs12121914.

[5] L. Soleimannejad, S. Ullah, R. Abedi, M. Dees, B. Koch, Evaluating the potential of Sentinel-2, Landsat-8, and IRS satellite images in tree species classification of hyrcanian forest of Iran using Random forest, J Sustain Forest (2019) 38:615–28. doi:10.1080/10549811.2019.1598443.

[6] E. Grabska, D. Frantz, K. Ostapowicz, Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians, Remote Sens Environ (2020) 251:112103. doi:10.1016/j.rse.2020.112103.

[7] C. Musi, S. Anggoro, Sunarsih, System dynamic modelling and simulation for cultivation of forest land: Case study Perum Perhutani, Central Java, Indonesia, J Ecol Eng (2017) 18:25–34. doi:10.12911/22998993/74307.