

ACCURACY ASSESSMENT OF AUTOMATIC IMAGE PROCESSING FOR LAND COVER CLASSIFICATION OF ST. PETERSBURG PROTECTED AREA*

Anna Bogoliubova¹, Przemysław Tymków²

¹ National Mineral Resources University

² Wrocław University of Environmental and Life Science

Abstract. This study analyzes the evaluation of land cover supervised classification quality. Authors put forward the hypothesis that the overall accuracy of image classification depends on its division into parts of the same area. The dependence is described by the logarithmic curve – $T = 4.3004 \cdot \ln(x) + 72.697$, because the determination coefficient is maximum ($R^2 = 0.9678$). The research area was the Yuntolovo reserve, the protected area near St. Petersburg (Russia). In order to increase the overall accuracy of the land cover automatic classification based on aerial images, a new methodology of data preprocessing was introduced. The proposed method of estimating the overall classification accuracy of land cover protected areas increases on average by 10% by dividing the source aerial image into no more than 10 equal parts. With further partitioning of the image into parts of the same area, the overall accuracy is slightly increased. Pixel-based image analysis of supervised classification and error matrix were evaluated using ILWIS 3.31 software and in our own software in .NET environment.

Key words: overall accuracy, automatic image processing, protected area, land cover/use, supervised classification

INTRODUCTION

Looking at a land cover map it is important to note that it is not a perfect representation of reality. There are always errors in maps and before we can evaluate the utility of a particular map we need to realize how accurate it really is and how accurate it should

* This work was supported by the National Science Centre project No. UMO-2011/01/D/ST10/07671

© Copyright by Wrocław University of Environmental and Life Sciences

Address correspondence to – Adres do korespondencji: Przemysław Tymków, Wydział Inżynierii Kształtowania Środowiska i Geodezji, Instytut Geodezji i Geoinformatyki, ul. Grunwaldzka 53, 50-357 Wrocław, e-mail: przemyslaw.tymkow@up.wroc.pl

be to sufficiently meet the requirements of the intended application. The most reliable method of checking the accuracy of a map in a quantitative way depends on designing and implementing an accuracy assessment. Without such an objective measure of the map's quality we might be limited to use qualitative statements from people who have used the map. These comments may provide a rough indication as to the quality of a map but they tend to be very subjective and can be misleading.

The results of an accuracy assessment typically provide us with an overall accuracy of the map and the accuracy for each class in the map. Accuracy assessment should be an important part of any classification. Proper accuracy assessment of a product created using remotely sensed data can be time consuming and costly. This is the primary reason why this important step is sometimes not taken in a mapping project or is substantially modified in order to save time or money. In these cases one can try to take an educated guess of what the accuracy is, based on the methods, experience, and data used but in the end it is only a guess. However, without any quantitative accuracy assessment we do not know how precise our classification is.

This study is based on previous research of the automatically performed supervised classification for the territory of the Yuntolovsky reserve, which is situated in St. Petersburg (Russia). Supervised classification is the process of using training samples, samples of known identity to classify pixels of unknown identity.

Considering the above facts, the objectives of this paper are: to compare the classification results with reference data and to provide a quantitative assessment using an error matrix (calculation of the omission error, commission error, overall accuracy and K-hat).

The main objectives of this study are: 1) to assess how well a classification worked; 2) to understand how to interpret the usefulness of image classification; 3) to prove the hypothesis the authors' put forward, namely that the overall accuracy of image classification depends on its divisions into parts of equal area for interpretation.

STUDY AREA

The protected areas have been founded in many countries around the world. They are protected by law. Scientists are studying these areas and people visit them. But inappropriate economic activity can have adverse effects on the natural landscape and lead to the destruction of links between its constituent components. To prevent this from happening, in these areas conservation measures should be introduced. They should aim at protecting the natural environment through better monitoring and control systems. Environmental protection measures rely on land use planning to study the dynamics of land use/cover and its changes and biodiversity of natural systems.

The object of this research is the natural complex of the Yuntolovsky reserve situated in St. Petersburg (Russia). The reserve was founded in 1990 and it is one of the protected natural areas in St. Petersburg. The area of the reserve is 1025.4 ha [Bogoliubova 2012]. The reserve lies in the northern part of St. Petersburg in the Primorsky district. Its territory is washed by the Lakhta Bay – the gulf of the Neva River. There is a very vast forest in the northern part of the Yuntolovo reserve. The Lakhta Bay and the Yuntolovo forest are considered to be the territory of the reserve. There are 337 species of vascular plants and

69 species of mosses in this area [Volkova 2005]. Not only some species of plants but all kinds of flora and fauna on this territory are protected. The area is famous for its varied flora and fauna.

BACKGROUND AND PREVIOUS WORK

Recognition of land cover types based on spectral reflectance characteristics of satellite data is one of the fundamental tasks of remote sensing. The resulting spatial information data about the types of land cover based on remote sensing facilitates creating thematic maps and digital databases, which are used, in particular, for optimal control of territories, land use/cover management, environmental protection and basic research in Earth sciences. Remote sensing is cost-effective and it is increasingly used to characterize urban land use/cover [Jensen and Cowen 1999].

Previous investigations were carried out in the Yuntolovsky nature reserve in the period from 2011 to 2012. The main objective of this study is to compare the algorithms of supervised and unsupervised classification techniques for the classification of the Yuntolovsky reserve in St. Petersburg (Russia) according to land cover/use based on aerial imagery. The following information was available for this research: aerial images of the Yuntolovsky reserve (Primorsky District of St. Petersburg) (scale 1: 5000, taken in 2005 and 2012); topographical maps (scale 1: 2000, made in 2002 and 2004; scale 1: 5000, made in 2005 and 2011); a thematic geobotanical map and a description (made in 2005).

Before starting work with non-metric aerial images, these images were georeferenced. For geometric correction, an aerial image was registered in the Local Coordinate System 1964 (rectangular coordinate system) via projective transformation with the use of the Nearest Neighbor resampling method using ArcGIS 10.0. The image was adjusted to a topographic map with the use of reference points (in this case the lake outline, meadows outline) acquired by digitization. Fourteen adjustment points were used. The accuracy of adjustment was approximately 0.3–1.0 meter and was accepted as adequate because the topographical map was created in 2011 and the image was taken in 2012. The spatial resolution of pixel was 0.5 m. On account of computational complexity of the recognition process, the mosaic was divided into 100×100 m squares [Tymkow 2009] in line with the grid on the topographical map, which constituted the data for classification [Bogoliubova and Tymkow 2014].

The next step of research was to create a nomenclature for land cover. A modified scheme of Anderson land use/cover classification system was adopted to examine multi-temporal land use/cover changes in the study area since this classification system is suitable and can efficiently be mapped from aerial imagery [Anderson et al. 1976]. Five different land use/cover types have been identified in this study as: forest and vegetation, water bodies, wetland/lowlands, artificial surface, and cultivated land. This land cover classification scheme indicates major categories of land cover [Bogoliubova and Tymkow 2014]. The classification of images using the maximum likelihood algorithm was performed in our own software in NET environment. Based on these experiments we assessed the land cover recognition accuracy which allowed for comparison of algorithms efficiency.

The overall accuracy for land cover map (2012) was 71.5% with the corresponding kappa statistics of 63.7%. A standard total accuracy for correct classification of land cover types has been set between 85 percent and 90 percent [Anderson et al. 1976]. Examination of the total accuracy of the derived maps showed that they did not meet this minimum requirement.

MATERIALS AND METHODS

Importance of accuracy assessment

Image classification cannot be considered complete until we obtain an estimate of its accuracy. The interpreter (decoder) has to determine exactly how the object classes in the image correspond to these classes on the earth (terrain) surface. In the study of image processing the term “accuracy” means a measure of consistency with reliable information in a spatial point with data on the classified image [Jensen 1996].

The classification accuracy is usually assessed by comparing the classification with some reference data that is believed to accurately reflect the true land cover. Sources of reference data include, among other things, ground truth, higher resolution satellite images, and maps derived from aerial photo interpretation. We should note that virtually all reference data (even ground truth data) are inaccurate to some degree [Lillesand 2004].

The accuracy assessment reflects the real difference between our classification and the reference data. Consequently, if reference data is highly inaccurate, assessment might indicate that classification results are poor, while in fact it is a correct classification. It is better to get less, but more accurate reference data. Also to be aware of temporal changes: if a satellite image or an aerial image was not taken at the same time when reference data were collected, apparent errors might occur due to the fact that landscape has changed. Ideally, selection of reference sites should be based on some random sampling design. Ground ruts made by cars do not provide a random design and the estimated accuracy measures from these types of sampling designs are usually biased [Lillesand 2004].

When using the supervised classification method, the simplest way to assess its accuracy is to compare the classified data with a training set. Accuracy assessment should not be based on the training pixels. The problem with using training pixels is that they are usually not randomly selected and that the classification is not independent of the training pixels. Using training pixels usually results in an overly optimistic accuracy assessment [Lillesand 2004]. If we assume that the training samples were formed correctly and accurately, the resulting accuracy is 100%. For this reason, a more reliable method of assessing the classification accuracy of remote sensing data is:

- to find data related to a known object;
- to split them into two parts;
- and then use one of them as the training set, and the second one to assess the classification accuracy [Bogoliubova 2012].

There are two primary components of errors in thematic maps, such as land cover maps; a position error and a thematic error. When we look at the accuracy of a land cover map we typically do not differentiate between position and thematic errors. The accuracy assessment sampling design usually has a built-in mechanism that allows for reasonable

position errors based on the input data and on the scale of the final map. In the end, however, the accuracy figures do not indicate if the error is due to a position error or a thematic error [Bogoliubova 2012].

The results of accuracy assessment are usually summarized in a confusion matrix, which describes not only the classification error for each class, but also errors related to incorrect classification. This contingency table is often referred to as a confusion matrix, misclassification matrix [Bogoliubova 2012], or error matrix [Rosenfield and Fitzpatrick-Lins 1986, Congalton 1991, Congalton and Green 1999]. The error matrix is composed of an equal number of rows and columns (a square matrix of $n \times n$ dimension), where n – the number of object classes on a control map or ground referencing information.

Evaluation of the classification quality

Another area that is continuously drawing an increased attention of remote sensing specialists is the classification accuracy [Lillesand et al. 2004]. Reviews of accuracy statistics can be misleading if we don't understand what they represent. In this section we will describe and compare some of the ways in which accuracy is represented. We will start with a simple example. ILWIS supplies a method of assessing accuracy by error matrix based on test areas (ground truth). By defining the ground truth mask, ILWIS generates an error matrix automatically.

Evaluation of the supervised classification quality for the training data does not pose any problems. However, it does not provide reliable information on the result quality. Verification of the recognition quality is conducted on a prepared test data set, that is, a vector set and the expected response of a classifier that have not been used previously in the training process [Tymków 2009].

In the analysis of remote sensing data we used two rasters: a thematic land cover maps derived from automatic supervised classification (Maximum Likelihood method) of images and a digital reference topographic map. When interpreting the results we assumed that the classified result is potentially inaccurate, and that the ground truth well reflects the real situation. Otherwise, if the ground truth is imperfect, we cannot speak about the "error", and speak about the "difference" between the two sets of data instead [Bogoliubova 2012]. The values in the cells along the diagonal represent counts for correctly classified pixels, where the reference data (column) land cover type matches the mapped land cover type (row) pixels.

Various measures can be applied to the quantitative description of classification quality. Researches [Adamczyk and Będkowski 2005] postulate that the error matrix method should be employed. The matrix determines a number of points belonging to class j , which were classified as points belonging to class i . The error matrix structure can be presented as follows [Kubik et al. 2008] (Tab. 1).

A number of parameters can be calculated on the basis of the defined error matrix [Congalton and Mead 1983, Congalton and Green 1999, Lillesand and Kiefer 2000, Huber-Moy et al. 2001, Adamczyk and Będkowski 2005, Kubik et al. 2008].

Table 1. The error matrix structure
Tabela 1. Struktura macierzy błędów

| Belongs to a class | According to the reference data | | | | | | Row total |
|-----------------------------|---------------------------------|----------|----------|----------|-------|----------|-----------|
| | 1 | 2 | 3 | ... | M | | |
| According to the classifier | 1 | D_{11} | N_{12} | N_{13} | ... | N_{1M} | R_1 |
| | 2 | N_{21} | D_{22} | N_{23} | ... | N_{2M} | R_2 |
| | 3 | N_{31} | N_{32} | D_{33} | ... | N_{3M} | R_3 |
| | ... | ... | ... | ... | ... | ... | ... |
| | M | N_{M1} | N_{M2} | N_{M3} | ... | D_{MM} | R_M |
| Column total | C_1 | C_2 | C_3 | ... | C_M | N | |

Cells located in the diagonal of the errors matrix, are the number of correctly classified pixels (D_{ii}). A measure for the overall classification accuracy (T) can be derived from this table by counting how many pixels were classified as the same in the satellite image or aerial image and on the ground (ΣD_{ii}) and dividing this value by the total number of pixels ($N = \Sigma R_i = \Sigma C_j$). It's expressed as a percentage:

$$T = \frac{\Sigma D_{ii}}{N}, \quad (1)$$

where: ΣD_{ii} – the total number of correctly classified pixels,
 N – total number of pixels in the error matrix.

The producer's accuracy is a reference-based accuracy that is computed by reviewing the predictions produced for a class and by establishing the percentage of correct predictions:

$$PA = \frac{D_{ij}}{R_i}, \quad (2)$$

where: PA – producer's accuracy,
 D_{ij} – number of correctly classified pixels in row i (in the diagonal cell),
 R_i – total number of pixels in row i .

The user's accuracy is a map-based accuracy that is computed by reviewing the reference data for a class and establishing the percentage of correct predictions for these samples. It can be calculated according to the following formula:

$$UA = \frac{D_{ij}}{C_j}, \quad (3)$$

where: UA – user's accuracy,
 D_{ij} – number of correctly classified pixels in column j (in the diagonal cell),
 C_j – total number of pixels in column j .

In order to assess the accuracy of land cover maps extracted from satellite data, the stratified random sample strategy was employed which is a measure of unbiased assessment [Jensen 2005]. So, it is clear that the accuracy assessment of the classification is

largely dependent on how the training sets were generated. More complete measure of the classification accuracy is Kappa coefficient, also known as Kappa hat or K-hat (\hat{K}) [Tymków 2009]. This coefficient compares the number of pixels in each cell in the error matrix with the possibility to distribute pixels as a random variable. A non-parametric Kappa test was used to measure the classification accuracy as it accounts for all the elements in the confusion matrix rather than the diagonal elements [Rosenfield and Fitzpatrick-Lins 1986]. The Kappa coefficient was calculated according to the formula (4):

$$\hat{K} = \frac{N \sum_{i=1}^m D_{ij} - \sum_{i=1}^m R_i \cdot C_j}{N^2 - \sum_{i=1}^m R_i \cdot C_j}, \quad (4)$$

where: \hat{K} – Kappa-coefficient,
 N – total number of pixels,
 m – number of classes,
 $\sum D_{ij}$ – total diagonal elements of an error matrix (the sum of correctly classified pixels in all images),
 R_i – total number of pixels in row i ,
 C_j – total number of pixels in column j .

The registered value reflects overall classification accuracy, and random coincidences – consistency assessment between an aerial image and a referenced raster with a random distribution of pixels in classes [Chandr and Gosh 2008]. To obtain this estimate the error matrix is considered. Its elements are based on the total number of pixels in the corresponding rows and columns.

Landis and Koch [1977] offer a different interpretation of agreement (Tab. 2) [Tymków 2009].

Table 2. Interpretation of agreement for Kappa coefficient

Tabela 2. Interpretacja współczynnika Kappa

| Value of Kappa-hat | Interpretation of agreement |
|-------------------------------|-----------------------------|
| $0,81 \leq \hat{K} \leq 1$ | almost perfect agreement |
| $0,61 \leq \hat{K} \leq 0,80$ | substantial agreement |
| $0,41 \leq \hat{K} \leq 0,60$ | moderate agreement |
| $0,21 \leq \hat{K} \leq 0,40$ | fair agreement |
| $0,0 \leq \hat{K} \leq 0,20$ | slight agreement |
| $\hat{K} < 0,0$ | poor agreement |

Sample evaluation of the classification quality

The cells of the error matrix contain the number of pixels, which is based on information received from the land cover map and the reference raster. Thus, if a pixel is classified as belonging to the class “water bodies” of the classified map and to the class of “forests and vegetation” in the reference raster, it will be considered in the cell of errors matrix and will be located in the first column of the second row (the cell is defined as $n_{1,2}$). The error matrix for land cover classification of the Yuntolovsky reserve (2012) on the strength of Maximum Likelihood algorithm is presented in Table 3.

The cells of the error matrix contain the number of pixels, which is based on information received from the land cover map and the reference raster. Thus, if a pixel is classified as belonging to the class of “water bodies” of the classified map and to the class of “forests and vegetation” in the reference raster, it will be considered in the cell of errors matrix and will be located in the first column of the second row (the cell is defined as $N_{1,2}$).

The sum of the diagonal elements of the errors matrix shows the total number of correctly classified pixels $\sum D_{ii} = 1.098.712.113$ pixels. It's expressed as a percentage:

$$T = \frac{1098712113}{1536000000} = 0,715 = 71,5\%.$$

The drawback of this measure is that it does not tell us anything about how precisely individual classes were categorized. The user's and producer's accuracy are two widely used measures of class accuracy. The producer's accuracy refers to the probability that a certain land cover of an area on the ground is classified as such, while the user's accuracy refers to the probability that a pixel labeled as a certain land cover class on the map is really this class on the earth surface.

The user's and producer's accuracy for any given class typically are not the same. In the above example, an estimate for the producer's accuracy of the “artificial surface” is:

$$PA = \frac{204757017}{338739836} = 60,4\%.$$

The producer's accuracy varies greatly between different classes of land cover. This is due to the fact that some classes are more spectrally distinguishable than others, and are classified more accurately.

While the user's accuracy is:

$$UA = \frac{204757017}{42415396} = 82,8\%.$$

As users of classification, we can expect that roughly 82.8% of all the pixels classified as “artificial surface” are indeed artificial surface on the ground. However, as producers, we are not quite settled by the fact that we only classified 60.4% of all the artificial surface pixels as such.

On the other hand, the producer's accuracy for a class may be high and the user's accuracy – low. This means that the interpreter does the quality work of classifying by selecting which training pixels belong to a particular class. However, the results of this classification were not accurate enough.

Table 3. Confusion matrix for pixel-based image classification
 Tabela 3. Macierz niezgodności dla klasyfikacji obrazu w podejściu per-piksel

| Remote sensing classification, pixels | Ground referencing information, pixels | | | | | | Accuracy criteria | | |
|---------------------------------------|--|-------------------------|-----------------|-----------------|--------------------|-----------------------|-------------------------|-------------------------|--|
| | Water bodies | Forests and Vegetations | Wetland/Lowland | Cultivated land | Artificial surface | Row total, pixels | Producer accuracy [%] | Omission error [%] | |
| Water bodies | 446 719 814 | 10 638 995 | 19 573 229 | 218 145 | 3 138 051 | 480 288 234 | 93.0 | 7.0 | |
| Forests and Vegetations | 15 762 970 | 185 437 482 | 6 205 003 | 33 839 888 | 3 601 810 | 244 847 154 | 75.7 | 24.3 | |
| Wetland/Lowland | 23 847 248 | 11 154 462 | 122 364 597 | 21 310 307 | 28 880 734 | 207 557 348 | 59.0 | 41.0 | |
| Cultivated land | 168 052 | 110 096 789 | 8 074 583 | 139 433 203 | 6 794 801 | 264 567 429 | 52.7 | 47.3 | |
| Artificial surface | 0 | 465 375 | 87 031 630 | 46 485 814 | 204 757 017 | 338 739 836 | 60.4 | 39.6 | |
| Column total, pixels | 486 498 084 | 317 793 103 | 243 249 042 | 241 287 356 | 247 172 414 | 1 536 000 000 | 1 098 712 113 | K=63.7% | |
| User accuracy [%] | 91.8 | 58.44 | 50.3 | 57.8 | 82.8 | <u>I=71.5%</u> | <u>MPA=68.2%</u> | <u>MUA=68.2%</u> | |
| Commission error [%] | 8.2 | 41.6 | 49.7 | 42.2 | 17.2 | - | - | - | |

The data in Table 3 shows that from 1.536.000.000 pixels only 1.098.712.113 pixels (sum of diagonal elements) are identified correctly. Thus, the overall classification accuracy (T) was 71.5%. Assuming a binomial distribution of this estimate a one-sided criterion can be formulated [Chandr and Gosh 2008]:

$$p = T - \left[1,645 \sqrt{\frac{(T)(q)}{N} + \frac{50}{N}} \right], \quad (5)$$

where: p – overall accuracy corresponding to 95% confidence interval,

T – the overall classification accuracy,

$q = 100 - T$,

N – sample size.

If the overall classification accuracy is higher than a predetermined threshold, the classification should be used with a significance level of 95%. Usually, the threshold value is used with 85% significance level. We calculate the one-sided test of significance for our case [Chandr and Gosh 2008]:

$$p = 71,5\% - \left[1,645 \sqrt{\frac{(71,5\%)(28,5\%)}{1536000000} + \frac{50}{1536000000}} \right] \approx 71,5\%.$$

Since in our case the overall accuracy value, the corresponding one-sided significance test, is 71.5%, the results of the classification should be recognized not quite satisfactory. This is understandable due to a large sample size ($N = 1.536.000.000$ pixels). Consequently, there is a necessity to improve the methodology to assess the overall accuracy of classification. Let us discuss this issue later.

The non-diagonal elements of the error matrix contain information about the commissions (false classification) error and omissions error in it. Omissions in classification are presented for each land cover class to the right of the diagonal elements. To obtain the omission estimate of the image classification for each class the values are summarized in the corresponding line, and then the resulting value is divided by the total number of pixels in its class. In our calculations, 133.982.819 pixels which belong to the class “artificial surface” were missing. Out of these, 465.375 pixels have been attributed to the “forests and vegetation” class and 87.031.630 pixels assigned to “wetlands” class and 46.485.814 pixels – to “cultivated land” class. Thus, the omission’s error of land cover image classification of the protected areas for the class “artificial surface” was 39.6%. [Bogoliubova 2012]

Similarly, to assess the commission classification we calculated the sum of nondiagonal elements in the column of the class, and then divided it by the total number of pixels in a given class of land cover. So, in the class “water bodies” there were 39.778.270 extra pixel, out of which 15.762.970 pixels in fact belonged to the class of the “forest and vegetation”, 23.847.248 pixels to the class “wetlands”, and 168.052 pixel belonged to the class of “cultivated land”. The analysis of experimental data shows that the commission error of classification was 8.2%.

Omissions and commissions errors of image classification for all land cover classes of the Yuntolovsky reserve of St. Petersburg (2012) are presented in Table 4.

Table 4. Omission and commission errors of image classification
 Tabela 4. Błędy nadmiaru i pominięcia dla obrazu klasyfikowanego

| Land cover type | Omission [%] | | | Commission [%] | | |
|-------------------------|----------------|---------------------------------|--------------------|----------------------|------------------------------------|----------------------|
| | Missing pixels | Total number of pixels in a row | Omission error [%] | False classification | Total number of pixels in a column | Commission error [%] |
| Water bodies | 33 568 420 | 480 288 234 | 7,0 | 39 778 270 | 486 498 084 | 8,2 |
| Forests and Vegetations | 59 409 671 | 244 847 154 | 24,3 | 132 355 621 | 317 793 103 | 41,6 |
| Wetland/Lowland | 85 192 751 | 207 557 348 | 41,0 | 120 884 445 | 243 249 042 | 49,7 |
| Cultivated land | 125 134 226 | 264 567 429 | 47,3 | 101 854 154 | 241 287 356 | 42,2 |
| Artificial surface | 133 982 819 | 338 739 836 | 39,6 | 42 415 396 | 247 172 414 | 17,2 |

As stated above the significance test p gives an evaluation of the land cover classification accuracy for the entire image. Significance test is required for each class of land cover. We formulated this significance test for each class as a two-sample test. The corresponding 95% confidence level for each class can be obtained using the formula that defines the limits of the confidence interval [Chandr and Gosh 2008]:

$$p = PA_i \pm \left[1,96 \sqrt{\frac{(T)(q)}{N} + \frac{50}{N}} \right], \quad (6)$$

where: p – classification accuracy for each class of land cover, corresponding to 95% confidence level, PA_i – classification accuracy of each class of land cover type, $q = 100 - T$, N – sample size.

Confidence intervals for omissions and false classification are presented in Table 5.

Note that all these accuracy measures are estimates for the true, unknown accuracies. Thus, it is reasonable to ask what the accuracy of these measures is. In the above example, is the user's accuracy of water bodies really 82.8% or could it also reasonably be 70%? In general, it is possible to construct confidence intervals for all of these accuracy measures to give an idea of these accuracies estimates. The width of these confidence intervals is influenced mainly by the sample size and by accuracy measures themselves.

Table 5. Confidence intervals for omissions and false classification
 Tabela 5. Przedziały ufności dla błędów ominięcia oraz błędnej klasyfikacji

| Land cover type | Correct classification | Omission [%] | | | False classification [%] | | |
|------------------------|------------------------|-------------------------------|-------------------------|-------------------------|------------------------------------|---------------------|-------------------------|
| | | Total number of pixels in row | Producer's accuracy [%] | 95% confidence interval | Total number of pixels in a column | User's accuracy [%] | 95% confidence interval |
| Water bodies | 446 719 814 | 480288 234 | 93.0 | 92.998–93.002 | 486 498 084 | 91.8 | 91.798–91.802 |
| Forests and Vegetation | 185 437 482 | 244 847 154 | 75.7 | 75.695–75.705 | 317 793 103 | 58.4 | 58.395–58.405 |
| Wetland/ Lowland | 122 364 597 | 207 557 348 | 59.0 | 58.993–59.007 | 243 249 042 | 50.3 | 50.294–50.306 |
| Cultivated land | 139 433 203 | 264 567 429 | 52.7 | 52.694–52.706 | 241 287 356 | 57.8 | 57.794–57.806 |
| Artificial surface | 204 757 017 | 338 739 836 | 60.4 | 60.395–60.405 | 247 172 414 | 82.8 | 82.795–82.805 |

Using a value of 85% as the minimum threshold of the correct classification, we can see that classes “forests and vegetation”, “wetlands”, “cultivated land”, “artificial surface” do not satisfy the formula. It is because the boundary values of the confidence interval are lower than 85%. This is understandable large sample size.

The value of Kappa coefficient for the aerial image of the Yuntolovsky reserve (2012) is $\hat{K} = 0.637$, which means that the previously obtained classification accuracy was 63.7%. Pixels over this area were randomly distributed into classes.

Therefore, based on Table 2, we have the substantial agreement of the automated interpretation of the aerial image reached by applying the maximum likelihood algorithm with the normal distribution function of data.

Methodology of accuracy assessment of automated interpretation of images by their division

The studies revealed that the value of the overall classification accuracy of the land cover types which amounts to 71.5% is not satisfactory. In order to increase the overall accuracy of the land cover automatic classification based on aerial images we used a new methodology of data preprocessing.

The methodology relies on the hypothesis that if a satellite or on aerial image is divided into m parts of equal area, to make for every m_i part of the training set, we perform classification based on the maximum likelihood method and evaluate the overall accuracy as the arithmetical mean. The result of overall accuracy increases with the division of the image into more parts. Obviously, this involves more time spent on the interpretation of these parts.

A hypothesis developed by the authors was tested in the experiment: the original aerial image of the Yuntolovsky reserve was divided into parts of the same area with a multiple of $i = 2$.

In the experiments we assume that the population is regarded as fixed. This makes sense since we have a certain classification that we want to compare with the true land cover. The random part comes into play when we select the sample. In simple random sampling, each sample is selected with equal probability (7%) and independently of the other sample. Consequently, samples which are uncorrelated and unbiased estimators are relatively easy to derive. Some scientists believe that spatial autocorrelation is a problem for deriving an unbiased estimator. However, it is samples that are random, not land cover and the errors made in the classification of results.

The source aerial image was consistently divided into 2, 4, 6, 8, 10, 12, 14, 16, 20 and 22 and 24 parts of equal area in consecutive order with the multiple of $i = 2$.

The supervised classification of the aerial image was performed according to the Maximum Likelihood method. For each m_i part the training set was created, which was 7% of each part of the source image. The experimental results of the study are shown in Table 6.

Table 6. The dependence of the overall accuracy value of the supervised classification from partitioning scheme of the aerial image

Tabela 6. Zależność wartości dokładności całkowitej klasyfikacji nadzorowanej od podziału zdjęcia lotniczego

| Number of m_i | Arithmetical mean value (average) of the overall classification accuracy [%] | Number of pixels in each part m_i of the aerial image | Number of pixels in the source aerial image |
|-----------------|--|---|---|
| 1 | 71.5 | 1 536 000 000 | 1 536 000 000 |
| 2 | 75.7 | 768 000 000 | 1 536 000 000 |
| 4 | 79.7 | 384 000 000 | 1 536 000 000 |
| 6 | 81.5 | 256 000 000 | 1 536 000 000 |
| 8 | 82.2 | 192 000 000 | 1 536 000 000 |
| 10 | 82.8 | 153 600 000 | 1 536 000 000 |
| 12 | 83.3 | 128 000 000 | 1 536 000 000 |
| 14 | 83.7 | 109 714 286 | 1 536 000 000 |
| 16 | 84.1 | 96 000 000 | 1 536 000 000 |
| 20 | 84.7 | 76 800 000 | 1 536 000 000 |

In Table 6 we can clearly see that the overall classification accuracy of a source aerial image obtained depends on the number of equal parts the source image was divided into. Thus, we notice that the overall supervised classification accuracy depends on the division of the source aerial image. The evaluation of the overall classification accuracy based on the division of the aerial image into a different number of parts is described in [Bogoliubova 2012].

RESULTS

The experiment results confirm the hypothesis we propose in this paper. The following results are shown in Figure 1 as the approximating curve, which given the high correlation values can describe the identified dependence.

The dependency is approximated by a logarithmic curve $T = a \cdot \ln x + b$, with coefficients $a = 4.3004$ and $b = 72.697$ where T – the overall classification accuracy, x – number of equal parts of the aerial image according to the partitioning scheme. Approximation of the experimental data in the ten-point scale reveals that the general formula of the overall classification accuracy depending on the number of equal parts of the aerial image with logarithmic and exponential approximations, respectively, is as follows:

$$T = 4.3004 \cdot \ln(x) + 72.697, \quad (7)$$

The approximating curve describes the identified dependence with the high value of the determination coefficient is $R^2 = 0.9678$.

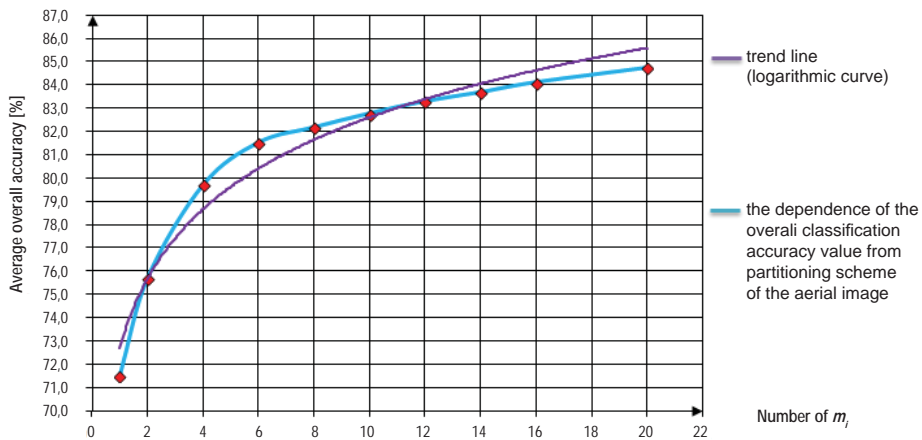


Fig. 1. The dependence of the overall classification accuracy value on the partitioning scheme of the aerial image (approximation logarithmic function)

Rys. 1. Zależność wartości dokładności całkowitej klasyfikacji nadzorowanej od podziału zdjęcia lotniczego (aproxymacja krzywą logarytmiczną)

In this paper we proposed, substantiated and implemented the method of estimating the overall accuracy of automated interpretation of aerial and satellite imagery. The method requires consistent division of the image into m equal parts with multiple 2. We also determined the dependence of the overall classification accuracy on the number of parts into which the original image is divided. This dependence shows that the accuracy of the classification increases with the division of image into parts, but the number of parts cannot be higher than 10. Further division does not significantly improve the accuracy.

DISCUSSION AND CONCLUSIONS

Errors in the classified map occur because the remotely-sensed data cannot capture the following classes:

- classes are not spectrally separable;
- atmospheric effects mask subtle differences;
- spatial scale of remote sensing instrument does not match classification scheme.

It is required to consider the following points when assessing the classification accuracy:

- choice of reference data is important;
- interaction between the sensor and desired classification scheme;
- error matrix is a foundation of accuracy assessment;
- all forms of accuracy assessment should be reported to a user;
- interpreting accuracy in classes can yield ideas for improvement of classification.

After analyzing the evaluation results of automatic classification of the protected areas land cover we can come to a conclusion that there is a clear classification of the experimental data of the contours of water bodies, forests and vegetation classes. And the “wetlands”, “cultivated land” and “artificial surface” classes overlap with one another.

This is due to the fact that the initial aerial image is the shadow of a cloud, which extends to the territory of forests, wetlands, and urban sprawl. The shadow gives the effect that the number of pixels in the test area classes, i.e. “wetlands”, “cultivated land” and “artificial surface” is almost the same.

The results are not satisfactory. Modification of the sample size, however, will affect the confidence interval. An increase in the sample size results in a higher precision of the estimated accuracy measures. Decreasing the sample size will reduce the precision.

Contrary to what has been stated in the literature, the total number of pixels in a satellite image has a minor (negligible) influence on the accuracy of the estimates. It is incorrect to say that one needs to increase the sample size because the area under investigation is very large.

The mean overall accuracy value was calculated for ten different data divisions. The diagram illustrating the dependency of this measure from the number of parts into which an image was divided is presented in Figure 1. The results were approximated with the logarithmic curve that is defined as follows:

$$T = 4.3004 \cdot \ln(x) + 72.697.$$

The dependence is described by the logarithmic curve because standard deviation is minimum (StDev = 3.94) and the determination coefficient is maximum ($R^2 = 0.9678$).

In order to increase the overall accuracy of the land cover automatic classification of an aerial image we used the proposed methodology of the overall classification accuracy evaluation. The proposed method of estimating the overall classification accuracy of the protected areas land cover increases its value on average by 10% by dividing the source aerial image into no more than 10 equal parts. With further partitioning of the image into parts of equal area the overall accuracy is slightly increased.

The result of overall accuracy increases with division of the image into equal area parts, and obviously there is also an increase in the time spent on the interpretation of such parts.

REFERENCES

- Adamczyk J., Będkowski K., 2005. *Metody cyfrowe w teledetekcji*. Wydawnictwo SGGW.
- Anderson J.R., Hardy E.E., Roach J.T., Witmer R.E., 1976. A land use and land cover classification system for use with remote sensor data. USGS Professional Paper 964, U.S. Geological Survey.
- Bogoliubova A.A., 2012. The aerospace monitoring of land cover protected areas in St.-Petersburg: Thesis...Ph.D. / National Mineral Resources University. St.-Petersburg, 144.
- Bogoliubova A., Tymków P., 2014. Land cover changes and dynamics of Yuntolovsky reserve, *Ejppau* 17(3), #03. Available Online: <http://www.ejppau.media.pl/volume17/issue3/art-03.html>
- Chandr A.M., Gosh C.K., 2008. Remote sensing and GIS. Moscow, Techno sphere, 312.
- Congalton R.G., 1991. A review of assessing the accuracy of classification of remotely sensed data. *Remote Sensing of Environment*, 37, 35–46.
- Congalton R.G., Green K., 1999. *Assessing the accuracy of remotely sensed data: Principles and practices*. Lewis Publishers, Boca Raton.
- Congalton R.G., Mead R.A., 1983. A quantitative method to test for consistency and correctness in photo interpretation. *Photogrammetric Engineering and Remote Sensing*, 49, 69–74.

- Jensen J.R., 1996. Introductory digital: A remote sensing perspective (2nd ed). Chapter 8, Thematic Information Extraction: Image Classification. Upper Saddle River, New Jersey: Prentice-Hall, 234–252.
- Jensen J.R., 2005. Introductory digital image processing: A remote sensing perspective (3rd ed.). Upper Saddle River, New Jersey, Prentice Hall.
- Jensen J.R., Cowen D.C., 1999. Remote sensing of urban/suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing*, 65 (5), 611–622.
- Hubert-Moy L., Cotonnac A., Du L., Chardin A., Perez P., 2001. A comparison of parametric classification procedures of remotely sensed data applied on different landscape units. *Remote Sensing of Environment*, 75, 174–187.
- Kubik T., Paluszynski W., Iwaniak A., Tymków P., 2008. Klasyfikacja obrazów rastrowych z wykorzystaniem sztucznych sieci neuronowych i statystycznych metod klasyfikacji. Wydawnictwo Uniwersytetu Przyrodniczego we Wrocławiu, 86.
- Landis J., Koch G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lillesand Thomas M., 2004. Remote Sensing and Image Interpretation (5th ed.). Chapter 7, Digital Image Processing. New York: John Wiley and Sons, 550–610.
- Lillesand T.M., Kiefer R.W., 2000. Remote Sensing and Image Interpretation, 4th ed. Wiley & Sons.
- Lillesand T.M., Kiefer R.W., Chipman J.W., 2004. Remote sensing and image interpretation. 5th Edn. John Wiley and Sons, Inc. USA.
- Rosenfield G.H., Fitzpatrick-Lins K., 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 52(2), 223–227.
- Tymków P., 2009. Application of photogrammetric and remote sensing methods for identification of resistance coefficients of high water flow in river valleys. Wydawnictwo Uniwersytetu Przyrodniczego we Wrocławiu, 102.
- Volkova E.A., 2005. Regional complex natural state reserve «Yuntolovsky» / Saint-Petersburg, 202.

OCENA DOKŁADNOŚCI AUTOMATYCZNEJ KLASYFIKACJI POKRYCIA TERENU DLA OBSZARU CHRONIONEGO SANKT PETERSBURGA

Streszczenie. W pracy dokonano analizy sposobów oceny jakości klasyfikacji pokrycia terenu na danych obrazowych. Autorzy wysunęli hipotezę, że ogólna dokładność klasyfikacji obrazu zależy od jego podziału w procesie klasyfikacji na podobszary. Zależność tę opisano krzywą logarymiczną $T = 4,3004 \cdot \ln(x) + 72,697$, dla której uzyskano najwyższy współczynnik determinacji ($R^2 = 0,9678$). Badania prowadzono dla rezerwatu Yuntolovo, chronionego obszaru w pobliżu Sankt Petersburga (Rosja). W celu zwiększenia ogólnej dokładności automatycznej klasyfikacji pokrycia terenu na podstawie zdjęć lotniczych autorzy zaproponowali nową metodologię wstępnego przetwarzania danych. Proponowana metoda, polegająca na podziale obrazu klasyfikowanego na nie więcej niż dziesięć równych części, poprawia ogólną dokładność klasyfikacji pokrycia obszarów lądowych średnio o 10%. Podział na większą liczbę części nie zwiększa już znacząco jakości klasyfikacji, a dodatkowo wprowadza niejednoznaczności spowodowane zmniejszaniem próby uczącej. Klasyfikację obrazów i analizę dokładności prowadzono z wykorzystaniem pakietu ILWIS 3.31 oraz autorskiego oprogramowania stworzonego w środowisku NET.

Słowa kluczowe: dokładność całkowita klasyfikacji, automatyczne przetwarzanie obrazów, obszary chronione, klasyfikacja pokrycia/użytkowania terenu, klasyfikacja nadzorowana

Accepted for print – Zaakceptowano do druku: 30.03.2014

For citation – Do cytowania: Bogoliubova A., Tymków P., 2014. Accuracy assessment of automatic image processing for land cover classification of St. Petersburg protected area. *Acta Sci. Pol. Geod. Descr. Terr.*, 13 (1-2), 5–22.