Mapping of the current state of forest vegetation in the north Baikal Region based on the merging of remote sensing data and ground observations

I N Vladimirov, D V Kobylkin and A A Sorokovoy

Sochava Institute of Geography SB RAS, Irkutsk, Russia

E-mail: garisson@irigs.irk.ru

Abstract. Geoinformation analysis and modeling study of the current state of vegetation (as exemplified by the southeastern macroslope of the Baikal Range, Eastern Siberia) were performed. The phytocenoses of the territory differ from each other in many features, namely in projective cover, ratio of ecological groups, confinement to landforms, epiterranean phytomass, etc. These details help to interpret the types of plant communities based on the values of the calculated vegetation index. We proposed a technique that assumes the joint use of the NDVI and NDWI indices, a digital elevation model (DEM) based on SRTM radar topographic survey data, forest inventory data and materials from field expeditions. The use of the DEM and the indices reflecting the physiological state of vegetation considering their availability of chlorophyll and water helps to arrange the communities under study into groups determined by moisture conditions and biomass supply. To determine the species composition of forest stand, we proposed a method based on the Boolean logic – a decision tree – a schematic representation in the form of a tree structure of a complex decision-making process used in a multi-step analysis. The results obtained show that the use of the relationship between the ratio of plant communities to the moisture factor and their spectral characteristics, taking into account the use of DEMs, enables us to create very informative maps, to improve the reliability of interpretation of satellite information, and to map territories not covered by ground-based surveys with the use of interpolation.

1. Introduction

Geoinformation analysis and modeling of vegetation include issues relating to the use of geographic information systems (GIS) and remote sensing data (RSD) with the aim of determining different vegetation state parameters, as well as mathematical methods and algorithms, and computational technologies connected with RSD.

An assessment of vegetation cover parameters is an integral part of geographical research [1]. Phytoindication methods in remote sensing are based on establishing relationships between spectral characteristics of an image and biological (phytomass resources, species composition) parameters of plant communities stands.

To determine bioparameters and to identify structures of the earth surface, multispectral high resolution satellite images are considered to be the most appropriate [2]. It is common knowledge [3] that spectral curves of natural objects are defined by three zones of the spectrum: green, red and near-infrared.
Multispectral and multitemporal satellite data have played a primary role in characterizing land cover change and deforestation rates [4], but are fast becoming a fundamental component of conservation planning and biodiversity assessment [5, 6]. Physiological state of vegetation is largely determined by the content of chlorophyll and moisture level of the green fractions of vegetation. Direct determination of absolute values of these parameters from remote sensing data is difficult to date and requires additional ground-based measurements. In this connection it is advisable to use relative indices, obtained from spectral values, which correlate with the level of chlorophyll and moisture supply to plants. The normalized difference vegetation index (NDVI) and the normalized difference water index (NDWI) can serve in this respect.

2. Materials and methods
To study and assess vegetation cover parameters we used diverse information, namely: satellite images, cartographic information, data of surface survey, library materials, and databases of GIS. They include: 1) moderate resolution satellite images (Landsat 5 TM, path 132, and row 22, acquired on 2010.08.23); 2) topographic maps of the 1:100 000 and 1:200 000 scales; 3) data of the NASA's Shuttle Radar Topography Mission (SRTM); 4) state inventory data on forest reserves, and forest assessment descriptions; 5) and field work materials.

To ensure comparability and opportunities for combined analysis of indicators of vegetation state, obtained from satellite data of high spatial resolution, to reduce the influence of random factors, as well as to ease the task of spatial modeling in GIS environment a technique has been proposed, which assumes a joint use of the NDVI and NDWI indices, digital elevation model (DEM) based on the data from the SRTM (http://www.jpl.nasa.gov/srtm/), forest management data (management plan and forest inventory data of the North Baikal Forestry (Severobaikalskii leskhoz)), and field work materials (geobotanical descriptions of the dominant types of vegetation in all altitudinal-zonal subdivisions) (figure 1).

![Figure 1](image-url)

**Figure 1.** The logical circuit of methodology for assessing and mapping forest vegetation with the use of RSD and GIS-technologies.

2.1. Study Area
The study area is located in the Baikal Range (54°30′-56°00′ N Lat., 108°20′-109°00′ E Long.). The Baikal Range (Baikal Mountains) rises steeply over the northwestern shore of Lake Baikal in southern Siberia, Russia (figure 2). Its highest peak is the Cherskii Mountain (2572 m). The mountains around Lake Baikal are densely wooded with Grey Alder (*Alnus incana*), Eurasian Aspen (*Populus tremula*),
Downy Birch (*Betula pubescens*), Siberian Larch (*Larix sibirica*), Siberian Fir (*Abies sibirica*), Siberian Spruce (*Picea obovata*), Scots Pine (*Pinus sylvestris*) and Siberian Pine (*Pinus sibirica*). According to the landscape-ecological regionalization of the Russian Federation, the territory under investigation is incorporated in subtaiga zone of the Middle-Siberian sector [7]. In accordance with the more detailed scheme of physical-geographical regionalization, this territory refers to the Baikal-Dzungur mountain-taiga physical-geographical region.

2.2. Methods

The vegetation index is an indicator that is calculated as a result of procedures with different spectral ranges of RSD, and that is related to the vegetation parameters in a given pixel of an image. The main assumption on the use of vegetation indices is that some mathematical operations with different bands of RSD can provide useful information about vegetation. This is confirmed by multiple empirical data [8, 9]. The second assumption is an idea that the open ground in an image will form a straight line (the so-called “soil line”) in the spectral space [10-14]. Almost all of the common vegetation indices use only the correlation of red and near-infrared bands, suggesting that in the near-infrared region there is a line of open soil. It is understood that this line means a zero amount of vegetation.

The NDVI is one of the most well-known indices derived from optical remote sensing imageries, has been extensively used to estimate plant biomass [15], patterns of productivity [16], growth status and spatial density distribution [17]. NDVI was first formulated by Rouse et al. [18] as the difference

![Figure 2. Location map of the study area in the Baikal region, in southern Siberia.](image)
between near-infrared and red visible reflectance values received by the sensors normalized over the sum of the two. Chlorophyll reflectance is about 20% in the red spectrum and 60% in the NIR spectrum and the contrast between the responses of both bands allows the quantification of the energy absorbed by chlorophyll, thereby providing indicative levels of different vegetation surfaces [19]. NDVI is moderately sensitive to changes in soil and atmospheric background.

Calculation of the NDVI is based on the two most stable (not dependent on other factors) sections of the spectral reflectance curve of vascular plants: the red spectral region (band 3 of Landsat 5 TM – 0.63-0.69 μm), which is a region of the maximum absorption of solar radiation by chlorophyll, and the near-infrared region (band 4 of Landsat 5 TM – 0.76-0.90 μm), which is a region of maximum reflection of cellular structures of a leaf, i.e. high photosynthetic activity leads to less reflection in the red spectral region and greater reflection in the infrared region. The NDVI is calculated as the ratio of the measured values of spectral brightness in the red (RED) and near-infrared regions (NIR) of spectrum by the following formula: \( \text{NDVI} = \frac{(\text{NIR} - \text{RED})}{(\text{NIR} + \text{RED})} \).

The NDWI was calculated according to the formula: \( \text{NDWI} = \frac{(\text{NIR} - \text{SWIR})}{(\text{NIR} + \text{SWIR})} \), where NIR is a value of the spectral brightness in the near-infrared spectral region, and SWIR is a value of the spectral brightness in the mid-infrared spectral region. The joint use of the values of brightness characteristics in the near-infrared (band 4 of Landsat 5 TM – 0.76-0.90 μm) and mid-infrared (band 5 of Landsat 5 TM – 1.55-1.75 μm) spectral regions makes it possible to reveal variations of vegetation cover, associated with moisture conditions [20].

The largest values of the NDWI index, connected with the moisture content, are generally observed for communities with hygrophilous vegetation (of swamp lands, meadows, etc.). The lowest values of the feature are registered in xerophytic communities.

For an objective identification of the features of the spatial association of plant communities, the DEM based on the SRTM data [21, 22] was used. The SRTM data is a matrix of heights with a mesh size of 3 arc seconds (about 90 m). It should be noted that the spatial resolution of SRTM in some way affects the subsequent calculations. This fully refers to the incline and slope aspect and other variables, the values of which respond to the specified dimensions of a cell of the model, and sets the limits on details of the simulated surface.

At the stage of preparation of materials for the study the import of data of SRTM to the MapInfo Vertical Mapper format was carried out for subsequent conversion of the bitmap representation of objects into the vector ones. The required projection and units of measurements of coordinates, distances and areas were set. To process and analyze the values of vegetation indices obtained digital elevation models at regular steps of 30 m were constructed, which corresponds to the spatial resolution of the survey data of Landsat 5 TM. The GRID of absolute altitude obtained was used to calculate the steepness and slope aspect. All the data obtained in the form of regular networks (altitude, steepness and exposure of slopes, and the values of the NDVI and NDWI indices) were converted into the vector form and represent an array of regularly distributed points. Thus, the data on the factors of analysis being considered were prepared. These data are summarized in tables of the MapInfo and Excel formats.

3. Results and discussion

The territory of the southeast macroslope of the Baikal Range under investigation is characterized by high capacity of habitats, and, consequently, the diversity of vegetation communities. Phytocoenoses differ from each other in a large number of features. Namely, cover density, correlation of ecological groups, relation to landforms, top phytomass, etc. All this allows us to interpret types of plant communities according to the values of calculated vegetation indices [23].

At the first stage of the thematic image analysis, the NDVI vegetation index is calculated and the results are classified according to vegetation types depending on the range of NDVI values. A set of classes, distinguished in the course of the analysis of satellite data, should provide a separation of forested areas from areas without forest cover, as well as a subdivision of forested areas into coniferous, deciduous and mixed stands. As a source of auxiliary data for image classification we used forest
management materials, reflecting the spatial distribution of the region's forests and their species composition, as well as the data of field research in key areas.

Digital multispectral satellite information reflects various factors of the landscape environment and can be interpreted as the coordinate space of complex factors of such an environment. This allows studying the patterns of pixel placement within the boundaries of the plot in this factorial ordination space. Each plot has its own frequency distribution of pixels according the index values (histogram). In the ordination space the plot is represented by a set of points. For its specific definition, a modal value is allocated (the optimum point is the weighted arithmetic mean) (table 1). The remaining index values are considered as a set of acceptable (relational) deviations, outside of which the unusual conditions of this location fall.

| Landscape                                      | Weighted arithmetic mean (average) of NDVI | Standard deviation |
|------------------------------------------------|-------------------------------------------|--------------------|
| Alpine tundra                                  | -0.28                                     | 0.00917            |
| Alpine meadows                                 | 0.21                                      | 0.00976            |
| Mountain steppes                              | -0.20                                     | 0.00636            |
| Sub-golets overgrowth of Siberian dwarf pine   | -0.12                                     | 0.01270            |
| Sub-golets spare forests and overgrowth of bushes | -0.04                                     | 0.01706            |
| Spruce forests                                 | 0.04                                      | 0.01819            |
| Fir forests                                    | 0.07                                      | 0.01261            |
| Siberian pine forests                          | 0.09                                      | 0.01527            |
| Larch forests                                  | 0.15                                      | 0.00703            |
| Pine forests                                   | 0.22                                      | 0.00462            |
| Birch and aspen forests                        | 0.29                                      | 0.00362            |
| Meadows and bogs                               | 0.23                                      | 0.02736            |

The proximity of patches in ordination space is not always due to their typological similarity in the composition of the stand or type of location. For example, it is often difficult to distinguish between Siberian stone pine and fir forests, pine and larch forests, etc. In disputable regions a patch is assigned a type, which manifests itself to the maximum (according to the number of pixels) within the boundaries of the contour in an image.

The NDVI values varied within individual phytocenoses ambiguously, which was determined, on the one hand, by the predominance in them of plants of various life forms and species, and, on the other hand, by the accumulation of biomass during the growing period. The largest values of the NDVI in the key areas in the period of maximum vegetation development are characteristic for patches with tree layer continuum, represented by light-coniferous forests (0.20-0.50); slightly lower index values correspond to small-leaved forests (0.10-0.20), and the values 0 to 0.10 correspond to burned areas, cutover areas, and open stands. The smallest amount of ground phytomass and the lowest values of the NDVI (from -0.20 to 0) are characteristic for dry stony steppes and golets (bald peak, barrens within the alpine tundra belt). Lichen vegetation on steep rocky-talus slopes of the northern and northwestern exposures has NDVI values of less than -0.20 (figure 3). The obtained NDVI values differ from the generally accepted ones, which is primarily associated with the mountainous terrain of the study area, relatively high absolute altitude, pronounced altitudinal zonation, and sparse vegetation cover.
Figure 3. Classified image obtained by the calculation results of the NDVI. 1 – vegetation with tree layer continuum, represented by light-coniferous forests; 2 – small-leaved forests; 3 – burned areas, cutover areas, and open stands; 4 – dry rocky steppes and golets; 5 – lichen vegetation on steep talus slopes of the northern and northwestern exposures.

At the second stage of the research the index of the moisture content NDWI is calculated. The joint use of the NDVI and NDWI indices, reflecting the physiological state of forests in terms of the chlorophyll and water supply to them, makes it possible to distribute the investigated communities into groups, which are determined by the moisture conditions and biomass reserves. The first variety corresponds to drier habitats, such as golets and steppe belts. The second variety of communities is confined to wet habitats, occupying topographic lows (coastal-aquatic vegetation, bog massifs, floodplain vegetation, etc.). The nival-glacial belt clearly stands out.

To differentiate the territory and to select similar locations of vegetation, maps of steepness and slope aspect were constructed based on the digital terrain model. Six gradations of steepness of slopes were distinguished: 1) leveled surfaces (0-2°); 2) gentle slopes (2-7°); 3) downward slopes (7-15°); 4) slopes of medium steepness (15-20°) 5) steep slopes (20-40°), and 6) very steep slopes (over 40°) (figure 4). Also, eight types of slope aspect were identified (figure 5).

To determine the species composition of forests, a technique (figure 6), based on the algebra of logic, is suggested; it is a decision tree, i.e. a schematic representation of a complex decision-making process, applied for multi-way analysis process, in the form of a tree structure. Decision tree learning is among the most popular machine learning techniques used for ecological modelling. Decision trees can be used to predict the value of one or several target (dependent) variables. They are hierarchical structures, where each internal node contains a test on an attribute (SRTM, NDVI, NDWI), each branch corresponding to
an outcome of the test, and each leaf node giving a prediction for the value of the class variable (forest types). The common way to induce decision trees is the so-called Top-Down Induction of Decision Tress.

![Figure 4. Differentiation of the territory under investigation according to the steepness of slopes.](image1)

![Figure 5. Differentiation of the territory under investigation according to the slope aspect.](image2)

![Figure 6. Decision tree for forest type determination.](image3)

The results, obtained in the course of the research, are confirmed by the data of the field work, conducted in the key areas. The first site is located on the shores of Lake Baikal in the area of the Kotelnikovskii Cape (figure 7a) and the estuarine parts of the Kurkula and Goryachaya river valleys. The second site is located in the middle reaches of the Kurkula river (figure 7b). The correlation between data from field mapping and NDVI data is high (0.45<r<0.73). The comparison of the results (figure 8) demonstrates good representativeness for the entire analysis area and can be applied in Northeast Asia.
Figure 7. Landscape structure of model sites, based on field research. a – model site No. 1, Kotelnikovskii, b – model site No. 2, Cape Kurkula and Goryachaya river valleys. 1 – snow patches; golets: 2 – of rocky watersheds and hanging circuses, 3 – of remnants of weathering and taluses, 4 – steep rocky-talus lichen, 5 – siberian dwarf pine, 6 – mountain-valley meadow-tundra; mountain-taiga: 7 – mountain-valley dark-coniferous, 8 – mountain-valley light-coniferous, 9 – slope light-coniferous of open stands and sparse woods, 10 – small-leaved on burned and cutover areas, 11 – slope light-coniferous, 12 – floodplain of large rivers; Piedmont: 13 – waterlogged yernik (birch shrubs) of valley bottoms, 14 – valley willow, 15 – of debris cones pine herbaceous, 16 – pseudo-sub-golets; mountain-steppe: 17 – slope steppificated.

Figure 8. Classified image of the NDVI, combined with the contours of the landscape patches, obtained in the course of the field work.

Figure 9. Current state of forest vegetation in the North Baikal Region. 1 – alpine tundra, 2 – alpine meadows, 3 – sub-golets spare forests and overgrowth of bushes, 4 – sub-golets overgrowth of Siberian dwarf pine, 5 – spruce forests, 6 – fir forests, 7 – siberian pine forests, 8 – larch forests, 9 – pine forests, 10 – birch and aspen forests, 11 – mountain steppes, 12 – meadows and bogs.
With the relationship between the correlation of landscapes and plant communities with NDVI and NDWI, taking into account the results from using the DEM and detailed data of course observations, it is possible to create the map of current state of forest vegetation (figure 9).

4. Conclusion
Rational nature management and conservancy issues have been debated in recent years with increasing frequency in Siberia. In this case, rationalization is not so much linked to the resource aspects of nature management as to its ecological characteristics, and to assessment of its implications for the environmental quality. All this applies primarily for Siberia’s vegetation providing the source of varied bioresources as well as an accurate indicator of the state of environment.

The results of the research show that the use of dependence between the relation of vegetation communities to the humidity factor and their spectral characteristics with due regard for the use of DEM makes it possible to map territories not covered by surface survey via interpolation. Such investigations not only provide an opportunity to create thematic maps that are rather rich in content, but also enable to improve reliability of the interpretation of satellite data [24].

Acknowledgements
The reported study was funded by RFBR according to the research project № 17-29-05089 and within the Russian State Assignment (Project Reg. No. AAAA-A17-117041910167-0).

References
[1] Vladimirov I N and Chudnenko A K 2009 Multilevel modeling of the forest resource dynamics Mathematical Modelling of Natural Phenomena 5(4) 72-88
[2] Vladimirov I N 2014 Predictive mapping of the restoration-age dynamics of taiga forests on the basis of remote sensing data and geographical knowledge Applied Ecology and Environmental Research 12(4) 807-24
[3] Schowengerdt R A 2007 Remote Sensing: Models and Methods for Image Processing (London: Academic Press) p 560
[4] Lu D, Mausel P, Brondizio E and Moran E 2004 Change detection techniques International Journal of Remote Sensing 25(12) 2365-407
[5] Sesnie S E, Gessler P E, Finegan B and Thessler S 2008 Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments Remote Sensing of Environment 112(5) 2145-59
[6] Stickler C M, Southworth J 2008 Application of multi-scale spatial and spectral analysis for predicting primate occurrence and habitat associations in Kibale National Park, Uganda Remote Sensing of Environment 112(5) 2170-86
[7] Isachenko A G 1996 Landscape regionalization of Russia as the basis for regional ecologo-geographical analysis Izvestia RGO 128(5) 12-24 (In Russian)
[8] Tucker C J 1979 Red and photographic infrared linear combinations for monitoring vegetation Remote Sensing of Environment 8(2) 127-50
[9] Chen W, Zhao X 2007 Estimation of forest parameters based on TM imagery and statistical analysis Journal of Forestry Research 18 241-4
[10] Richardson A J and Wiegand C L 1977 Distinguishing vegetation from soil background information Photogrammetric Engineering and Remote Sensing 43(12) 1541-52
[11] Baret F, Jacquemoud S and Hanocq J F 1993 About the soil line concept in remote sensing Advances in Space Research 13(5) 281-4
[12] Yoshioka H, Miura T, Demattê J A M, Batchily K and Huete A R 2010 Soil line influences on two-band vegetation indices and vegetation isolines: a numerical study Remote Sensing 2(2) 545-61
[13] Xu D, Guo X A 2013 Study of soil line simulation from landsat images in mixed grassland Remote Sensing 5(9) 4533-50
[14] Ahmadian N, Demattê J A M, Xu D, Borg E and Zölitz R 2016 a new concept of soil line retrieval from landsat 8 images for estimating plant biophysical parameters Remote Sensing 8(9) 738
[15] Prince S D 1991 Satellite remote sensing of primary production: comparison of results for Sahelian grassland 1981-1988 International Journal of Remote Sensing 12(6) 1301-11
[16] Gowd S N and Dye D G 1987 Evaluating North American net primary productivity with satellite observations Advanced Space Resources 7(11) 165-74
[17] Purevdorj T S, Tateishi R, Ishiyama T and Honda Y 1998 Relationships between percent vegetation cover and vegetation indices International Journal of Remote Sensing 19(18) 3519-35
[18] Rouse J W, Haas R H, Schell J A and Deering D W 1973 Monitoring vegetation systems in the great plains with ERTS Third ERTS Symp. NASA SP 351(1) 309-17
[19] Tucker C J and Sellers P J 1986 Satellite remote-sensing of primary production International Journal of Remote Sensing 7(11) 1395-416
[20] Ceccato P, Flasse S, Tarantola S, Jacquemoud S and Gregoire J-M 2001 Detecting vegetation leaf water content using reflectance in the optical domain Remote Sensing of Environment 77(1) 22-33
[21] Rabus B, Eineder M, Roth A and Bamler R 2003 The shuttle radar topography mission – a new class of digital elevation models acquired by spaceborne radar ISPRS Journal of Photogrammetry and Remote Sensing 57(4) 241-62.
[22] Luedeling E and Buerkert A 2008 Typology of oases in northern Oman based on Landsat and SRTM imagery and geological survey data Remote Sensing of Environment 112(3) 1181-95
[23] Vladimirov I N, Sofronov A P, Sorokovoi A A, Kobylin D V and Frolov A A 2014 Structure of vegetation cover in the western part of the Upper Angara depression Geography and Natural Resources 2(35) 143-51
[24] Vladimirov I N 2018 The ecological potential of Baikal region’s geosystems IOP Conf. Series: Earth and Environmental Science 190 012017