Vegetation cover analysis of the mountainous part of north-eastern Siberia by means of geoinformation modelling and machine learning (basic principles, approaches, technology and relation to geosystem science)

Moisey Zakharov[1,3], *Mikhail* Cherosov[2*], *Elena* Troeva[2], and *Sebastien* Gadal[1,3]

[1]Institute of Natural Sciences, North-Eastern Federal University, Yakutsk, Russia
[2]Institute for Biological Problems of Cryolithozone, SB RASciences, Yakutsk, Russia
[3]Aix-Marseille University, UMR ESPACE CNRS, Aix-en-Provence, France

Abstract. For the first time, the geoinformation modelling and machine learning approaches have been used to study the vegetation cover of the mountainous part of North-Eastern Siberia – the Orulgan medium-altitude mountain landscape province. These technologies allowed us to distinguish a number of mapping units that were used for creation and analysis of 1:100 000 scale vegetation map of the interpreted key area. Based on the studies, we decided upon the basic principles, approaches and technologies that would serve as a methodology basis for the further studies of vegetation cover of the large region. Relief, slope aspect, genetic types of sediments, and moisture conditions were selected as supplementary factors to the vegetative indices for differentiation of both plant communities and vegetation map units.

Achieving reliable and unbiased data on spatial organization of the environment and vegetation diversity of any region of both Russia and the world is the primary task of regional biogeographical research activity.

Present-day vegetation mapping should involve the adapted approaches of geoinformation modelling and large amount of remote sensing data for machine learning and the multi-factor analysis of the classified vegetation cover. Landscape modelling process implies knowledge and data collection and processing them into an information complex. This approach underlies the methodology in geoinformatics and remote sensing technologies [1, 2]. Vegetation science (geobotany), having similar ideology, should follow it as well.

The study aimed to elaborate the general approach of geoinformation modelling and machine learning for the study of vegetation cover of the frozen landscapes of mountainous provinces of North-Eastern Siberia (the case study of the Orulgan middle-altitude mountain landscape province).

---

[*]Corresponding author: cherosov@mail.ru

We used the ASTER GDEM in GeoTIFF format with 30-m resolution, a series of Landsat 8 OLI and Sentinel 2 satellite images, vegetative indices (NDVI and GNDVI) and multi-component land cover reclassification for the recognition of vegetation units in QGIS environment with the following modules: Semi-Automatic Classification, Minimum Distance, Maximum Likelihood, and Spectral Angle Mapper (main). Additionally such software as GRASS GIS, IlWIS GIS, and Orfeo Toolbox were used with a number of machine learning algorithms (Support Vector Machine, Random Forest), SAGA GIS (to work with the digital elevation model), as well as TerrSet with the tools of IDRISI GIS Analysis and IDRISI Image Processing, and a number of vertical applications. We also used the Google cloud infrastructure Google Earth Engine (GEE) that provides free access to the big satellite imagery dataset with possibility to make a query using various filters.

We performed the geoinformation modeling according to [1-5] and the user manuals.

The approach of frozen landscapes geoinformation modelling was improved for more precise mapping purposes. It features the higher level of automatic classification of vegetation units and speeds up the mapping process significantly.

For vegetation mapping of the large territory of the Orulgan middle-altitude mountain province we analyzed the temporal series of 4500 Landsat 8 and Sentinel 2 images. Processing was conducted in GEE environment. This Google database comprises 2888 Sentinel 2 and 1249 Landsat 8 OLI datasets to cover the vegetation period of the studied territory.

Table 1 represents the criteria to derive vegetation mapping units based on the geoinformation modelling of landscape components. Those factors that determine the plant community formation – relief, slope aspect, sediment genetic types, and moisture conditions – also were used for vegetation cover differentiation on the map supplementing the analysis of vegetative indices.

Table 1. Criteria for distinguishing the vegetation mapping units based on landscape geoinformation modelling.

Vegetation	Typological landscape complexes	Relief	Genetic stratigraphic sequences	Vegetation examples
Biome	Landscape type (subtype)	Macro-relief	Genetic type of sediments	Mountain sparse forests
A complex of geobotanical categories of various level	Locality type	Meso- and macro-relief	Genetic type of sediments	*Larix* sparse forests and open stands of lichen and green-moss types
Association groups, plant community complexes (phytocoenoc hora)	Terrain type	Micro- and meso-relief	Genetic type of sediments	*Vaccinium vitis-idaea* lichen-green moss *Larix* sparse forests

The key area to study the spatial structure of the Orulgan middle-altitude mountain landscape province is situated in Eveno-Bytantaysky national ulus (municipal region), in the basin of the Bytantay River and its tributaries the Ulakhan-Sakkyryr, Achchygyy-

Sakkyryr, and Tumara. It stretches between N 68 ° 24.1 '– 67 ° 34.3' and E 129 ° 14.5 ' – 131 ° 32.0' with the total area of ca. 6000 km^2.

We created 387 training data and 550 validation points for the studied area. The combined geoinformation analysis of both satellite imagery and digital elevation model allowed us to distinguish the map units at the level of plant association groups (we used the ecological-phytocoenotic approach of vegetation classification, though other approaches are possible to be used in further works) and meso-relief.

Machine learning (ML) can be generally considered as a class of artificial intelligence intended for design, development and application of algorithms and methods that allow computers to learn.

We generated the training points by means of the random point instrument and classified them based on visual interpretation by the experts in geobotany of Google Earth images of high resolution and vegetation types published in [6]. The validation data were defined in the same way.

We compiled the database of interpreted satellite images in RGB-format. It includes: original B, G, R, and NIR bands; and such vegetation indices (VI) as NDVI and GNDVI. The best result was achieved when using the CVM algorithm for classification purposes (table 2). The processed spatial data yielded 7 vegetation categories: dwarf shrub - green moss - lichen *Larix* sparse forests, lichen *Larix* sparse forests, *Sphagnum Larix* open stands with peat bogs, *Salix* and *Duschekia* communities with patches of *Larix*, *Populus* and *Chosenia* forests, dwarf shrub and lichen mountain tundra, dwarf shrub (*Dryas*) mountain tundra, epilithic lichen communities and river pebble.

Table 2. Vegetation classification results using various ML algorithms.

Number of training points	Number of validation points	Learning algorithm	Overall accuracy, %
387	550	SVM	81.2
		ML	72.7
		MD	56.4
		SAM	23.4

Having verified the approach in the studied key area (vegetation interpretation, spatial structure analysis, substantiation of the distinguished mapping units), we applied it for the whole province.

As a result, the cloud infrastructure Google Earth Engine and remote sensing data allowed us to achieve more precise information on the spatial structure of the Orulgan middle-altitude mountain province. The analysis revealed the predomination of mountain tundra and mountain desert vegetation throughout the whole province. It is interesting that its portion reduced from 70% on the frozen landscapes map to 43% on our map, while the portion of the northern mountain *Larix* sparse forests increased from 16.5% to 42%, especially in valley complexes and on western macroslope. It proves the relevance to attribute the province to the group of mountain tundra and mountain sparse forest nature complexes in the area of continuous distribution of perennially frozen grounds according to the frozen-landscape regionalization.

The VI values allowed us to reveal both homogenous (phytocoenomera) and heterogenous (phytocoenochora) territorial contours for each mapped category.

In total, 9 mapping units were distinguished for the province's territory represented by phytocoenochoras (groups of plant associations) and the higher rank units (formations and groups of formations).

GIS modeling clearly displayed the combination of *Eriophorum* bogs and *Sphagnum Larix* open stands. On the other hand, the *Pinus pumila* shrub belt is not discerned well due to mixing with *Larix* sparse forests. E.G. Nikolin, the expert in mountain flora and vegetation, confirms this fact questioning the distinction of the mountain shrub belt in the western part of the Verkhoyansk Range. This altitudinal belt is more or less clearly expressed in the Central and Eastern Verkhoyansk Range [7]. In our case study, we regarded this mapping unit as the dwarf shrub - lichen *Larix* sparse forests with *Pinus pumila* in combination with *Duschekia fruticosa* and fruticose *Betula* shrubs. The groups of plant associations are characteristic for the slopes of southern aspect. Their upper part is covered with *P. pumila* stands that down the slope are gradually substituted by *Larix* forests with increasing canopy closure. The forests gradually transfer to valley complexes covered with *Duschekia fruticosa* and fruticose *Betula* shrubs ('yernik'). The map clearly reflects the latitudinal vegetation distribution patterns, when the *Larix* sparse forests increase in their area from the north southwards. The latitudinal patterns are expressed as a significant distinction between the eastern and western macroslopes. The western macroslope is influenced by the warming effect of the Lena River and warm air masses from the central part of Yakutia. The eastern macroslope is more exposed to cold air masses from the Arctic Ocean.

The epilithic lichen communities of fell-fields (including bare rocks) are characteristic for the near-to-summit parts of mountain ridges, increasing in their areas northward. The ridges in the southern part of the province are mainly covered with mountain lichen and dwarf shrub - green moss tundra. The map analysis proves the general latitudinal pattern of reducing the portion of the forest and mountain shrub belts northward and, on the contrary, increasing areas with mountain tundra and epilithic vegetation.

The discerned valley complexes in mountain regions feature the expressed intrazonality of vegetation. This shows the significance of paragenetic and paradynamic relations in a landscape.

The conducted study, as well as the experience of the authors, made possible to shape the concept of using the geoinformation modelling for vegetation cover studies of physiographically similar mountainous territories both of Siberia and the Russian Federation as a whole. The concept includes basic principles, approaches, technologies, as well as the defined set of main factors of vegetation differentiation (relief, slope aspect, genetic types of sediments, and moisture conditions). Our plans are to apply the approach for such regions as the Salair ridge (South Siberia), the mountains of the Republic of Buryatia as well as the lowland territory of Rostov Oblast in the course of joint researches. The developed technology of automated vegetation mapping is also applicable for observing dynamics of plant communities in mountains.

References

1.   E.A. Istomina, Geoinformation modelling and mapping the landscape complexes of Cisbaikalia (Barnaul, 2006)

2.   A. A. Glotov, Geoinformation modeling of the evolution of valley-river landscapes of Voronezh Oblast (Voronezh, 2013)

3.   E. A. Stytsenko, Development of the automatic interpretation of vegetation cover with the complex use of multi-seasonal satellite imagery (Moscow, 2013)

4.   S. Gadal, M. Zakharov, J. Kamičaitytė, Y. Danilov, *Alas Landscape Modeling by Remote Sensing Image Analysis and Geographic Ontology: Study Case of Central Yakutia (Russia)*, in Proceedings of the 6th International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM 2020), INSTICC, Apr 2020 (2020)

5.   M. Zakharov, S. Gadal, Y. Danilov, J. Kamičaitytė, *Mapping Siberian Arctic Mountain Permafrost Landscapes by Machine Learning Multi-Sensors Remote Sensing: Example of Adycha River Valley*, in Proceedings of the 7th International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM 2021), INSTICC, Apr 2021, Online streaming, Czech Republic (2021)

6.   The Atlas of Agriculture of the Yakutian ASSR (Moscow, 1989)

7.   E. G. Nikolin, The flora of the Verkhoyansk Range and its spatial organization (Yakutsk, 2012)