Determination of the state of forests based on a regular grid of ground-based sample plots and Sentinel-2B satellite imagery using the k-NN ("nearest neighbor") method

A Alekseev*, D Chernikhovskii, L Vetrov, M Gurjanov and I Nikiforchin

Department of forest inventory, management and GIS, Saint-Petersburg State Forest Technical University, 5 Institutsky Lane, St. Petersburg 194021, Russian Federation

*Corresponding email: a_s_alekseev@mail.ru

Abstract. In the paper on the base of regular grid of sample plots and remote sensing data from Sentinel – 2B the tree stands health status was estimated in continuous way for the study area. This study used a Sentinel-2B image (date of shooting July 31, 2018). The following spectral channels were selected: 8 – Near Infrared, 4 – Red, 3 – Green, 2 – Blue. After the initial processing, a single multi-zone image with four channels was formed, which was later used for classification. The following settings of the kNN classifier plugin were used to classify the territory based on the state of the stands: the number of nearest neighbours k = 5, the number of spectral channels – four, and the classification mode – for continuous values. The quality of determining the state of tree stands based on the k-NN classification of the territory was determined by comparing its results with the data of the control sample plots by calculating systematic and random errors. As a result of the calculations the following values of errors were obtained: the systematic error was equal to zero meanwhile random appears to be equal 1 damage class, both may be considered as satisfactory and used classification method as practically applicable.

1. Introduction

Determination of the state of forests currently has a value comparable in relevance to the assessment of the taxation characteristics of plantings for the needs of determining the reserves of forest resources as sources of reproducible raw materials for circulating bioeconomy. The state of forests changes under the influence of a large number of factors of natural and anthropogenic origin, such as forest fires, the appearance of pests and diseases of the forest, adverse weather events (windfalls, snow breaks, abnormal frosts, desiccation, flooding), pollution of the atmosphere, soil and water, and others. Damaged forests can occupy large areas, for example, forest fires can spread over areas of millions of hectares. In this regard impossible to determine the damage to forests in large areas without the use of remote sensing methods and the use of special procedures for statistical processing of remote data in combination with ground data. Materials of remote sensing of the Earth when solving problems of determining the state of forests necessarily need to be calibrated and verified on the basis of data from ground-based sample areas, without which they can give a distorted picture of the state of forests.

Among the methods of automated classification of remote sensing materials a special position is occupied by the "nearest neighbor" method or the k-NN method (k-nearest neighbors). An important feature of the method is that it allows combining the use of remote sensing materials and data from ground-based test areas in one procedure. This method was first proposed for the study of forests by
Finnish researchers in the 1990s [1-7]. As the practical interest in the k-NN method grew in the early 2000s a number of studies were conducted aimed at a comprehensive study of its capabilities and limitations [8-14]. The main objective of this study was to develop a methodology for assessing the state of forest tree stands in the border zone of Russia and Finland using a regular network of sample areas, remote sensing materials and geoinformation technologies.

2. Methods and Materials
During the summer field season of 2019 the state of forest tree stands in the border zone of Russia and Finland was assessed. The objects of the study were the sample plots regularly laid out under the ICP-Forests methodology on the territory of the Vyborg district of the Leningrad Region (see figure 1 and table 1). The method of conducting field work is described in detail in [15-17].

![Figure 1. Regular grid of sample plots colored according dominating tree species (orange – Scots pine, violet – Norway spruce, blue – Birch, green – aspen).](image)

As can be seen from table 1 main part of the sample plots are dominated by Scots pine – 59.6% and Norway spruce – 28.3%. Broadleaves species are presented totally on 12.1% of sample plots. The values of the average damage classes, according to the data of the surveyed sample plots were in the range from 0.28 to 0.69 damage class, i.e. there were no severely damaged, dying and dead tree stands (see table 1).

| Tree species       | Number of sample plots | %     | Mean damage class |
|--------------------|------------------------|-------|-------------------|
| Scots pine         | 59                     | 59.6  | 0.69              |
| Norway spruce      | 28                     | 28.3  | 0.40              |
| Birch              | 11                     | 11.1  | 0.28              |
| Aspen              | 1                      | 1.0   | 0.30              |
| Total              | 99                     | 100.0 | 0.54              |

To increase the representation of the tree stands with bad health status forest allotments with severe damaged, dying and dead tree stands were added to the ground based data. The materials of forest management were used from the standard geoinformation databases of the North-Western forest management district. With the help of the WinGIS and PLP-2015 programs needed tree stands were selected for which the degree of damage was indicated during the forest survey of 2019 as severe damaged, drying as well as dead tree stands. A set of allotment polygons with severely damaged, drying and dead plantings was exported to GIS. Using high spatial resolution images of open access (Bing, Google and Yandex) the suitability of the selected plots for further analysis was evaluated.
degree of damage to the plantings, the configuration of the allotment and the proportion of the damaged part of the allotment were visually assessed. A total of 153 ground sample areas were used in the processing, which were divided into two equal parts. Training trial areas were used to classify the territory according to the degree of vegetation damage, and control areas were used to assess the quality of the classification. The classification of the territory by the state of vegetation using the k-NN method was performed on the basis of only training part of sample plots and its quality was checked by comparing the classification results with the data of the control sample areas.

The quality of determining the state of tree stands based on the k-NN classification of the territory was determined by comparing its results with the data of the control sample plots by calculating systematic and random errors. Systematic (bias) and random (standard deviation) errors were determined by the following formulas:

\[
\Delta x = \frac{\sum_{i=1}^{n}(x_r - x_g)}{n},
\]

\[
SD = \frac{\left(\sum_{i=1}^{n}(\Delta x - \Delta x_i)^2\right)^{1/2}}{n-1},
\]

where, \(\Delta x\) is a systematic error in determining the damage class of tree stands, points; \(x_r\) and \(x_g\) are the damage classes of tree stands determined by the k-NN method using remote sensing materials and ground-based methods, respectively; \(n\) is the number of control sample plots.

\(SD\) is a random error in determining the state class of stands, points; \(\Delta x_i = x_r - x_g\) is an error in determining the tree stand damage class by ground and remote methods for the control test sample plot \(i\), points.

In this study Sentinel-2B image was used (date of shooting July 31, 2018) in which 4 following spectral channels were selected: 8 – Near Infrared (NIR), 4 – red (Red), 3 – green (Green), 2 – blue (Blue). The images in these channels had a spatial resolution of 10 m. After the initial processing, a single multi-spectral image with four channels was formed which was later used for classification by k-NN method. Sentinel - 2B image combined with a map with location of the ground sample plot network is presented on figure 2.

![Figure 2. Sentinel - 2B image combined with a map with location of the ground sample plot network.](image-url)
QGIS 3.10.11 - A Coruna GIS with the Semiautomatic classification plugin was used for preprocessing of survey materials (correction and conversion of brightness, atmospheric correction, and multi-spectral image formation) [18]. The classification of satellite images by the k-NN method was carried out using the QGIS Desktop 2.14.12 GIS with the k-NN classifier plugin (developed by Ventspils University College) [19]. The following settings of the k-NN classifier plugin were used to classify the territory based on the state of the stands: the number of nearest neighbor \( k = 5 \), the number of spectral channels – four, and the classification mode – for continuous values.

3. Results and Discussion
As a result of the calculations the following values of errors were obtained: the systematic error was equal to zero \( (\Delta x = 0 \text{ points}) \) and the random error was equal to 1 point \( (SD = 1 \text{ point}) \). This error values means that bias is absent in the classification done and as a result of random error tree stands health status may be mistakenly determine on 1 point better or worse comparatively with true value. In other words healthy tree stands may be classified as slightly damaged and wise versa, slightly damaged as damaged moderately and wise versa, moderately damaged as damaged severely and wise versa, severely damaged as dead tree stand and wise versa. Such random errors may be considered as are not very significant. Thus, it can be concluded that the application of the k-NN method to classify the studied territory by the health state of forest vegetation gives a positive result because the errors both systematic and random appears to be acceptable.

On figure 3 the thematic map presents the classification of study area located on near border territory of Russia and Finland according to damage classes of tree stands.

Figure 3. Thematic map of the study area colored according damage class of forest tree stands (Light shades correspond to healthy stands, dark shades correspond to damaged ones. The darker the shade, the worse the condition of the tree stands).

The proposed method of tree stands classification according its health status can be used in solving many important tasks of sustainable forest management, such as

- Planning the use, reproduction, protection and conservation of the forests taking into account their condition.
- Identification of forest areas that is completely healthy or damaged in various degrees of severity.
Planning of forest management operations for damaged forests in order to preserve their stability and sustainability, for example, clear cutting and subsequent renewal of dead tree stands, sanitary logging in damaged stands, etc.

Identification of the causes of forest damage and development of measures for their elimination.

The proposed method for classifying forests according to the degree of damage is essentially based on the joint use of data from ground-based sample areas and remote sensing materials, which significantly increases the reliability of the results obtained.

4. Conclusion

The k-NN method of forest classification based on the joint use of remote sensing data and ground-based sample areas has been widely used in national forest inventories, especially in Finland. With its help, objective data for large areas were obtained during the forest inventory and the method proved its practical effectiveness. When conducting a forest inventory, the main taxation indicators of stands are determined, such as tree stands species composition, their age, average heights, diameters and wood stock. In this study, an attempt was made to apply the k-NN method to assess the condition and extent of damage to tree stands, which are also of great importance for solving some specific problems of sustainable forest management. The results of the application of this method for assessing the state of forests in large areas using remote sensing materials were positive. A very important result is that the method does not give systematic errors, and a random error creates uncertainty only when determining neighboring classes of damage to tree stands, which is not critical for its application.

As a result, the method can be recommended for solving many problems of forestry related to monitoring the state of forests. In particular, it can be used to organize continuous monitoring of the health status of forests on the basis of obtaining up-to-date materials of remote sensing of the Earth and constantly informing interested groups of the population about the results obtained using modern information technology capabilities.

Acknowledgements

This study was carried out with the financial support of the project No. 21-16-00065 “The role of insects and pathogens in the weakening and death of coniferous stands in the North-West of the Russian Federation: quantification and monitoring” funded by the Russian Science Foundation.

5. References

[1] Haapanen R and Ek A 2001 Software and Instructions for kNN Applications in Forest Resources. Description and Estimation Staff paper series 152 Available at: https://conservancy.umn.edu/handle/11299/37198

[2] Tomppo E, Czaplewski R and Mäkisara K 2002 The role of remote sensing in global forest assessment Forest Resources Assessment Working Paper 61

[3] Kangas A and Maltamo M 2006 Forest Inventory: Methodology and Applications Edited by Kangas A and Maltamo M (Dordrecht: Springer) p 368

[4] Gjertsen A 2007 Accuracy of forest mapping based on Landsat TM data and a kNN-based method Remote Sensing of Environment 110 pp 420–430

[5] Meng Q, Cieszwski C J, Madden M and Borders B 2007 K nearest neighbor method for forest inventory using remote sensing data GIScience&Remote Sensing 44 pp 149-165

[6] Ösakabaşi F 2008 Classification of forest areas by k nearest neighbor method (Antalya: Middle East Technical University) 101

[7] Tomppo E, Haakana M, Katila M and Peräsaari J 2008 Multi-Source National Forest Inventory: Methods and Applications, vol 18 ed K Gadow T Pukkala M Tome (Springer Science & Business Media) p 388

[8] Beaudoin A, Bernier P Y, Guindon L, Villemare P, Guo X J, Stinson G, Bergeron T, Magnussen S and Hall R J 2014 Mapping attributes of Canada’s forests at moderate
resolution through kNN and MODIS imagery Can. J. For. Res. 44 pp 521-532
[9] Franco-Lopez H, Ek A R, Bauer M E and Franco-Lopez H 2001 Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method Remote Sensing of Environment 77 pp 251–274
[10] Haapanen R, Ek A R, Bauer M E and Finley A O 2004 Delineation of forest/nonforest land use classes using nearest neighbor methods Remote Sensing of Environment 89 pp 265-271
[11] Gjertsen A K 2005 Accuracy of forest mapping based on Landsat TM data and a kNN-based method Proc. of ForestSat 2005 Borås ed H Olsson Jönköping: Skogsstyrelsen pp 7-11
[12] Koukal T, Suppan F and Schneider W 2005 The impact of relative radiometric calibration on the accuracy of kNN-predictions of forest attributes Proc. of ForestSat 2005 Borås ed H Olsson Jönköping: Skogsstyrelsen pp 17-21
[13] McInerney D, Pekkarinen A and Haakana M 2005 Combining Landsat ETM+ with field data for Ireland’s National Forest Inventory – a pilot study for co. Clare Proc. of ForestSat 2005 Borås ed H Olsson Jönköping: Skogsstyrelsen pp 12–16
[14] Chernikhovskii D M and Alekseev A S 2019 The Method for Determining Forest Characteristics Based on Earth Remote Sensing Materials, Forest Management Data and the k-NN Algorithm (Case Study of Lodofinopolskoe Forest District of Leningrad Region) Lesnoy Zhurnal [Forestry Journal] 4 pp 45–65
[15] Alekseev A 2003 Forest Ecosystems Monitoring Text Book 2nd ed. [in Russian – Monitoring lesnykh ekosistem 2-ye izdaniye (Uchebnoye posobiye)] (Saint-Petersburg: Saint-Petersburg State Forest Technical Academy Publishing) p 116
[16] ICP Forests Manual available at: http://icp-forests.net/page/icp-forests-manual
[17] Alekseev A, Vetrov L, Gurjanov M, Nikiforchin I, Chernikhovsky D and Chernov I 2020 Analysis of the tree stands health status in the near border area of Russia and Finland based on the regular grid of sample plots and GIS technologies IOP Conference Series: Earth and Environmental Science 507 012001. DOI:10.1088/1755-1315/507/1/012001 pp 1-8
[18] Congedo L 2020 Semi-Automatic Classification Plugin Documentation Release 7.0.0.1. p 197 Available at: http://dx.doi.org/10.13140/RG.2.2.25480.65286/1
[19] Supervised kNN classifier (kNN) 2017 p 19 Available at: https://github.com/m6ev/kNN/find /master