Savelieva E., Bolshov L., Pozdnukhov A., Timonin V., Kanevski M., Chernov S., Murray Ch., Thorne P., Xie Y. ZONATION OF HANFORD FORMATION AT THE HANFORD SITE. Preprint IBRAE-2003-07. Moscow: Nuclear Safety Institute RAS, March 2003. 55 p.

Abstract

The current work is devoted to application of different approaches to classification (zonation) of the Hanford formation – the uppermost geologic layer in the Hanford site (Washington, USA). Approaches based on machine learning (artificial neural networks and statistical learning theory) were used together with geostatistical methods (indicator approach) for the current task. The problem was significantly complicated by low number of samples, especially for some classes. Additional available information was also used. Advantages and drawbacks of used methods are discussed in the paper.

Obtained results were compared with a map built by experts-geologists using additional geologic knowledge on the region under study. The comparison was performed with the help of the simplest methods developed for image analysis.

Савельева Е., Большов Л., Позднухов А., Тимонин В., Каневский М., Чернов С., Мюррей К., Торн П., Ксай Ю. ЗОНИРОВАНИЕ ХЭНФОРДСКОЙ ОСАДОЧНОЙ ФОРМАЦИИ НА ХЭНФОРДСКОМ ПОЛИГОНЕ. Препринт № IBRAE-2003-07. Москва: Институт проблем безопасного развития атомной энергетики РАН, 2003. 55 с.

Аннотация

Работа посвящена использованию различных подходов для классификации (зонирования) хэнфордской осадочной породы – верхнего геологического слоя Хэнфорского полигона (шт. Вашингтон, США). В работе обсуждены достоинства и недостатки различных подходов, основанных на методах машинного обучения (искусственные нейронные сети и статистическая теория обучения) и на методах геостатистики (индикаторный подход). Проблема зонирования усложняется малым количеством исходных измерений, особенно для некоторых классов. При зонировании использовалась дополнительная информация.

Результаты, полученные различными методами, сравнивались с картой, построенной экспертами-геологами с учетом знания геологических особенностей исследуемого региона. Для сравнения использовались простейшие подходы, разработанные для анализа изображений.

Zonation of Hanford formation sediment at the Hanford Site

E. Savelieva[1], L. Bolshov[1], A. Pozdnukhov[1], V.Timonin[1], M. Kanevski[1], S. Chernov[1], C. Murray[2], P. Thorne[2], Y. Xie[2]

[1] THE INSTITUTE OF NUCLEAR SAFETY
113191, Moscow, B. Tulskaya, 52

[2] PACIFIC NORTHWEST NATIONAL LABORATORY
Richland, Washington, USA

tel.: (095) 955-22-31, fax: (095) 958-11-51, E-mail: esav@ibrae.ac.ru;
http://www.ibrae.ac.ru/~mkanev

Contents

Introduction

Hanford site aquifer (gravel and sand) units are not homogeneous systems. Hydraulic parameters differ within an aquifer unit and are spatially variable. Nevertheless, zones with similar values of hydraulic parameters (parameter zones) can be distinguished. This parameter zonation approach is an alternative to analysis of the spatial variation of the continuous hydraulic parameters. The parameter zonation approach is primarily motivated by the lack of measurements that would be needed for direct spatial modeling of the hydraulic properties.

The main task of the current report is to adapt/develop techniques, based on geostatistical approaches and machine learning algorithms, for modeling of heterogeneous media in order to represent the spatial variability and uncertainty of the parameter zones. The zonation problem was formulated as a multiclass classification task. The case study containing the application of methods and their comparison was performed for the Hanford formation (Unit 1) zonation at the Hanford site.

The following methods, some of which are extensions of binary classification approaches, were generalized to solve the multiclass classification task:

- *k* nearest neighbors (KNN),
- probabilistic neural network (PNN),
- geostatistical methods (indicator approach: indicator kriging, simulations),
- artificial neural network multiplayer perceptron (MLP)
- statistical learning theory: Support Vector Machine (SVM).

All methods were compared using different approaches selected for the current case. The small number of initial data, particularly for some zones of Unit 1, does not allow for use of the traditional approach for extraction of a subset of the data for validation.

1. Description of data

The Hanford formation (U1) is the uppermost sedimentary unit included in hydrogeologic models of the Hanford site. U1 is presented by 5 zones with different hydraulic properties: class 1 – Hanford formation Gravel Type 1; class 2 – Hanford formation Gravel Type 2; class 3 – Hanford formation Gravel Type 3; class 4 – Hanford formation Sand; class 5 – Hanford formation Silt.

The initial data on U1 zonation is a set of 225 samples. Each sample is ascribed to a class (parameter zone), and this ascribing is accompanied by a measure of data quality (expert information) that addresses the certainty with with the class assignment was made (the higher the value assigned – the lower the confidence in the assignment). Initial data on classes and confidences are presented in Fig. 1 by the help of Voronoi polygons, where each polygon takes on the value of the class assignment or rhe data quality.

In addition to the measured data we are supplied with additional important information about the zonation based on geological interpretation of the data (see Fig. 2). This information can be considered and used in a several different ways. One interesting approach is based on using this information as additional soft information for machine learning algorithms. Another approach is related to the use of multiple point geostastistics. In this case, geological drawings are used as reference images to develop joint probability density functions. Here again, machine learning algorithms can be used as nonlinear estimators. At present, such studies are outside of the scope of the work. In the current work, the geological map is used for comparison of the results provided by different models. One type of comparison is connected with the spatial correlation structure of the geologic interpretation. Class indicator variogram roses of the geologic interpretation of the spatial distribution of the U1 zones are presented in Figs. 3 – 5.

The U1 zonation data set is impossible to partition into training and validation data subsets that would each retain the characteristics of the whole data set. This is caused by the different number of samples interpreted as belonging to a class. There are very few representatives of class2 (6), of class4 (15) and especially of class5 (only 1 member). So, the whole data set was used for training. Comparison of methods can be performed: 1) on the base of the accuracy test (application of the method to the training data), which can be used as a criteria with some caution, because the accuracy test gives only the empirical error for the training set and not the generalization error; and 2) comparisons/intercomparisons with geological map.

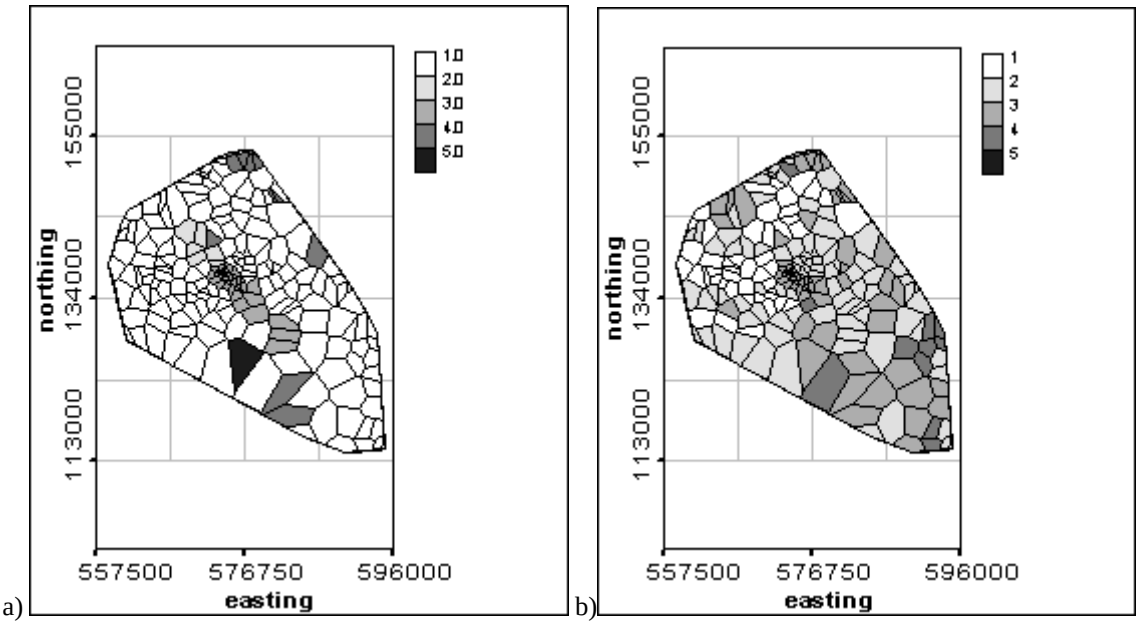


Fig. 1. Initial data: classes (a) and confidences (b)

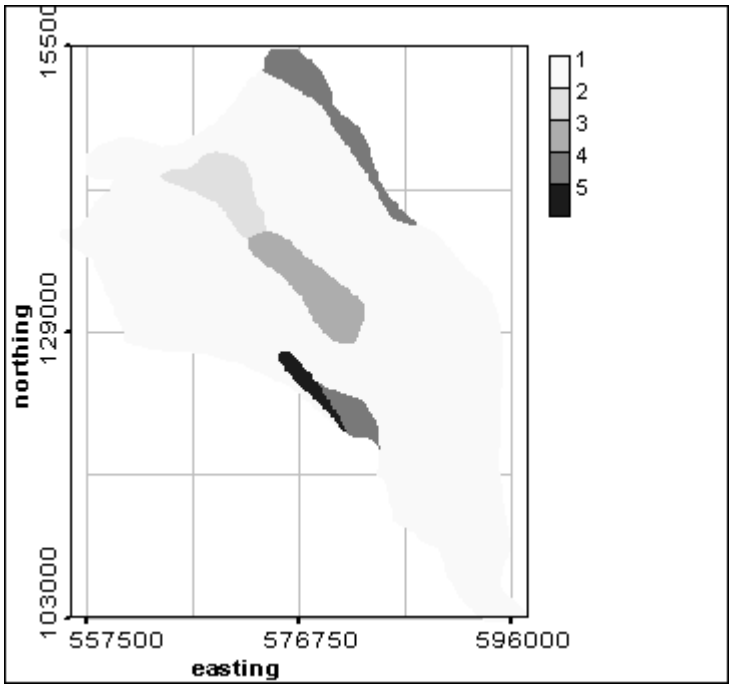


Fig. 2. Geologically based Unit 1 zonation

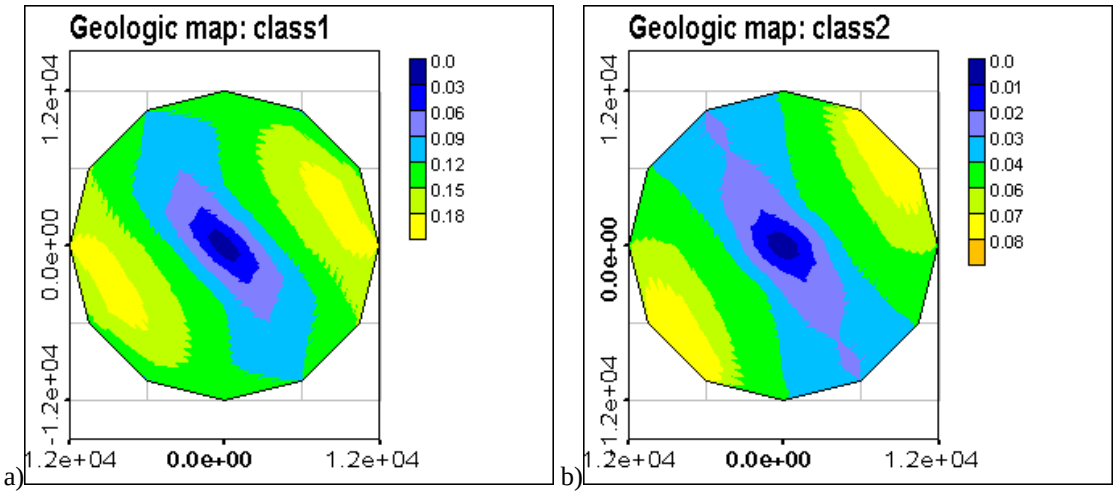


Fig. 3. Class indicator variogram rose for a priori geologic data: class1 (a) and class2 (b)

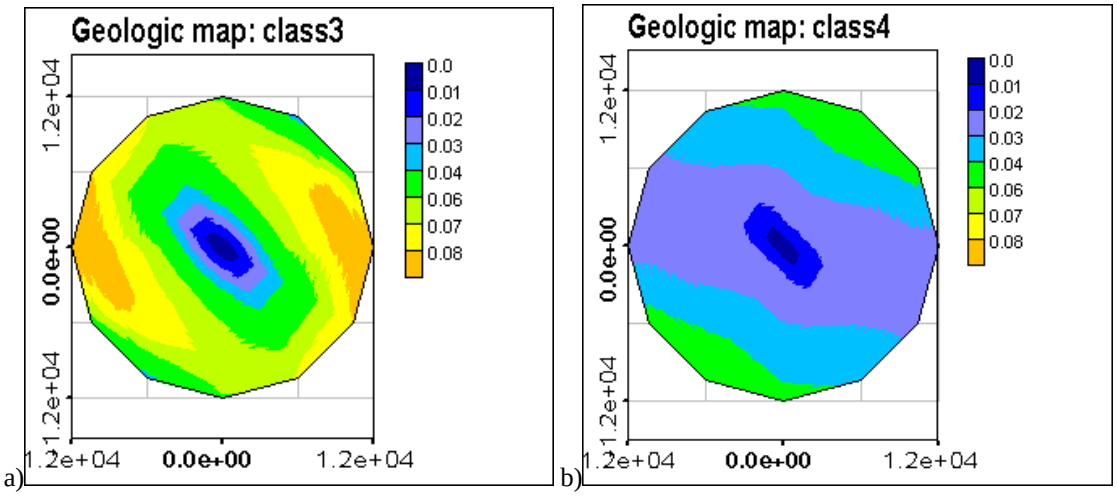


Fig. 4. Class indicator variogram rose for a priori geologic data: class3 (a) and class4 (b)

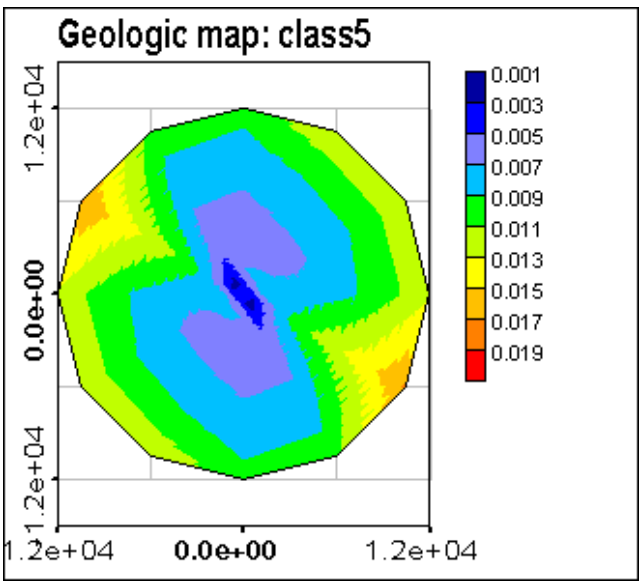


Fig. 5. Class indicator variogram rose for a priori geologic data: class5

2. K-nearest neighbors for U1 zonation

The optimal number of neighbors was searched using leave-one-out cross-validation and a bootstrap with 10 and 50 resamplings. The curves presenting these procedures are shown in Fig. 6. The minimums of curves provided by various approaches differ from each other. According to these results several candidates were selected as an optimal number of neighbors: 9 neighbors (based on cross-validation), 12 neighbors (based on a bootstrap with 50 resamplings) and 13 neighbors (based on a bootstrap with 10 resamplings). Results of the accuracy test for different numbers of neighbors are presented in Table 1. It is evident that there are many misclassifications when applying this method that tend to be distributed around the boundaries of the class zones.

Table 1. KNN accuracy test errors

Number of neighbors	Number of misclassifications	Error (%)
9	19	8.3
12	20	8.73
13	22	9.6

The results on the regular rectandular grid for 1 neighbor (k=1 corresponds to the Voronoi polygons) (see Fig. 7), and for 9 and 12 neighbors (see Fig. 8) also are of a questionable quality. 9 and 12 neighbors provide very high smoothing – all small clusters of the different classes disappear. 1NN is much closer to the geologic classification (at least all clusters are present), but it does not have the spatial orientation of the classes expected by the geologists – compare visibly with Fig.1a and Fig.2.

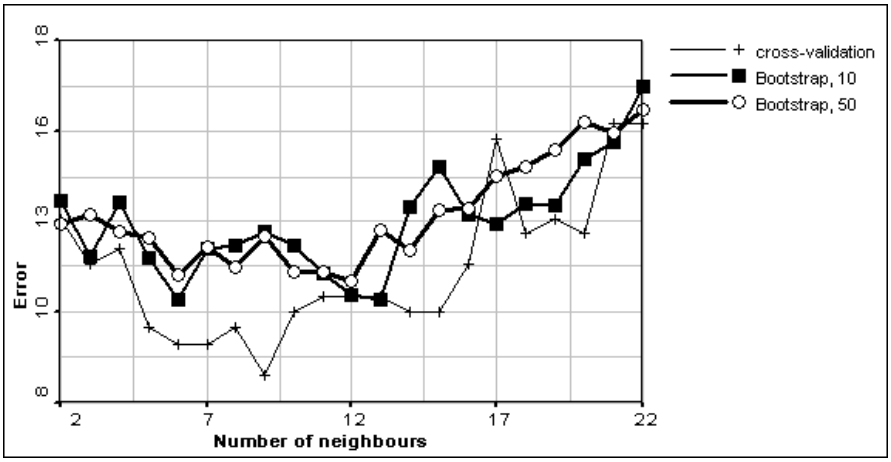


Fig. 6. Error curves to select the optimal number of neighbors for U1 zonation problem

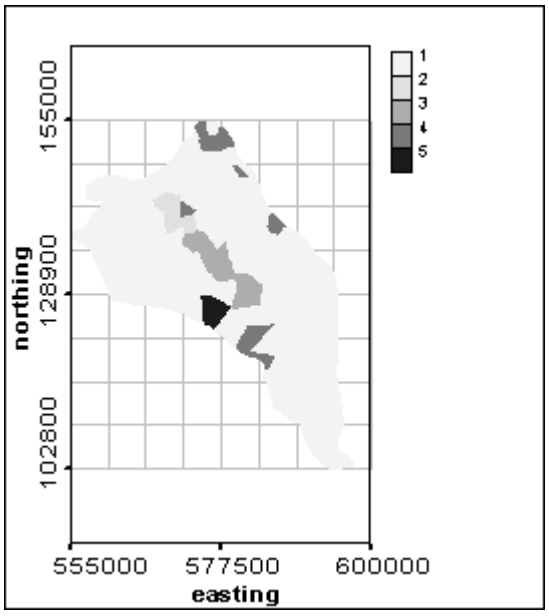
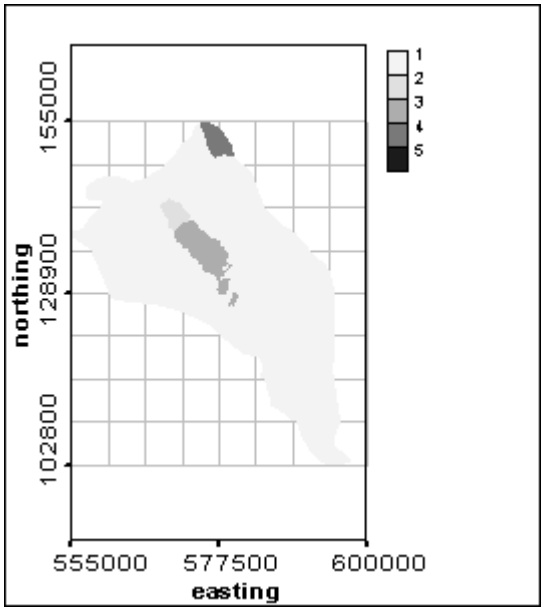


Fig. 7.U1 zonation by 1-NN method

a)



b)

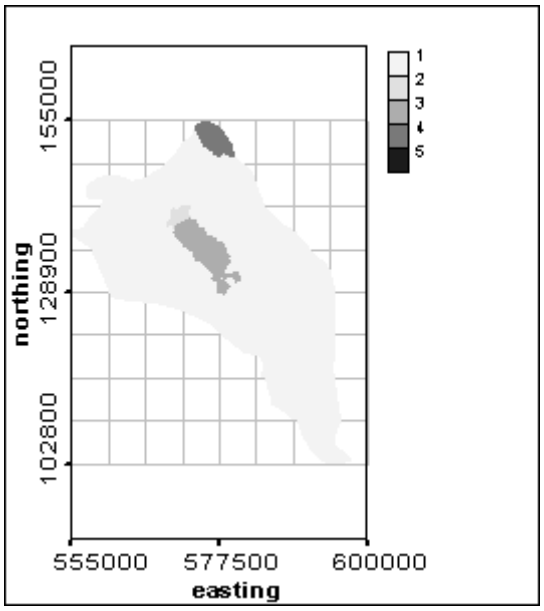


Fig. 8. U1 zonation by K-NN method with a) 9 neighbors; b) 12 neighbors

3. Probabilistic Neural Network for U1 zonation

The PNN approach does not need any generalization for the multiclass case, for there is no limitation on the number of classes in the basic theory. The detailed theory on PNN can be found in [1].

The PNN training procedure includes the following steps:

1. Perform leave-*k*-out cross-validation to obtain different kernel bandwidths in the X and Y directions with orientation of X-axis in west direction (an example is presented in Fig. 9);
2. Perform leave-*k*-out cross-validation on the rotation of the coordinate axes to select the best orientation (an example is presented in Fig. 10). Rotation is performed in the counterclockwise direction starting from the east-west direction corresponding to the X-axis (considered as 0 degrees);
3. Repetition of the first step and second steps until there are no changes in both kernel bandwidth and orientation;
4. Final gradient descent tuning of kernel bandwidths (an example is presented in Fig. 11).

Initially leave-*k*-out cross-validation was performed for *k=1*. Class size was not used while initializing the a priori class probabilities. The PNN parameters obtained after leave-1-out cross-validation are presented in Table 2. The sigma values indicated in Table 2 correspond to the kernel bandwidths.

Table 2. PNN parameters after leave-1-out cross-validation training

X-sigma	Y-sigma	Orientation	Training error
129.607	499.166	22	0%

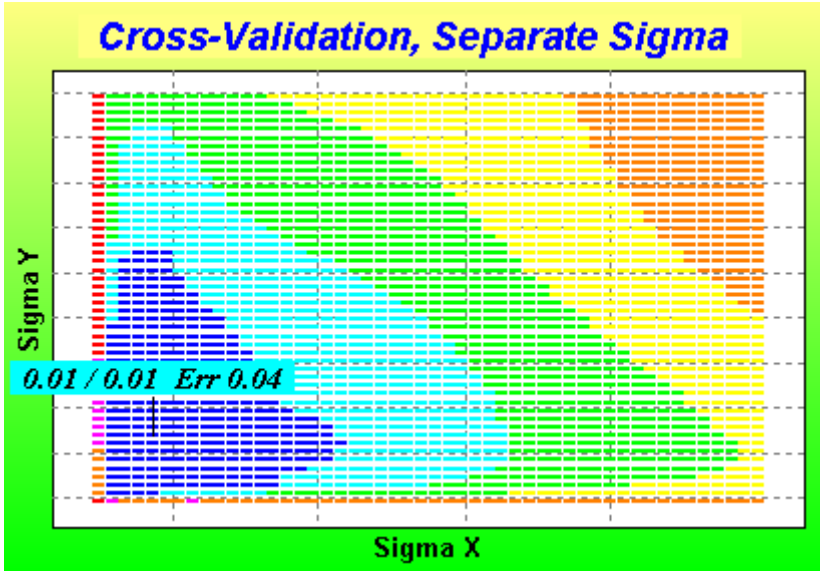


Fig. 9. Error surface for PNN kernel bandwidth adaptation by leave-one-out out cross-validation with X-axes oriented towards east-west direction

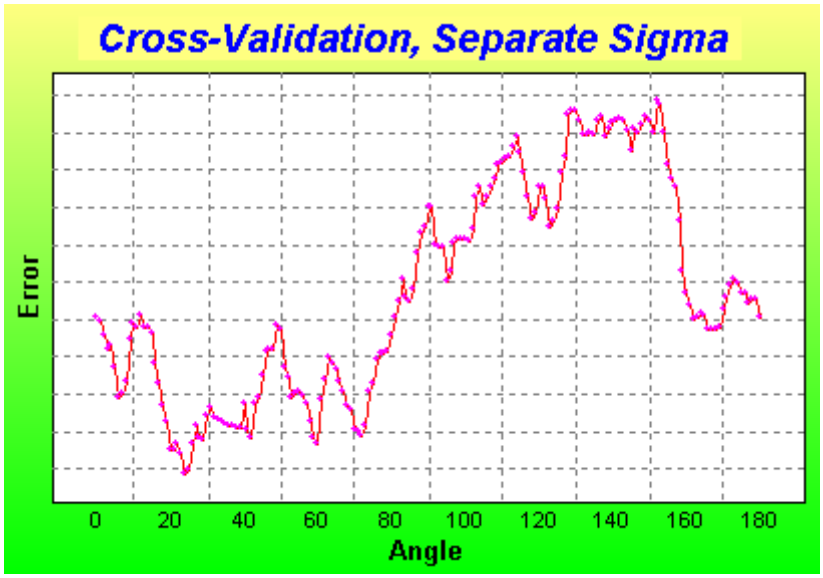


Fig. 10. Error curve for the coordinate system orientation after leave-one-out out cross-validation

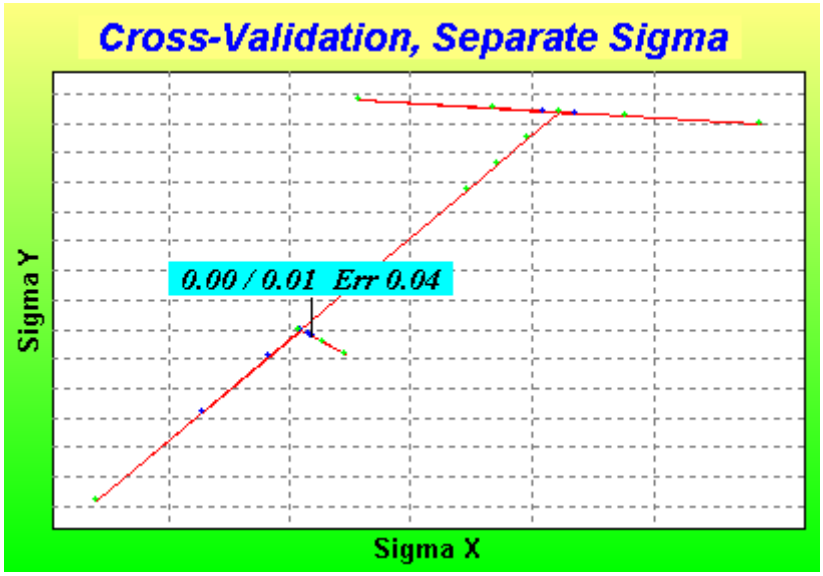


Fig. 11. Final adjusting of PNN kernel bandwidths by gradient decent method performed for preliminary results of leave-one-out cross-validation

The accuracy test for PNN was very successful with only 1 misclassification: a point belonging to class 1 was misclassified as class 4. This misclassification was not unexpected, because the data point is surrounded only by members of class 4. Results of U1 zonation using PNN on a regular grid are presented in Figs. 12 and 13, which show the classification made on the decision basis of a 0.5 threshold (Fig. 12) and the probability of the class winner being present (Fig. 13).

This PNN provides a zonation pattern similar to the geologic interpretation. However, the uncertainty zones provided by PNN seem to be too narrow (approximately one pixel thick and nearly invisible). The narrowness of the uncertainty bands may be the result of overfitting.

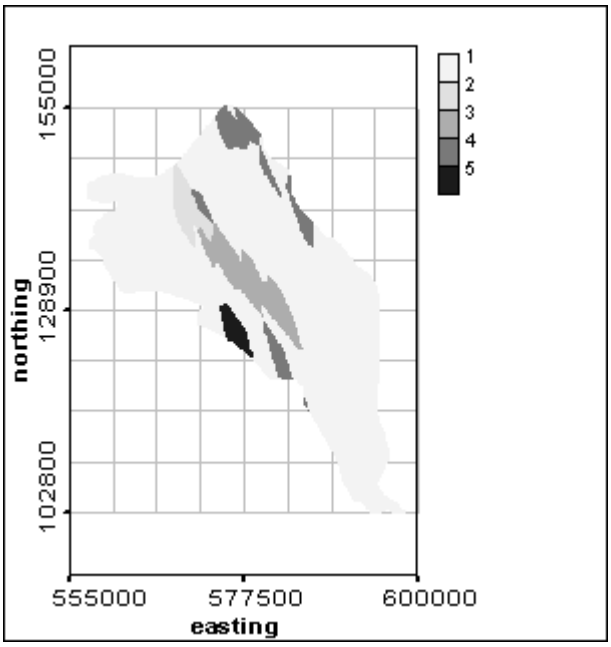


Fig. 12. U1 zonation by PNN trained using leave-one-out cross-validation

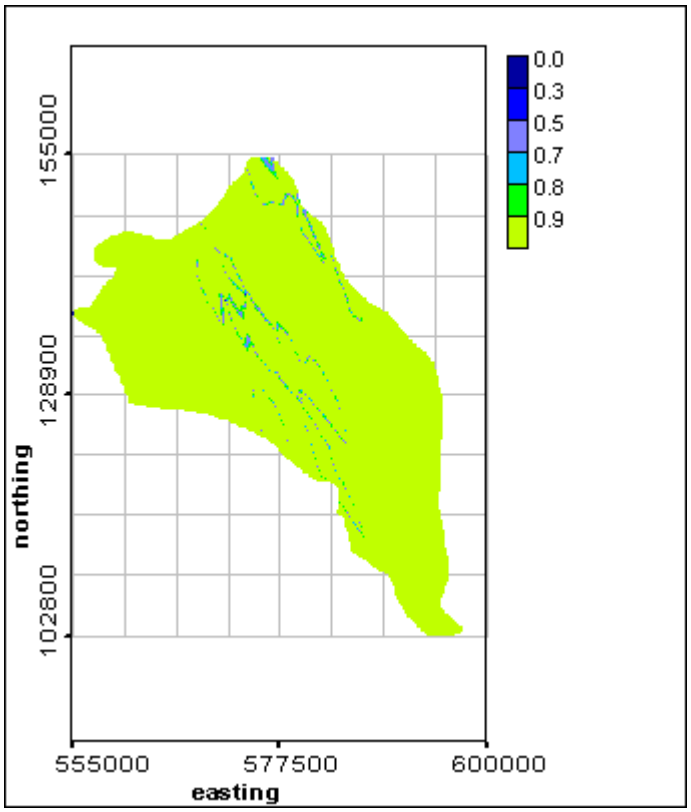


Fig. 13. Probability of the class winner according to the PNN trained using leave-one-out cross-validation

Because PNN trained using leave-one-out cross-validation appears to be overfitting the data, other methods of training PNN were also performed. Leave-k-out cross-validation methodology was used to select PNN parameters and to compare different results. Using more than one point in cross-validation introduces some regularization in the training procedure and it was expected to help in overcoming the overfitting.

The PNN tuning was performed with k set equal to 20, 15, 10, 8, 5, 4, 3, and 2. The training procedure was as described above. Table 3 summarizes the PNN parameters obtained after training, the cross-validation errors, and the results of accuracy tests.

The difference between the two errors presented in Table 3 is of great importance. The accuracy test error is related to the question "how well was the training dataset learned." Statistically, it corresponds to the Mean Square Error of the residuals. The accuracy test error, on the other hand, is the number of misclassifications found using the standard leave-1-out cross-validation procedure after the parameters were established using the leave-k-out training procedure for a given k.

The cross-validation error is the final (the smallest one) error achieved during the training procedure corresponding to the training algorithm. It is the averaged misclassification over all sets of k-left-out samples. It can be written as:

$$CVerror = \frac{\sum_{i=1}^{n} m_i}{kn},$$

where k is the number of samples being removed from the dataset using the leave-k-out algorithm, n is the number of cross-validation steps and m_i are numbers of misclassifications in each step. When the data are not clustered, the cross-validation error can be considered as an unbiased estimation of the generalization error, i.e., the prediction error. However, the U1 zonation data are clustered, so the estimate of the prediction error is biased. The dependence of this bias from k is unknown, so for the clustered data cross-validation error is only the method training stop point. But still it gives some feeling of what accuracy can be expected.

The U1 zonation results on the prediction grid obtained using PNN trained on different values of k are presented in Figs. 14 – 22. For each value of k, the figures provide both the classification result and the probability of the class winner being found for each node of the grid. For leave-k-out PNN, the uncertainty zones are very narrow again for nearly all values of k. This suggests that the narrow uncertainty zones are due to the configuation of the initial data, so that the classes appear to be easily separable by PNN.

Table 3. Parameters of PNN after training by leave-*k*-out cross-validation approach

k	σ_X	σ_Y	Orientation	CV error	Accuracy test (misclassifications)
20	77.265	407.958	41	0.41	0
15	364	131.138	164	0.23	1
10	77.265	1054.34	32	0.18	3
8	114.603	488.64	33	0.16	1
6	116.531	730.382	33	0.11	4
5	77.26	609.952	31	0.099	1
4	111.301	527.423	19	0.089	1
3	260.566	566.119	0	0.089	5
2	112.884	488.967	20	0.51	1

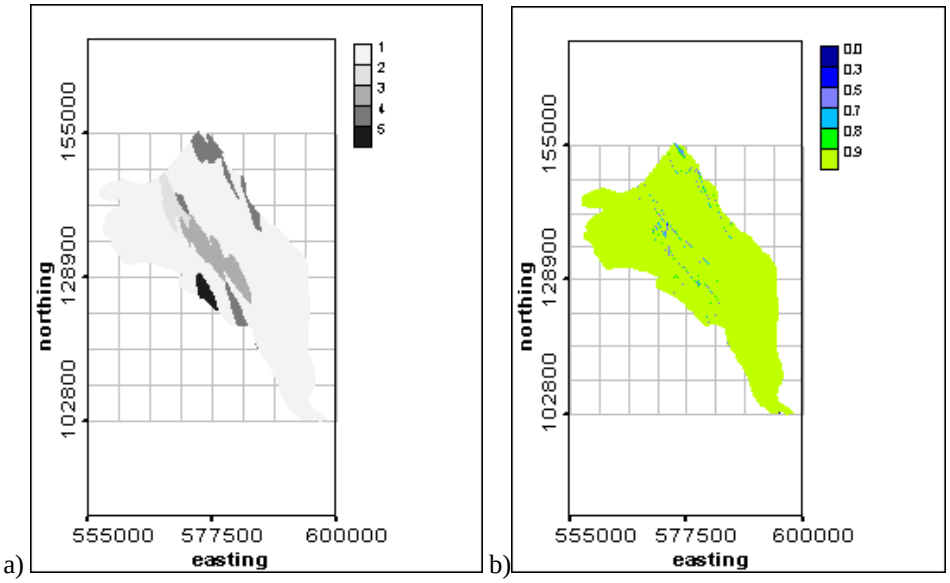


Fig. 14. U1 zonation by PNN trained by leave-2-out cross-validation: classification (a), probability of the class winner (b)

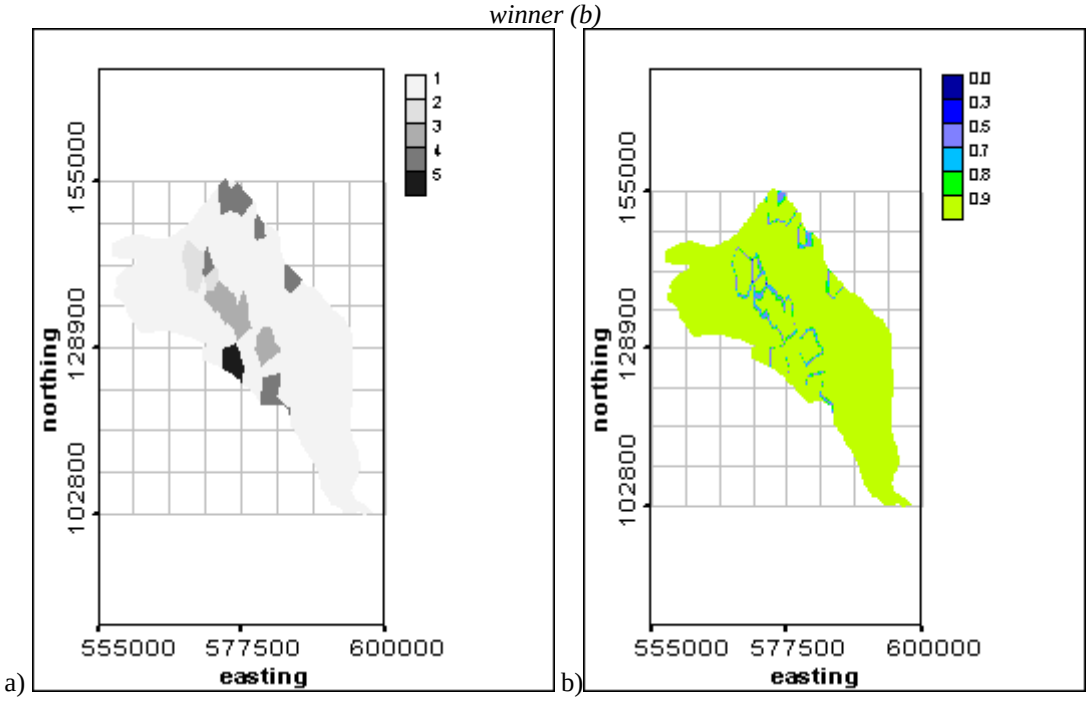


Fig. 15. U1 zonation by PNN trained by leave-3-out cross-validation: classification (a), probability of the class winner (b)

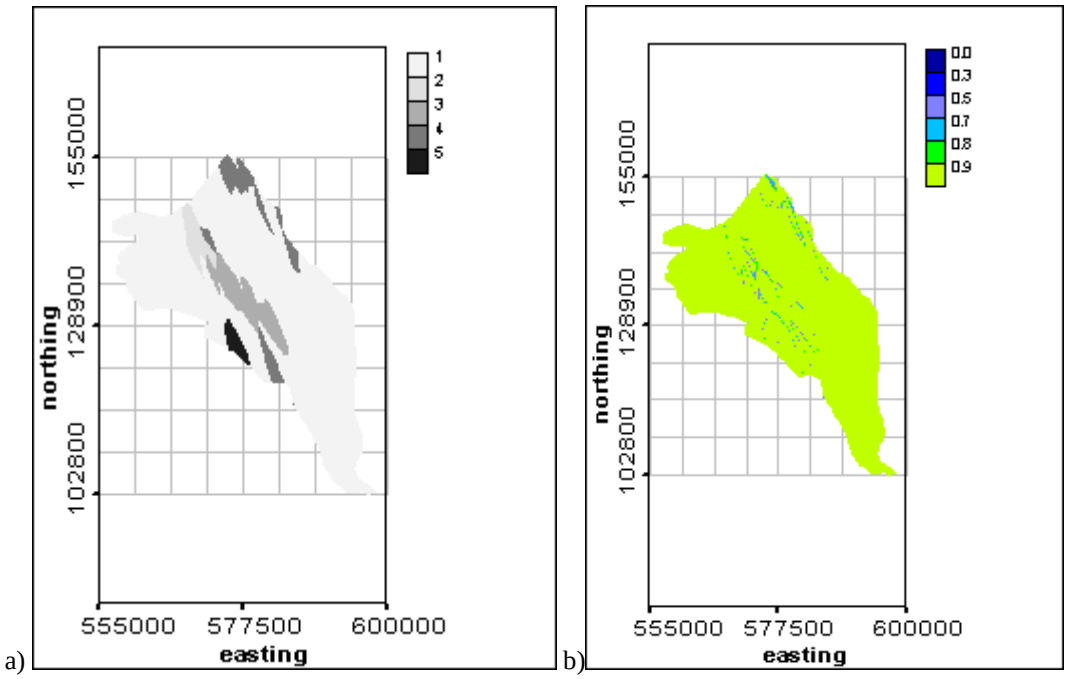


Fig. 16. U1 zonation by PNN trained by leave-4-out cross-validation: classification (a), probability of the class winner (b)

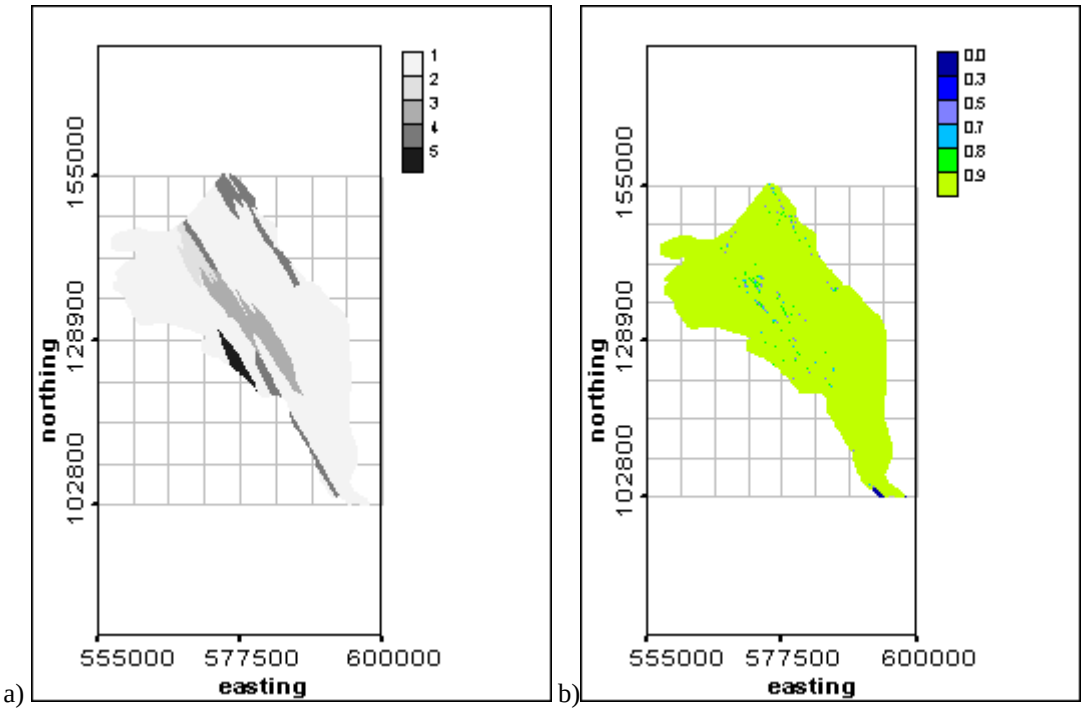


Fig. 17. U1 zonation by PNN trained by leave-5-out cross-validation: classification (a), probability of the class winner (b)

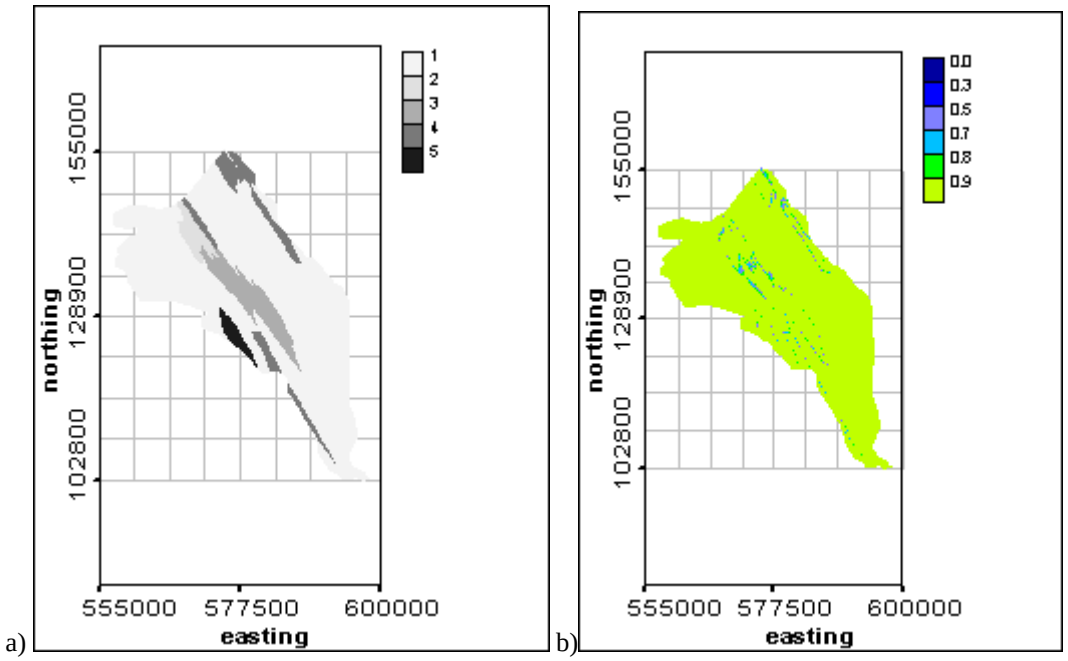


Fig. 18. U1 zonation by PNN trained by leave-6-out cross-validation: classification (a), probability of the class winner (b)

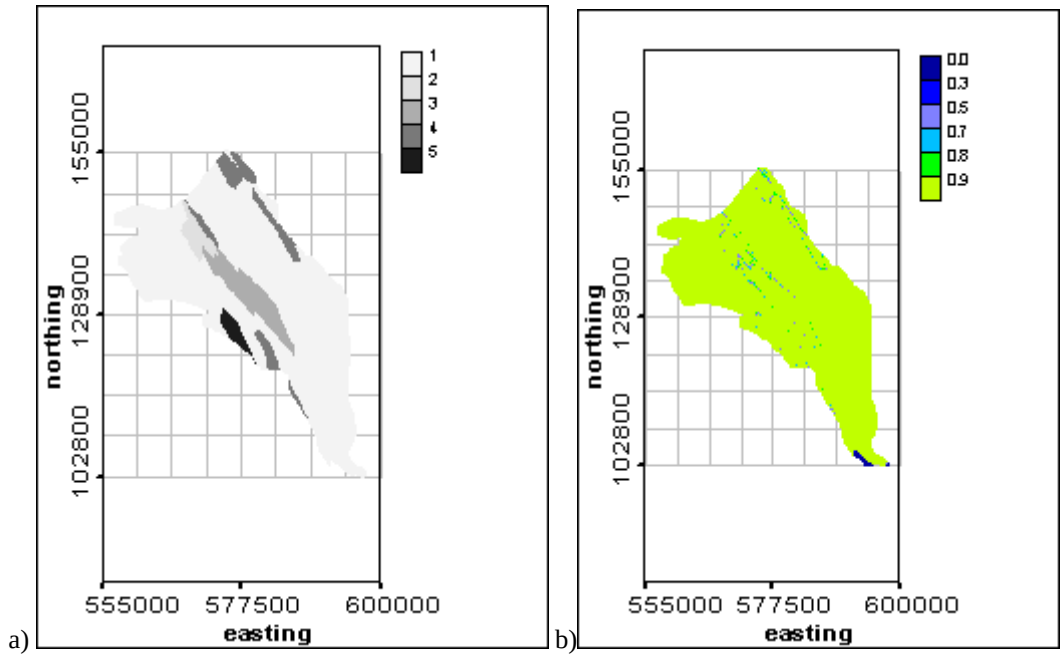


Fig. 19. U1 zonation by PNN trained by leave-8-out cross-validation: classification (a), probability of the class winner (b)

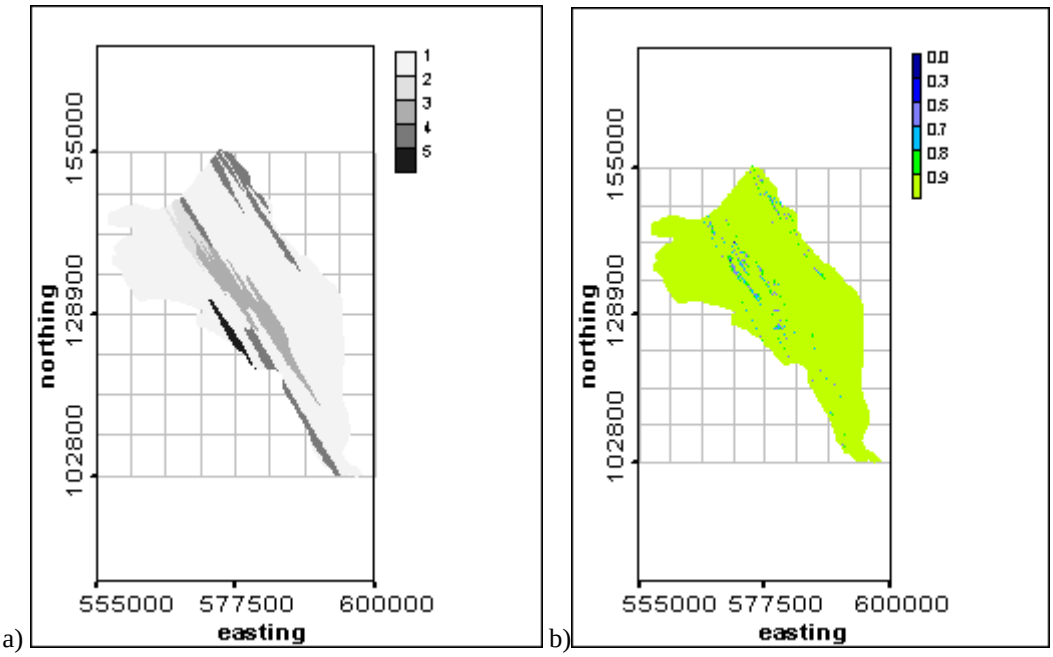


Fig. 20. U1 zonation by PNN trained by leave-10-out cross-validation: classification (a), probability of the class winner (b)

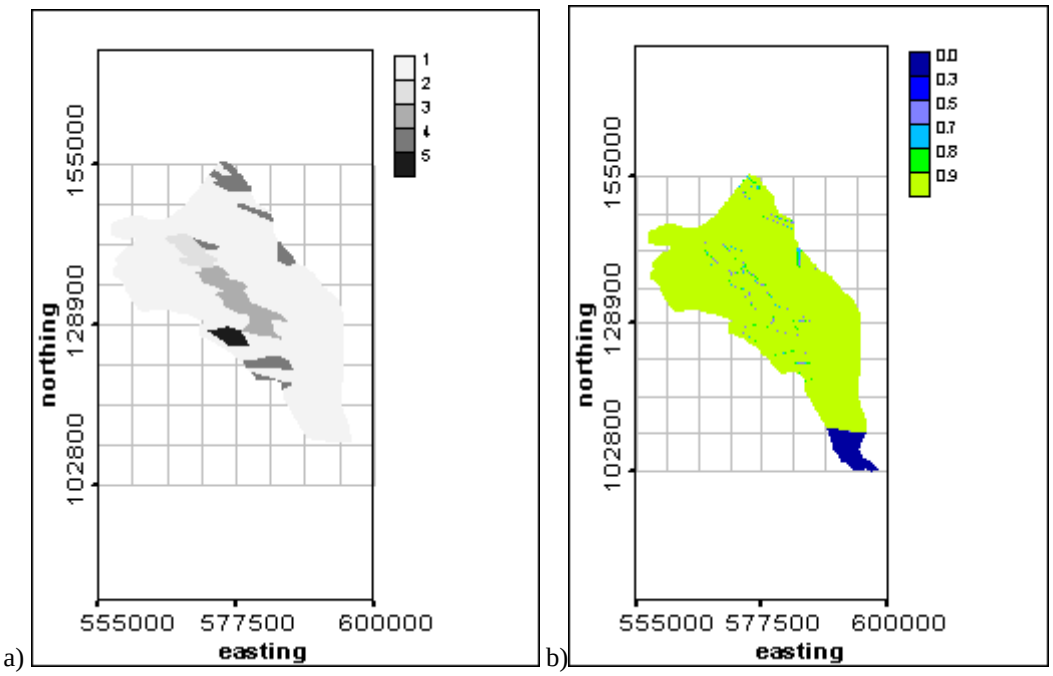


Fig. 21. U1 zonation by PNN trained by leave-15-out cross-validation: classification (a), probability of the class winner (b)

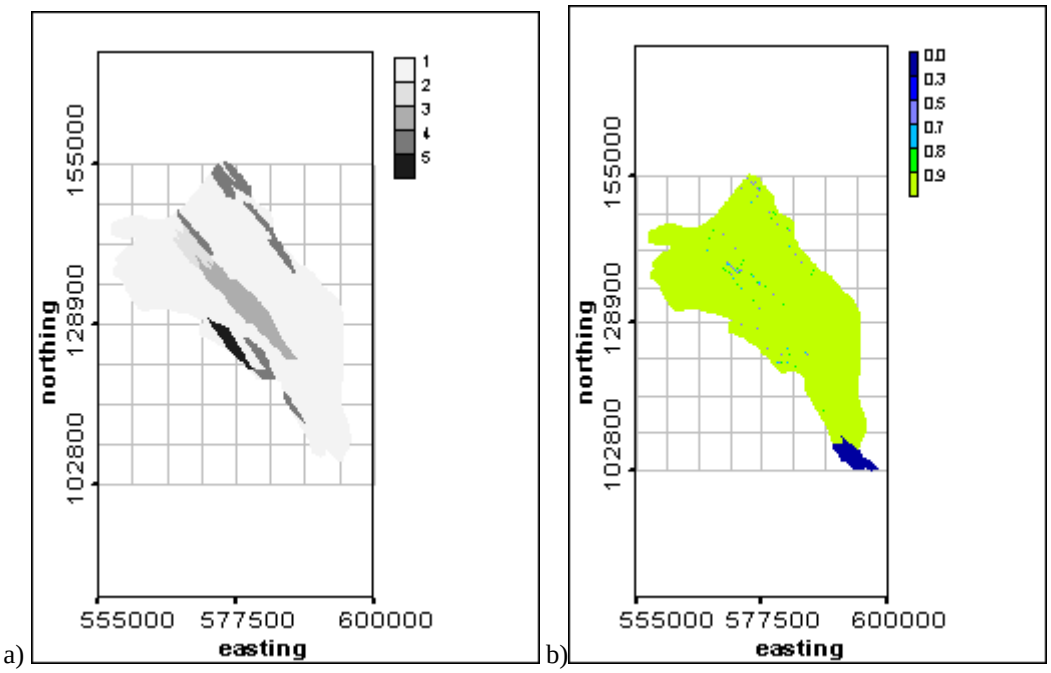


Fig. 22. U1 zonation by PNN trained by leave-20-out cross-validation: classification (a), probability of the class winner (b)

The results obtained using different values of k look rather different and it is not easy to compare them and to select the best. The conclusions on the quality of the result can be based on:

1. results of the accuracy test – but here there is the possibility of overfitting;
2. results of prediction on a subset of data not used during prediction phase; and
3. qualitative and quantitative comparisons of the results on the prediction grid with the interpretation prepared by a geologist using expert knowledge of the geologic structure of the site.

According to the accuracy test, models trained using the k-leave-out cross-validation approach with *k* set equal to 3, 6, and 10 are worse in comparison with the others. The other choices of *k* seem to provide not bad accuracy test result, but that could be the result of overfitting.

Two other comparison approaches are discussed in the following sections.

3.1. Comparison of results on the testing data set

One of the possible approaches to compare methods deals with their application to a validation dataset that was not used in the training procedure. Here, this commonly used approach was modified slightly, because the original data contain too few representatives of 3 of the classes (6 in class2, 15 in class4 and 1 in class5) and only 2 classes (1 and 3) have sufficient number of samples to be partitioned into training and testing data subsets. Thus a traditional validation procedure, where the trained model is applied to data not used in the training procedure, is impossible. But during PNN prediction together with the trained model original data are used as model centers. So it is possible to select a part of data belonging to classes 1 and 3 (where there is enough data for such partitioning) and to check the model performance for the data from those classes. This approach has a disadvantage – testing is performed only for 2 classes, but still it opens a possibility for some intercomparison between different PNN models.

This testing was performed for kernel models presented above (in section 3) by leave-*k*-out cross-validation training procedures. The comparison is performed between a set of models with known parameters. Results of this testing procedure are presented in Table 4. Below PNN models are called according to their number, which depends on the type of model training (*k* from leave-*k*-out cross-validation).

Models 15, 10, 4 and 1 produced the worst results. As the best one can select models 20 and 8. The same models were among those that performed well on the accuracy test, too.

Table 4. Testing of different PNN models on testing data subset

k	σ_X	σ_Y	Orientation (degrees)	Number of misclassifications	Classification error (%)
20	77.265	407.958	41	9	6.9
15	364	131.138	164	15	11.5
10	77.265	1054.34	32	18	13.7
8	114.603	488.64	33	8	6.1
6	116.531	730.382	33	12	9.2
5	77.26	609.952	31	13	9.9
4	111.301	527.423	19	14	10.7
3	260.566	566.119	0	12	9.2
2	112.884	488.967	20	13	9.9
1	129.607	499.166	22	19	14.5

3.2. Comparison with geologic map

The comparison of PNN model results with the a priori geologic interpretation (see Fig. 2) can be made qualitatively – the visual comparison of such parameters as orientation of classes, mutual position of classes, special spots of classes, etc. Based on such visual comparison, the following conclusions can be made:

1. The main orientation of the main class 2 and class 3 area is caught in all results.
2. The orientation of the class 5 area is wrong for the results obtained after training with leave-3-out and leave-15-out cross-validation (see Figs. 15 and 21).
3. After training with leave-15-out and leave-20-out cross-validation the south-west corner rests un-estimated – maximum probability for all classes is lower than 0.3 (see Figs. 21 and 22).
4. The areas of the classes from the results obtained after training with leave-5-out and leave-10-out cross-validation seem to be very narrow and long (the same can be seen from sigma values) – see Figs. 17 and 20. On the contrary, areas of the classes obtained after training with leave-3-out cross-validation are too round (see Fig. 15).
5. According to the visual comparison, the best results are obtained after training with leave-2-out and leave-4-out cross-validation (see Figs. 14 and 16). They have similar kernel parameters and orientation of coordinate axes.

The quantitative comparison of the results with the a priori geologic interpretation can be performed, for example, using a measure like the Hamming distance (misclassification matrix) – the simplest measure of the identity of two patterns. The Hamming distance is considered here as the number of non-equal pixels. Hamming distances between the geologic map and the PNN classification results of U1 zonation are summarized in Tables 5 – general difference and 6 – difference per class (how many pixels of a class in geologic map are assigned to any other class by a model). Zones where PNN classification differs from the geologic map are plotted in dark in Figs. 23 – 27.

According to the Hamming distance, the closest results to the geologic map are the results provided by PNN models 2, 4 and 8; the most differences are found for results from PNN models 1, 10 and 15. Special attention can be drawn to the results from PNN model 20. The southeast corner (where samples are absent) was not classified by this method, so this corner contributes a lot to the Hamming distance (see Figure 24b). However, the Hamming distance between the result of PNN model 20 and the geologic map is not the worst one. After cutting this south corner (see rectangle in Fig. 24b) the Hamming distance decreases to 9.12% (3678 pixels from 40329). So, this model is good, too.

After looking at all comparisons that were made (accuracy test, testing on a subset of data, visual and quantitative comparison with a priori geologic map), two PNN models can be selected as the “best”, models 8 and 20, and thus preferable for usage. Three PNN models can be marked as non-preferable for usage (models 3, 10 and 15). These models have parameters that are much different from the parameters of the other models and from the geologic interpretation: models 3 and 15 present a different orientation of kernel ellipse, while model 10 has a different correlation between σ_X and σ_Y.

So, for further usage a set of PNN models 1, 2, 4, 5, 6, 8 and 20 are proposed.

Table 5. Hamming distance between a priori geologic map and PNN predictions

k	σ_X	σ_Y	Orientation (in degrees)	Number of differences (in pixels)	Measure of difference (in %)
20	77.265	407.958	41	4553	10.98
15	364	131.138	164	5892	14.21
10	77.265	1054.34	32	5793	13.98
8	114.603	488.64	33	4028	9.72
6	116.531	730.382	33	4426	10.68
5	77.26	609.952	31	4900	11.82
4	111.301	527.423	19	4131	9.97
3	260.566	566.119	0	4196	10.12
2	112.884	488.967	20	4057	9.79
1	129.607	499.166	22	6669	16.1

Table 6. Hamming distance between a priori geologic map and PNN predictions per class

k	σ_X	σ_Y	Orientation (in degrees)	Class1	Class2	Class3	Class4	Class5
20	77.265	407.958	41	3094	722	124	553	60
15	364	131.138	164	3870	666	302	748	306
10	77.265	1054.34	32	4014	689	325	660	105
8	114.603	488.64	33	2570	677	113	602	66
6	116.531	730.382	33	3024	655	138	579	30
5	77.26	609.952	31	3301	673	216	652	58
4	111.301	527.423	19	2382	692	224	712	121
3	260.566	566.119	0	2283	667	470	603	173
2	112.884	488.967	20	2329	692	210	707	119
1	129.607	499.166	22	4608	682	297	744	338

a) b)

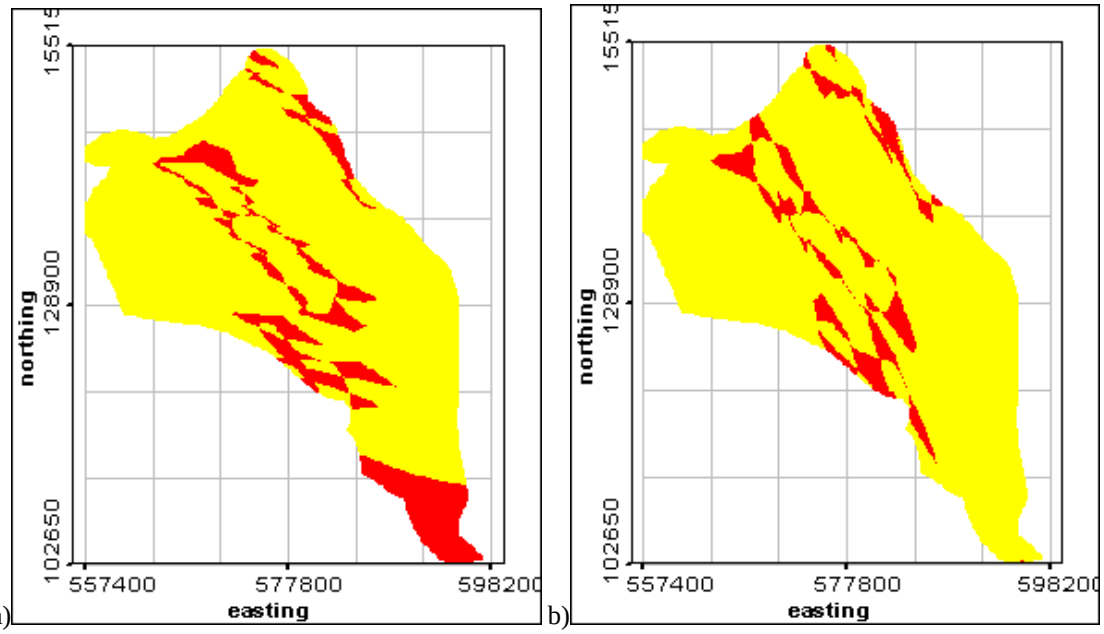


Fig. 23. Zones of difference between a priori geologic map and U1 zonation by PNN models 1 (a) and 2 (b). Areas of difference are dark

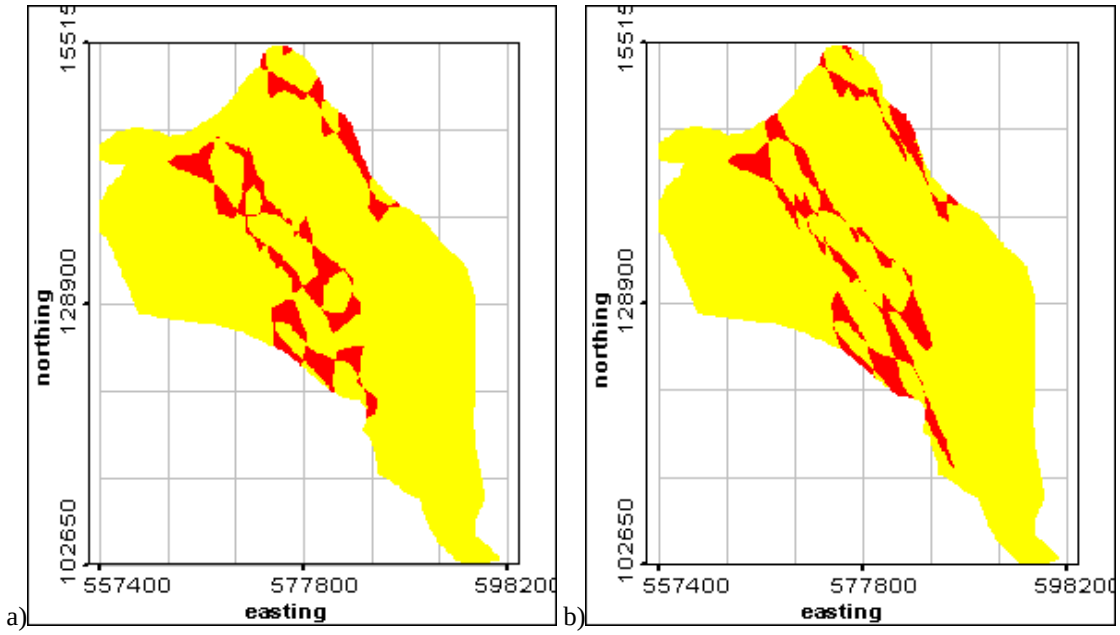


Fig. 24. Zones of difference between a priori geologic map and U1 zonation by PNN models 3 (a) and 4 (b) Areas of difference are dark

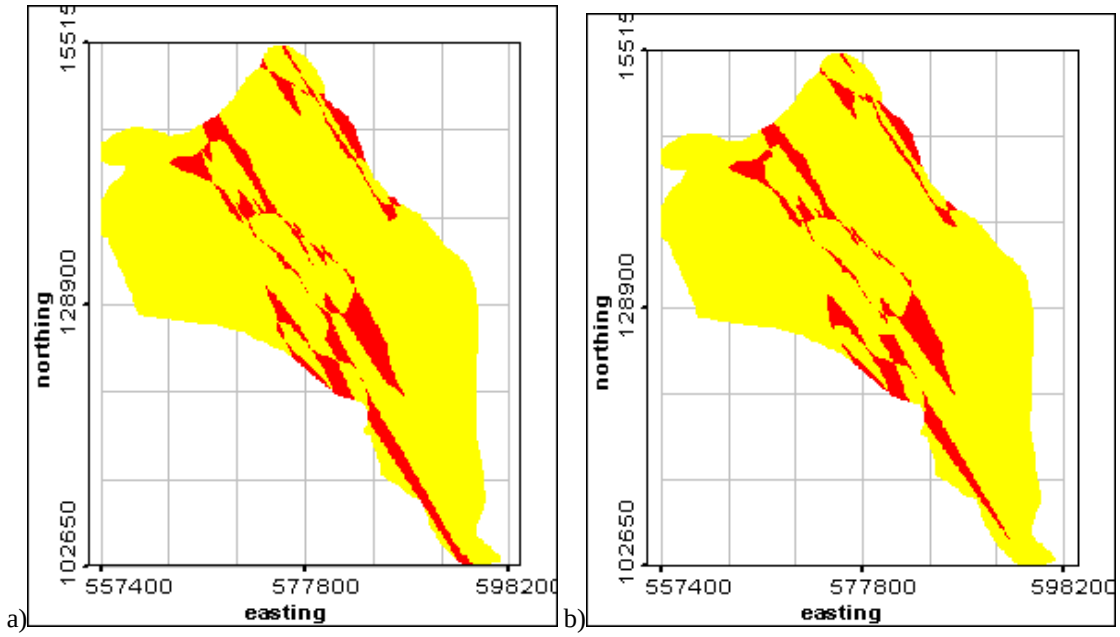


Fig. 25. Zones of difference between a priori geologic map and U1 zonation by PNN models 5 (a) and 6 (b) Areas of difference are dark

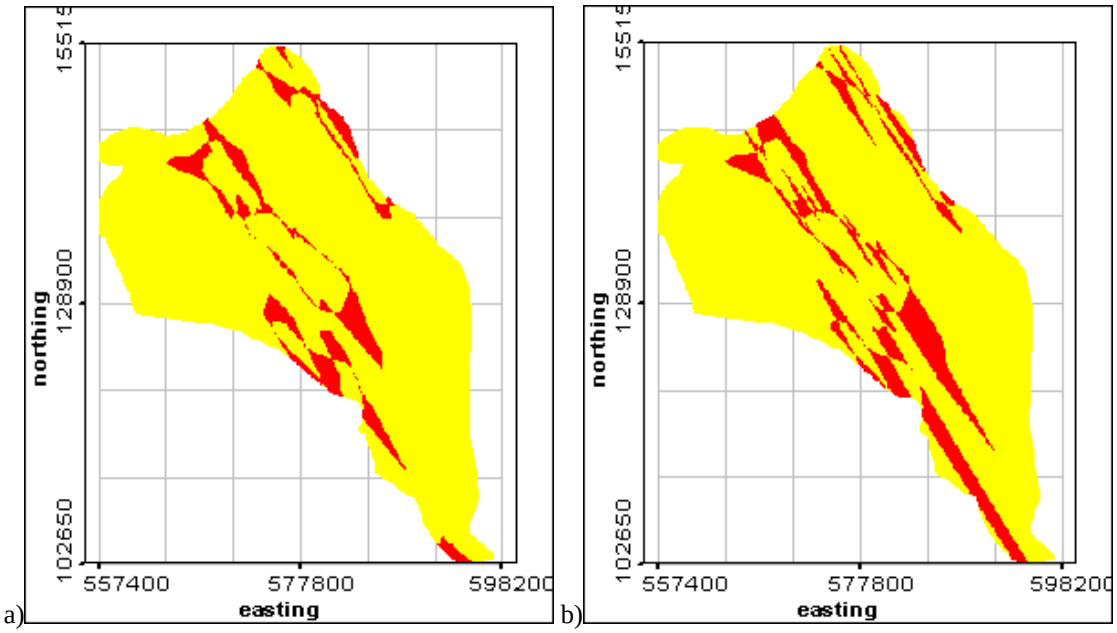


Fig. 26. Zones of difference between a priori geologic map and U1 zonation by PNN models 8 (a) and 10 (b) Areas of difference are dark

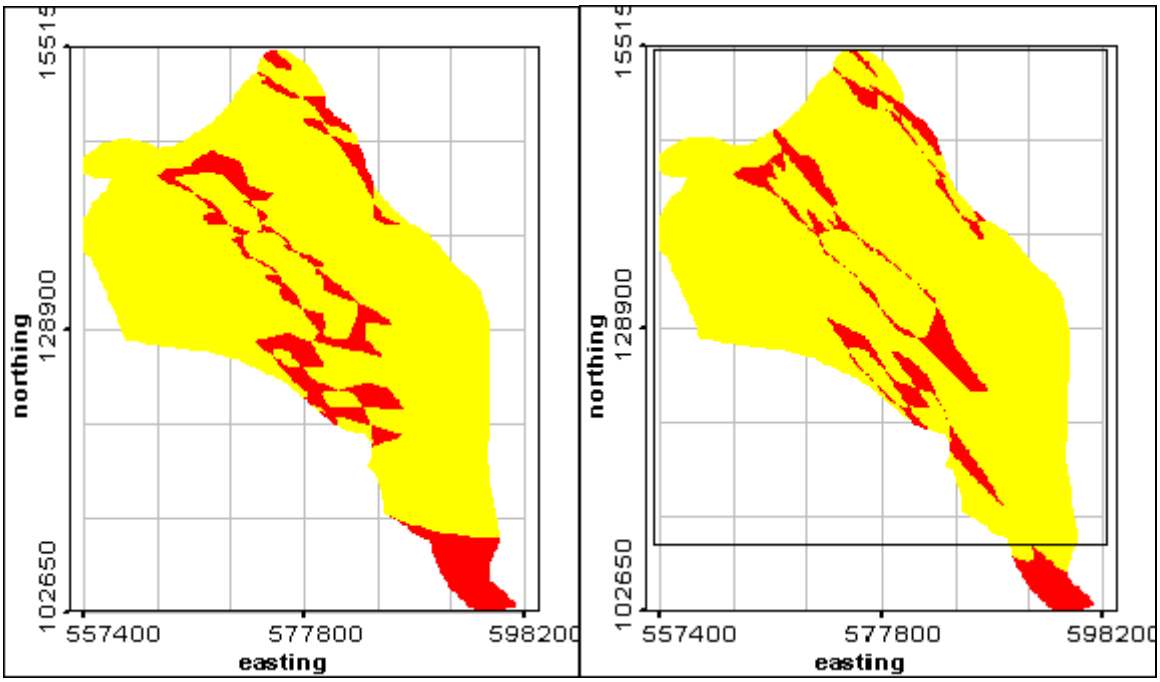


Fig. 27. Zones of difference between a priori geologic map and U1 zonation by PNN models 15 (a) and 20 (b) Areas of difference are dark

3.3. Comparison of spatial correlation structure

Another possible approach is to compare class indicator variograms of the results obtained from PNN with the class indicator variograms of the geologic map (see Figs. 3 – 5). The class indicator variograms can be used to examine the correspondence of the spatial correlation structure: anisotropy, level of class variability, etc.

In the current work, the first attempt to compare results using class indicator variogram roses is presented. For comparison we selected the 4 most characteristic PNN results (different and presenting all the main areas of variation): PNN2 and PNN3 both good according to comparison tests but with different kernel ellipse orientation; PNN10 the one with the strongest difference between kernel bandwidths in perpendicular directions (and as the result very long and thin areas for classes 2 and 3) and PNN15 as one absolutely different from the others. Variograms for PNN estimates are calculated for the same lags and directions as the variograms for expert's geologic map. Class indicator variogram roses for results provided by these models are presented below in Figs. 28 – 32. Based on visual comparison of class indicator variograms the following conclusions can be made:

1. PNN2 provides the best correspondence with geologic map class indicator spatial correlation structure among these four results.
2. PNN3 has a structure that is too isotropic for small lags.
3. PNN10 always provides too strong an anisotropy.
4. PNN15 is not bad according to class indicator spatial correlation structure.
5. The main orientation of the anisotropy seems to be the same as for geologic map. But the values differ.

Quantitatively, for class 2, indicator variogram values of PNN classification results are lower then indicator variogram values of geologic interpretation. For classes 3, 4 and 5 on the contrary, the indicator variogram values of PNN classification results are higher than indicator variogram values of geologic prediction. Such type of comparison of the results is useful, but the problem is that we again deal with some kind of visual (not formal) comparison. To perform more formal class indicator variogram comparison a special RMSE index for indicator class variogram was introduced. This index is calculated by the formula:

$$Index = \sqrt{\frac{1}{N_L N_D} \sum_{directions} \sum_{lags} \left(\gamma_{I\_type} - \gamma_{I\_method} \right)^2}, \quad (1)$$

where N_L is number of lags, N_D is number of directions, γ_{I_type} is value of class indicator variogram for a priori geologic map and γ_{I_method} is value of class indicator variogram for result of a method estimated for the same lag and direction as γ_{I_type}. The summation is made for the whole set of lags and directions, where variogram values are estimated. In fact, this index can be considered as a distance between variograms.

The results of the index calculations for predictions of the different PNN models are summarized in Table 7. The results show good correspondance of PNN indicator variograms with indicator variograms of expert's map (small values of index). The worst coincidence shows the PNN10, which is not good according to all characteristics.

Table 7. RMSE class indicator variogram index for predictions of different PNN models

Method	Class1	Class2	Class3	Class4	Class5
PNN2	0.017	0.011	0.016	0.01	0.006
PNN3	0.017	0.009	0.008	0.01	0.007
PNN10	0.04	0.011	0.032	0.025	0.004
PNN15	0.014	0.011	0.013	0.01	0.005

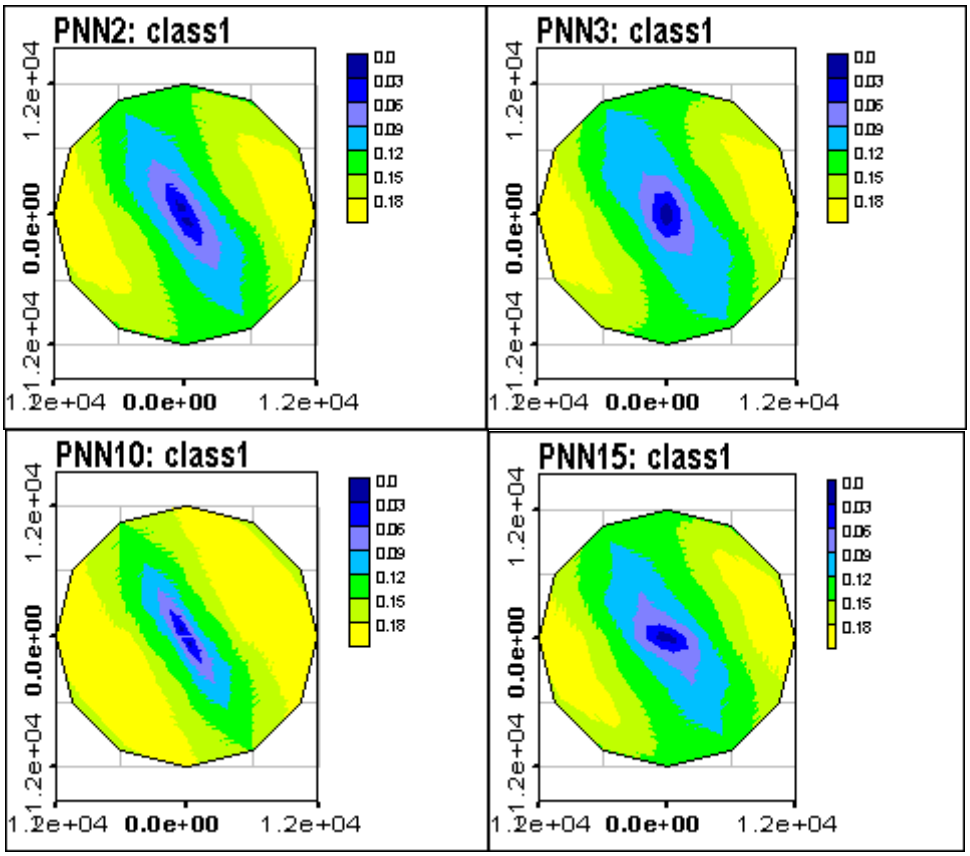


Fig. 28. Class indicator variogram roses (class1) for results of different PNN models

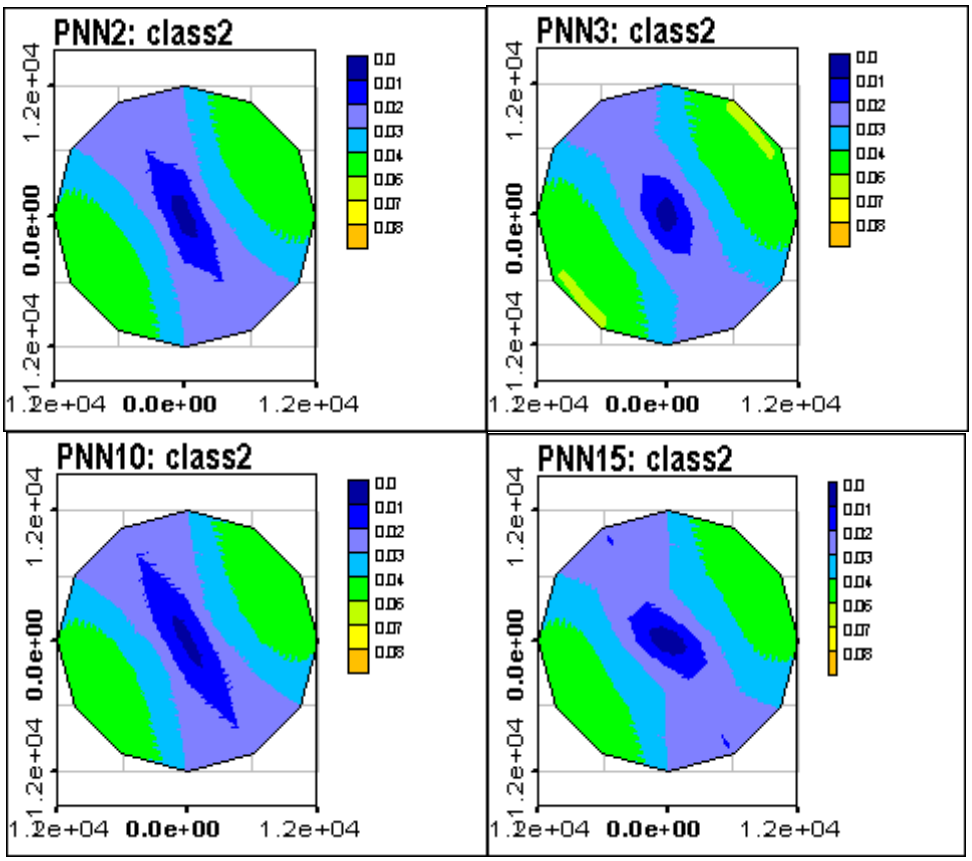


Fig. 29. Class indicator variogram roses (class2) for results of different PNN models

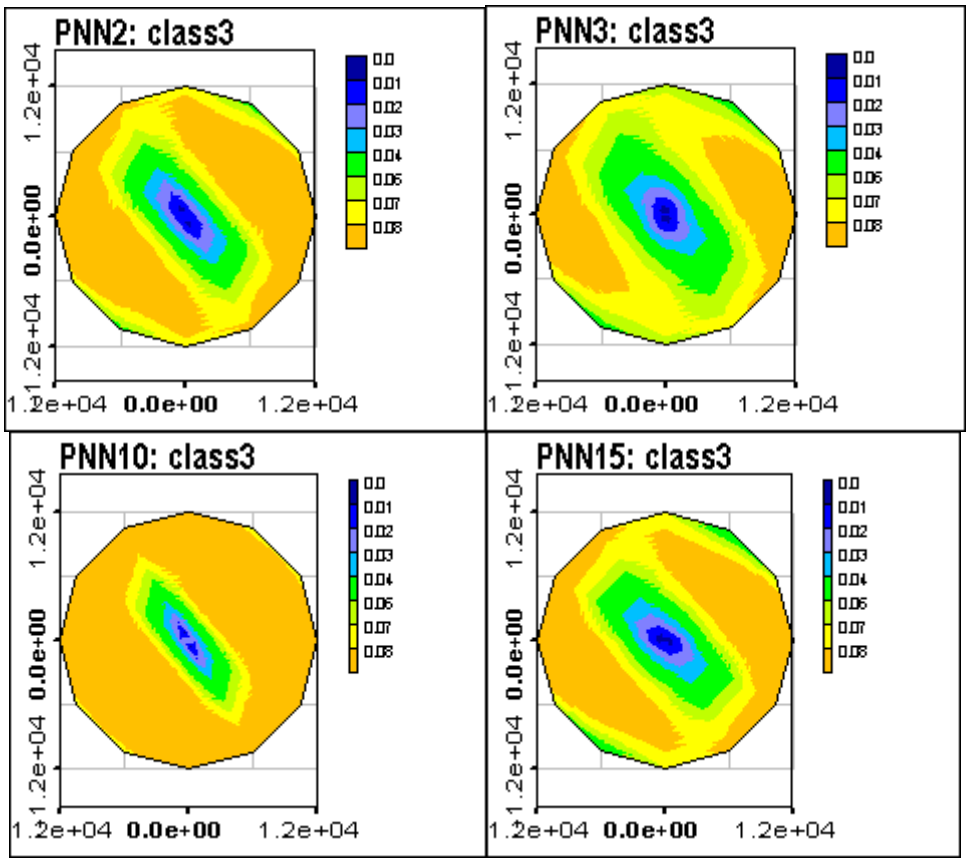


Fig. 30. Class indicator variogram roses (class3) for results of different PNN models

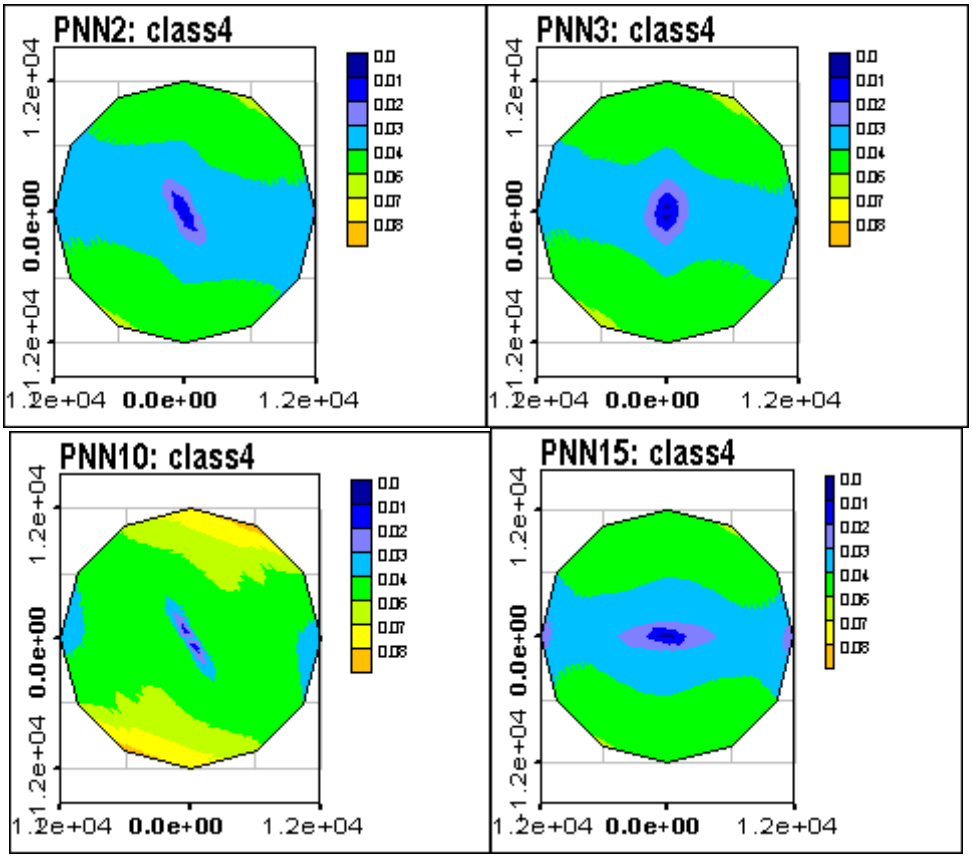


Fig. 31. Class indicator variogram roses (class4) for results of different PNN models

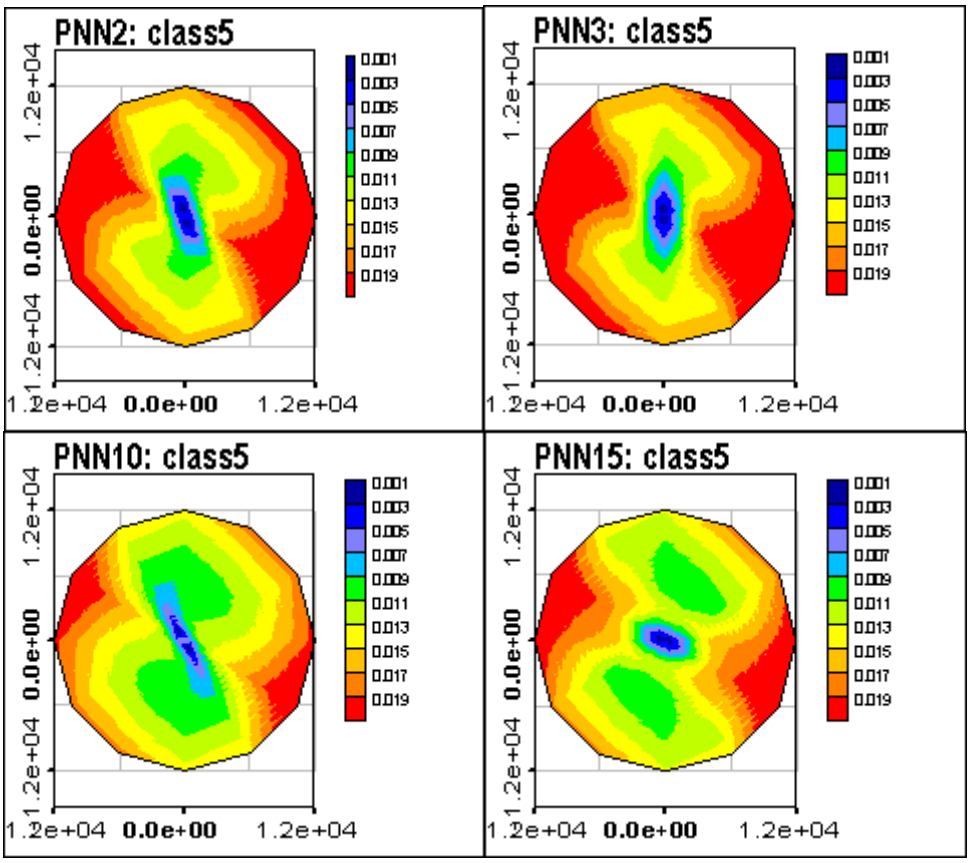


Fig. 32. Class indicator variogram roses (class5) for results of different PNN models

4. Multilayer perceptron for U1 zonation

The performance of an MLP for a multiclass classification task is presented below. The posterior probability of a class in our case is an internal product of MLP – the first part of an output:

$$s_i(g(X,W)) = P_{est}(y = i|X).$$

And inserted in the MLP classification rule, this provides the final output:

$$f(X) = I\left[s_i(g(X,W)) - \theta\right],$$

where $s(t)$ is a usual logistic sigmoid, $g(t)$ is the parametrizing of current MLP output, $\boldsymbol{X}$ is a training data set of size N, W is a set of MLP links' weights, θ is the classification threshold and $I(t_i)$ is a function identifying the number (class) providing the maximum t_i. It was proven [3] that first part of MLP output ($s_i(g(X,W))$) trained to minimize the empirical risk objective function are the probabilities under a number of constraints:the set X is composed of independently and identically distributed samples; training size is statistically large enough; global minimum of empirical risk function is achieved. Proposing such constraints to be satisfied we can treat these outputs as class probabilities. In our case the independence of samples is not satisfied, so we need to treat the output of MLP as a probability very carefully.

For the current classification task, several MLP's with different architectures were used. Most of the MLP's had 1 hidden layer with 3, 5, 7, 8, 9 or 10 hidden neurons. Also, two different MLP's with 2 hidden layers with 5 hidden neurons in each were tried. All these MLP's have 5 outputs, each producing the probability of a class (a value treated as a probability). The final problem is to assign a location to a certain class, so these 5 outputs

were generalized into one output providing the class winner. This solution is made based on the maximum class probability passing the probability threshold.

Training was performed using a weighted cost function with two types of weighting.

1. The first type of weights was obtained based on the number of class members (weight was inversely proportional to the class size). These weights, after normalization to a sum of 1, are presented in Table 8.
2. Another type of weights was based on additional soft information – the sample confidence. One MLP without a weighted cost function was used as well.

Results of accuracy tests for the different MLP's are summarized in Table 9. Results of U1 zonation by different MLP models are presented in Figs. 33 – 41. The results obtained are very different from one another, reflecting the flexibility of the MLP algorithm.

Table 8. Weights of class members for MLP training

Class	Size	Weight
Gravel 1 (1)	165	0.001212
Gravel 2 (2)	6	0.033333
Gravel 3 (3)	43	0.004651
Sand (4)	15	0.013333
Silt (5)	1	0.200000

Table 9. MLP accuracy test results

Number of model	MLP architecture	Type of weights	Number of misclassifications	Error (%)
1	2-3-5	Class size	32	13.91
2	2-5-5	Class size	26	11.3
3	2-7-5	Class size	16	6.96
4	2-8-5	Class size	25	10.87
5	2-9-5	Class size	19	8.26
6	2-10-5	Class size	23	10
7	2-5-5-5	Class size	3	1.3
8	2-5-5-5	NONE	1	0.43
9	2-9-5	Confidence	5	2.17
10	2-10-5	Confidence	8	3.48

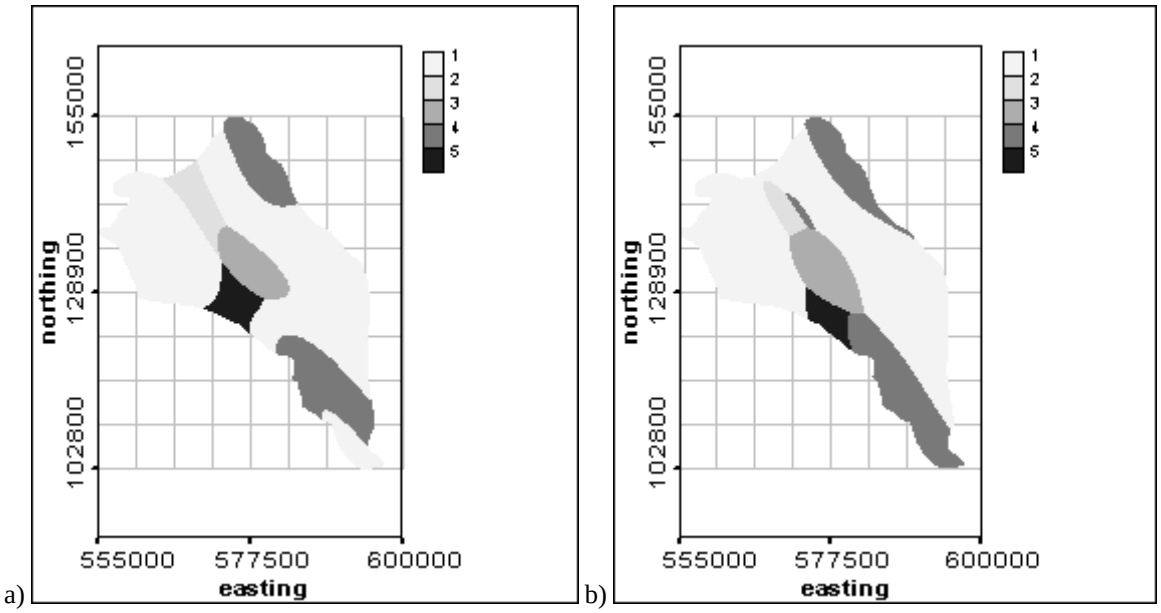


Fig. 33. U1 zonation by MLP models 1 (a) and 2 (b)

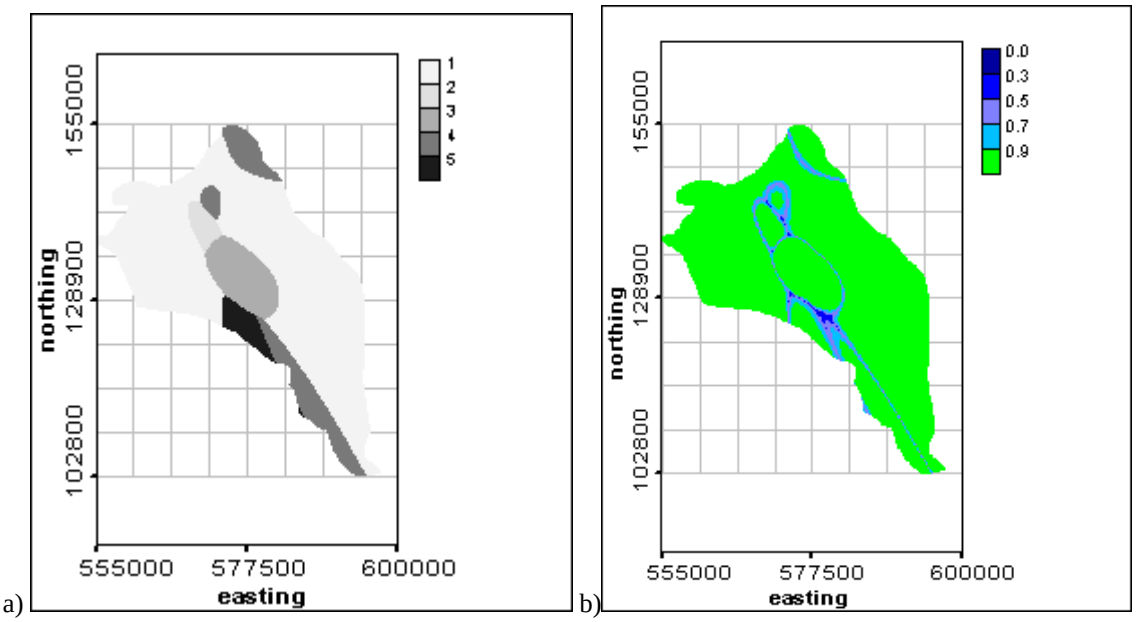


Fig. 34. U1 zonation by MLP model 3: classes (a) and the class winner probability (b)

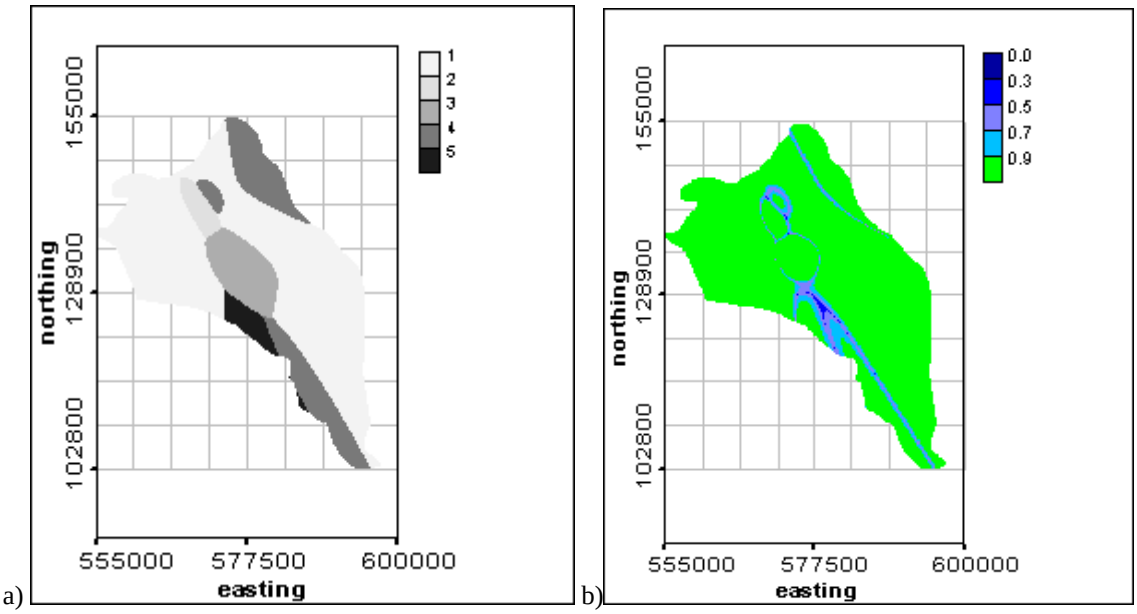


Fig. 35. U1 zonation by MLP model 4: classes (a) and the class winner probability (b)

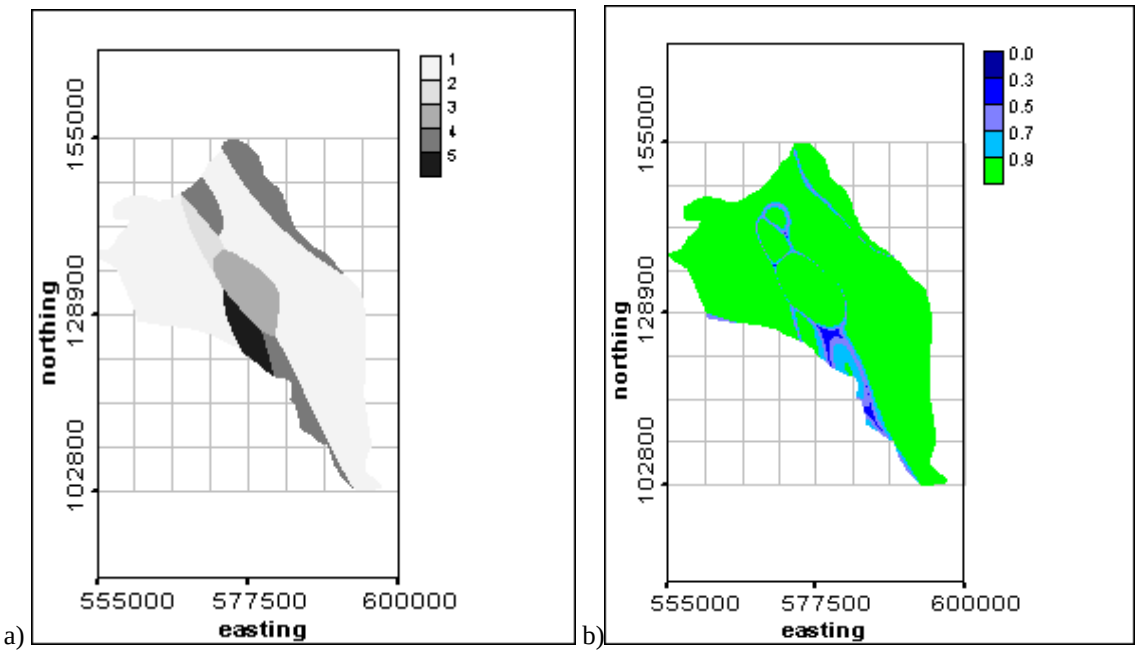


Fig. 36. U1 zonation by MLP model 5: classes (a) and the class winner probability (b)

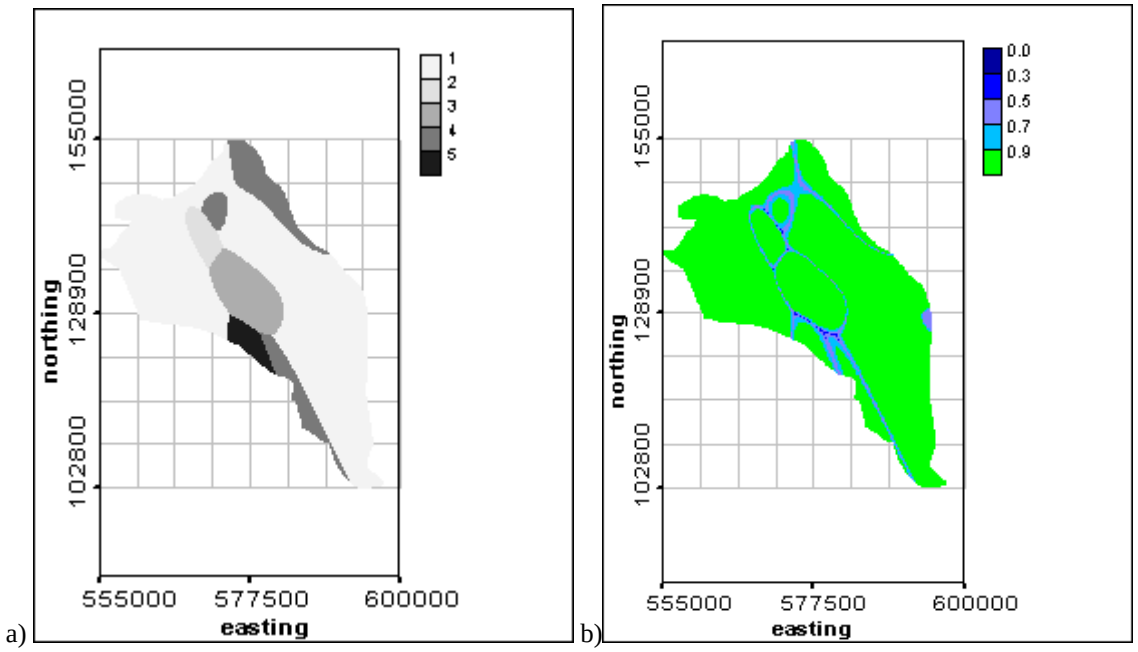


Fig. 37. U1 zonation by MLP model 6: classes (a) and the class winner probability (b)

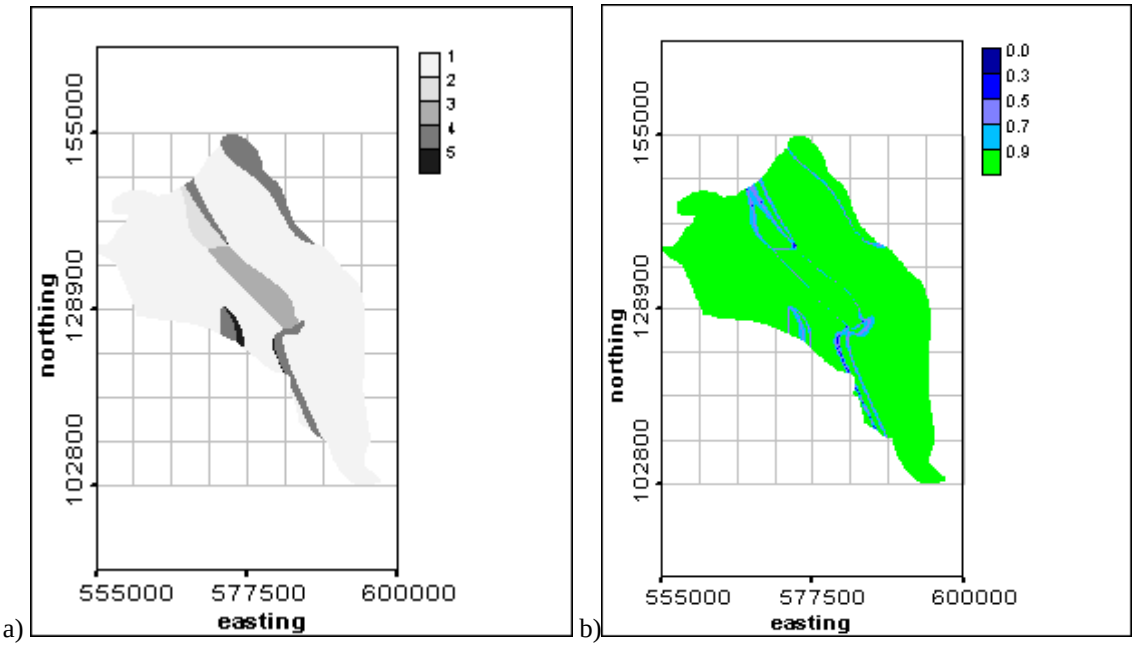


Fig. 38. U1 zonation by MLP model 7: classes (a) and the class winner probability (b)

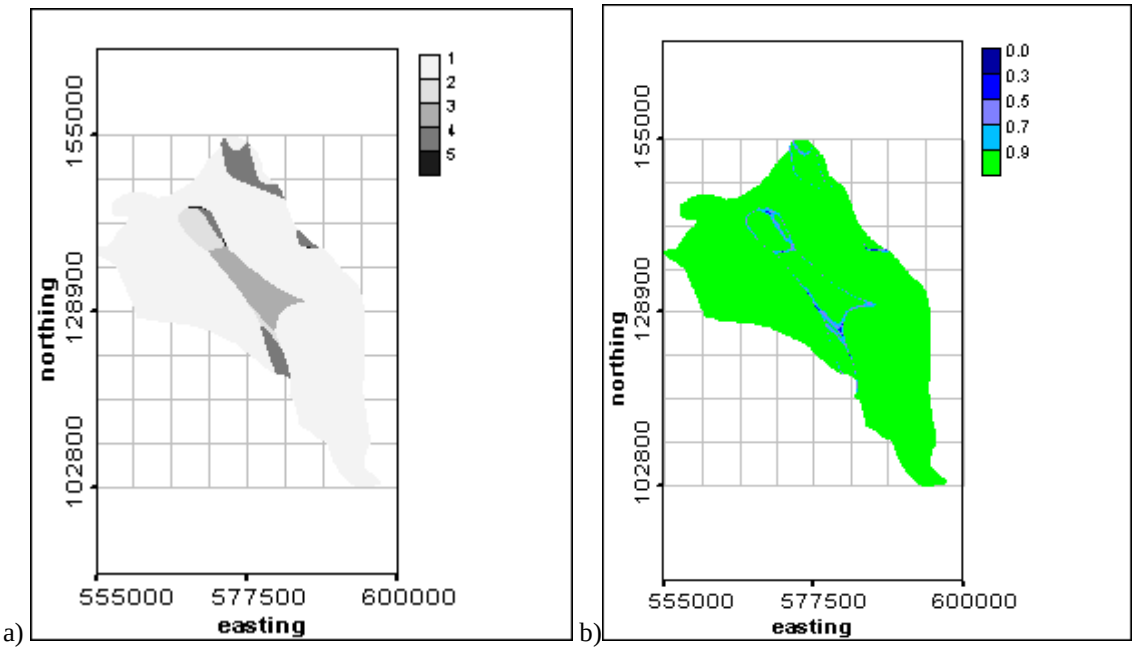


Fig. 39. U1 zonation by MLP model 8: classes (a) and the class winner probability (b)

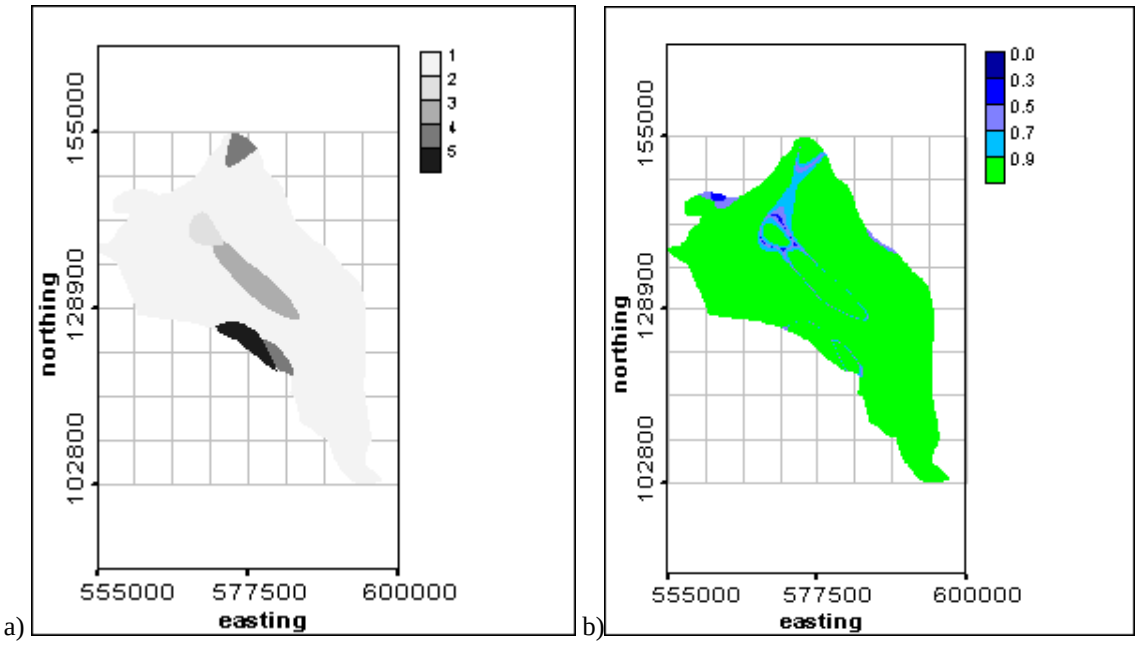


Fig. 40. U1 zonation by MLP model 9: classes (a) and the class winner probability (b)

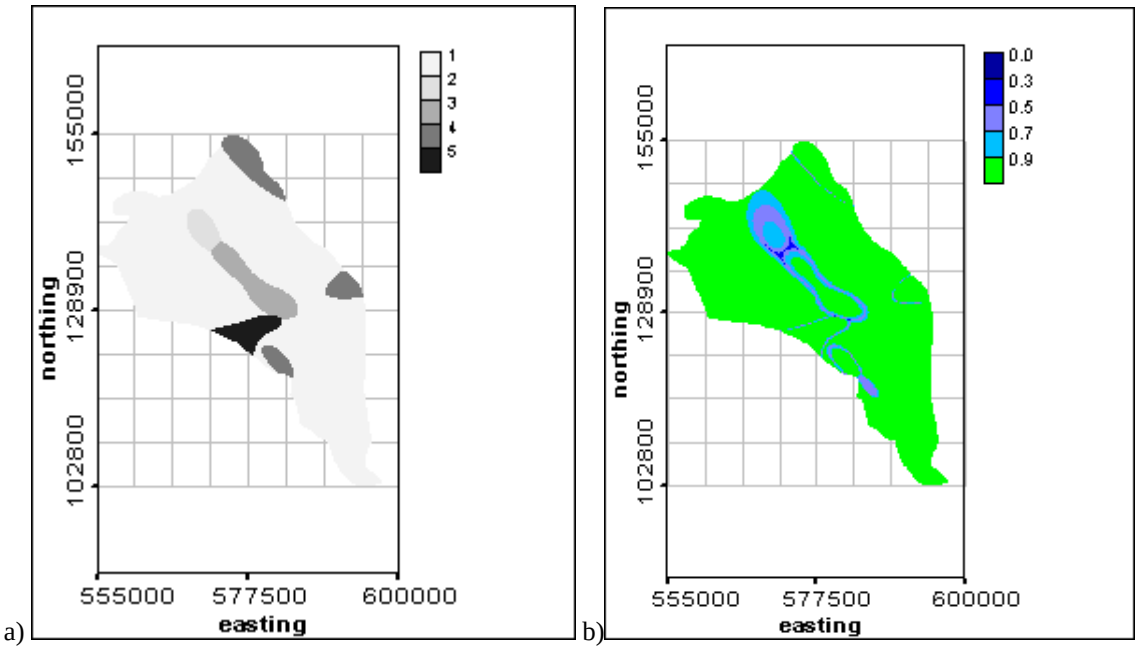


Fig. 41. U1 zonation by MLP model 10: classes (a) and the class winner probability (b)

The comparison of MLP models presented above can be made based on 2 aspects – accuracy test error and similarity to the a priori geologic classification map. Zones of difference between MLP classification results and the geologic map are presented in Figs 42 – 46.

According to the accuracy test results for MLP with 1 hidden layer, using class size as weights for cost functions during training are worse than using the sample confidence data, and among them the worst are the results of models 1, 2 and 4. They also visually differ stronger from the geological classification – all patches of the U1 facies are too thick and round (see Figs. 33 – 37 and 43, 45, 46).

Using sample confidence data (models 9 and 10) improves the results of accuracy test and the resulting figures look more like the a priori map, and they have no penetration of class 4 among classes 2 and 3 (see Figs. 40, 41 and also 42 and 46). However, they do not reproduce the long zone of class 4 along the border in the north, which can be seen in the results of MLP with 2 hidden layers – models 7 and 8 (see Figs. 38, 39 and 44).

The quantitative aggregated measure of zero dimension (the number) of the difference between MLP classification results and geologic map (Hamming distance) is presented in Tables 10 and 11, the general difference and number of pixels of a class in geologic map assigned to any other class by a model correspondingly. According to the Hamming distance, the models that are closer to the geological information are MLP models 7, 8, 9 and 10. The results that are most distant from the geologic map are the results of MLP models 1, 2 and 4. Here we see high correspondence between the results of the accuracy test and the similarity with the a priori geologic map.

Table 10. Hamming distance between a priori geologic map and MLP predictions

Model number	MLP architecture	Weighting	Number of differences (in pixels)	Measure of difference (in %)
10	2-10-5	confidence	3872	9.34
6	2-10-5	class size	6660	16.07
1	2-3-5	class size	10058	24.26
2	2-5-5	class size	10461	25.24
7	2-5-5-5	class size	3989	9.62
8	2-5-5-5	none	3702	8.93
3	2-7-5	class size	7452	17.98
4	2-8-5	class size	9175	22.13
9	2-9-5	confidence	3321	8.01
5	2-9-5	class size	6862	16.55

Table 11. Hamming distance between a priori geologic map and MLP predictions per class

Model number	MLP architecture	Weighting	Class1	Class2	Class3	Class4	Class5
10	2-10-5	confidence	2503	683	197	361	128
6	2-10-5	class size	5842	564	56	198	0
1	2-3-5	class size	9143	258	58	408	191
2	2-5-5	class size	9508	572	279	94	8
7	2-5-5-5	class size	2514	652	148	374	301
8	2-5-5-5	none	1817	565	67	837	416
3	2-7-5	class size	6043	770	87	552	0
4	2-8-5	class size	8252	480	105	338	0
9	2-9-5	confidence	1182	859	211	1054	15
5	2-9-5	class size	5977	616	86	165	18

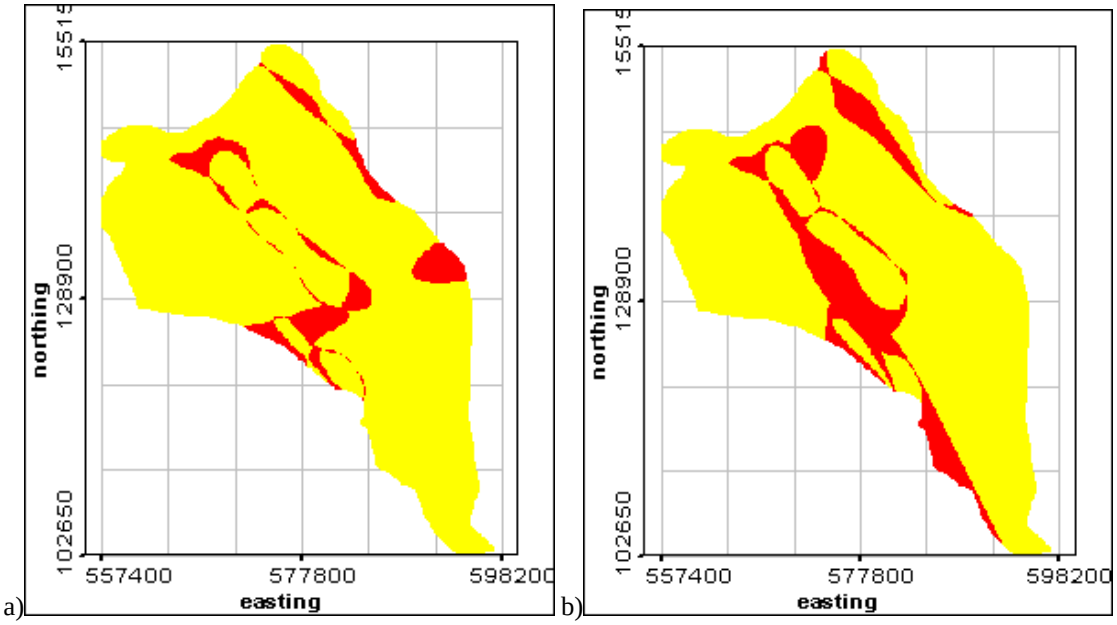


Fig. 42. Zones of difference between a priori geologic map and U1 zonation by MLP 2-10-5 trained with weights according to confidence (a) and class size (b). Areas of difference are dark

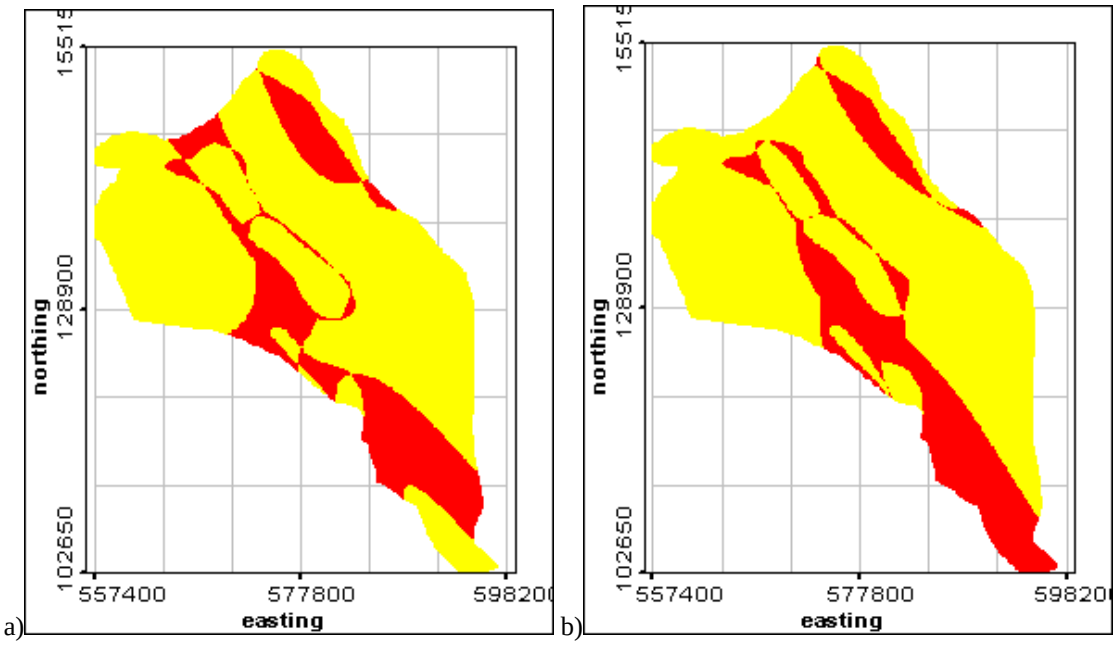


Fig. 43. Zones of difference between a priori geologic map and MLP 2-3-5 (a) and MLP 2-5-5 (b) trained with weights according to class size. Areas of difference are dark

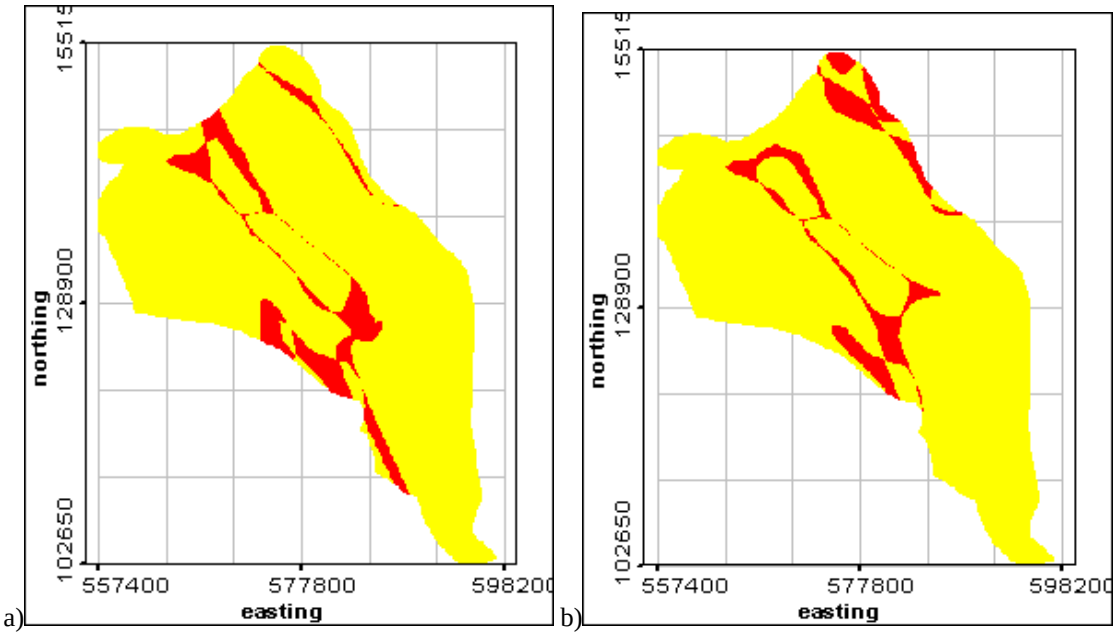


Fig. 44. Zones of difference between a priori geologic map and MLP 2-5-5-1 trained with weights corresponding to class size (a) and without weights (b). Areas of difference are dark

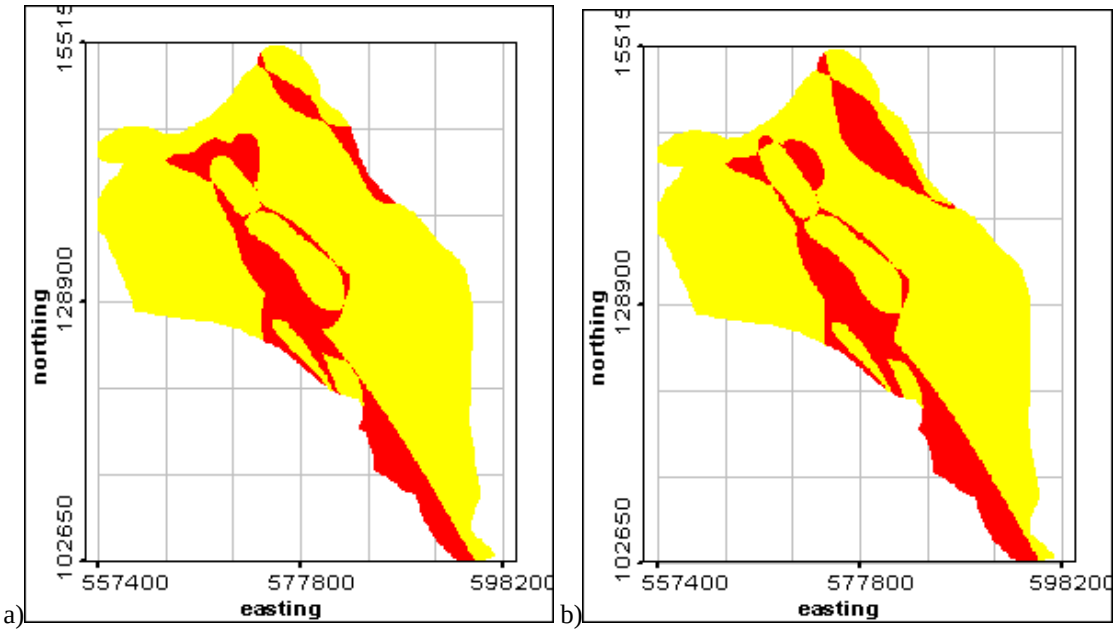


Fig. 45. Zones of difference between a priori geologic map and MLP 2-7-5 (a) and MLP 2-8-5 (b) trained with weights corresponding to class size. Areas of difference are dark

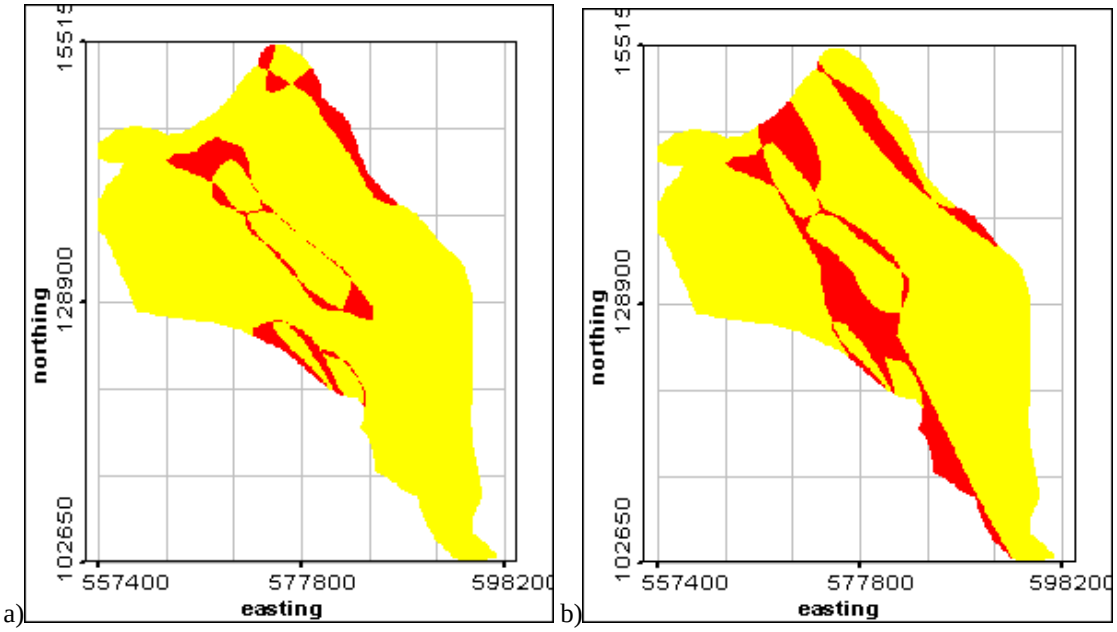


Fig. 46. Zones of difference between a priori geologic map and MLP 2-9-5 trained with weights according to confidence (a) and class size (b). Areas of difference are dark

Four characteristic MLP models were selected for comparison of the class indicator variograms: MLP1 (2-8-5) – rather low MLP, not too good according to all tests, but the best among poor models and MLP2 (2-5-5-5), MLP3 (2-5-5-5) and MLP4 (2-9-5) – all showing good results according to previous tests, but trained using different goal functions (weighted according to class size, not weighted and weighted according to sample confidence, respectively). The class indicator variogram reproduction for different MLP models are rather peculiar – each MLP model looks as if it is correctly oriented for only one class – for example MLP1 looks very good for class2, MLP2 looks good for class1 and not bad for class5, MLP3 looks very good for class4 and MLP4 looks good for class3.

RMSE index values for the class indicator variograms are summarized in Table 12.

Table 12. RMSE index for class indicator variograms of MLP results

Method	Class1	Class2	Class3	Class4	Class5
MLP1	0.12	0.007	0.04	0.062	0.019
MLP2	0.022	0.012	0.011	0.03	0.0028
MLP3	0.009	0.0065	0.011	0.009	0.008
MLP4	0.018	0.018	0.006	0.013	0.011

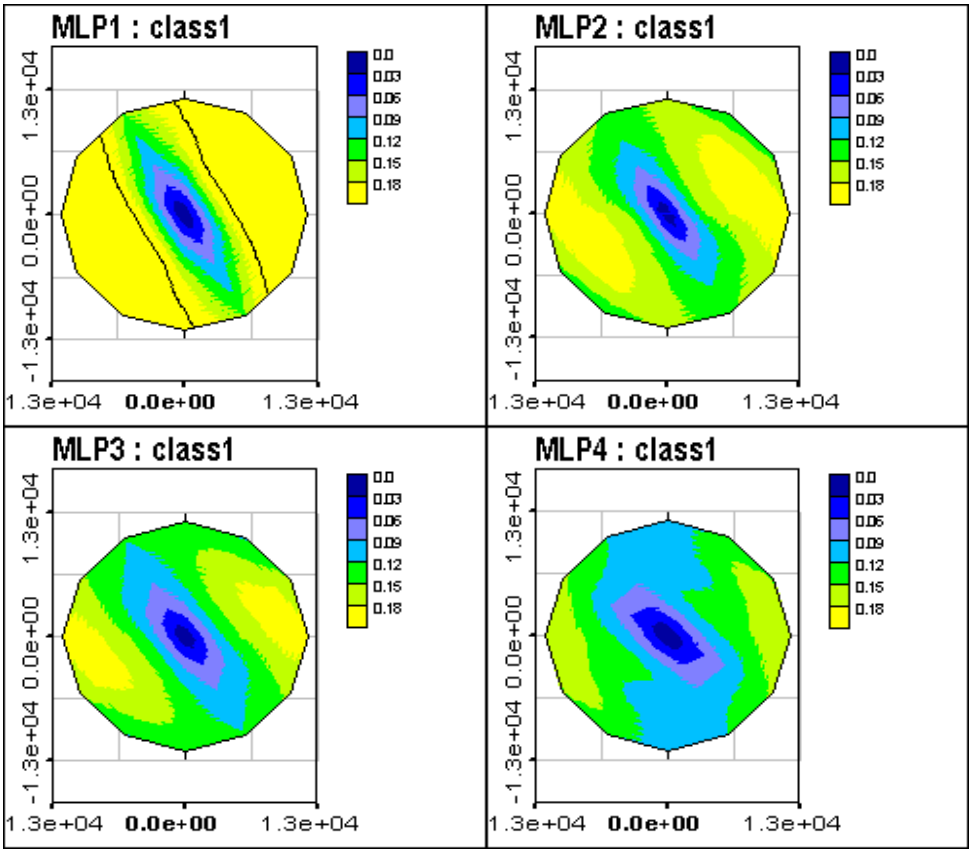


Fig. 47. Class indicator variogram roses (class1) for results of several MLP models

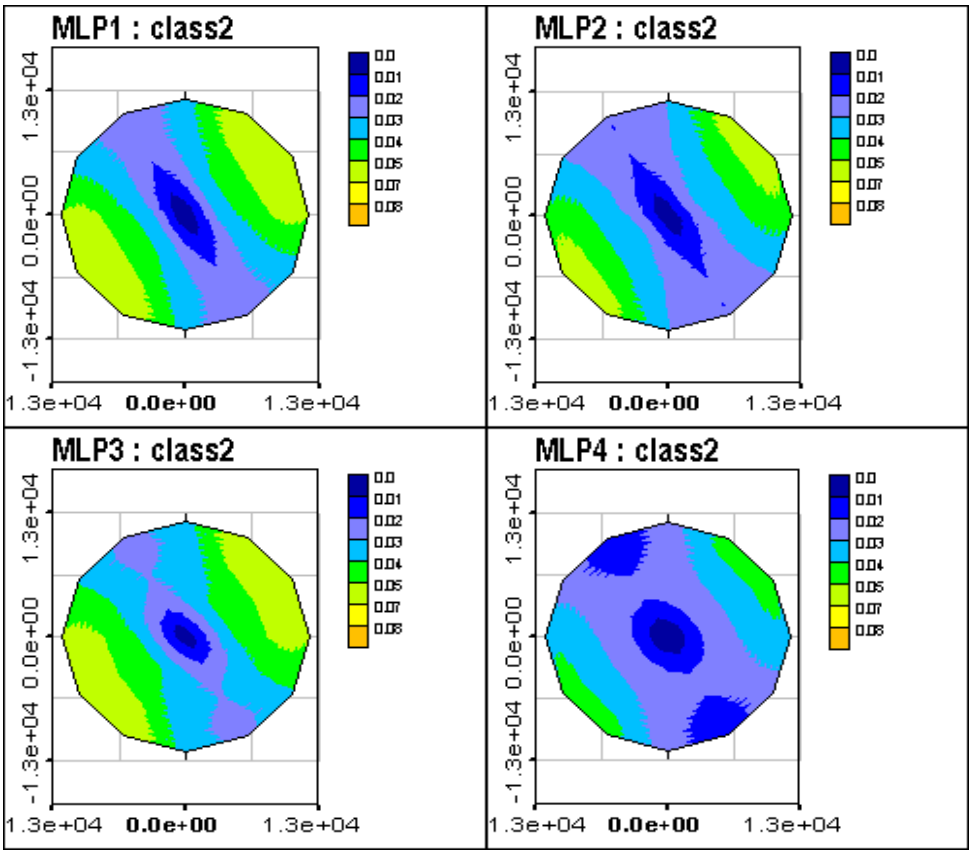


Fig. 48. Class indicator variogram roses (class2) for results of several MLP models

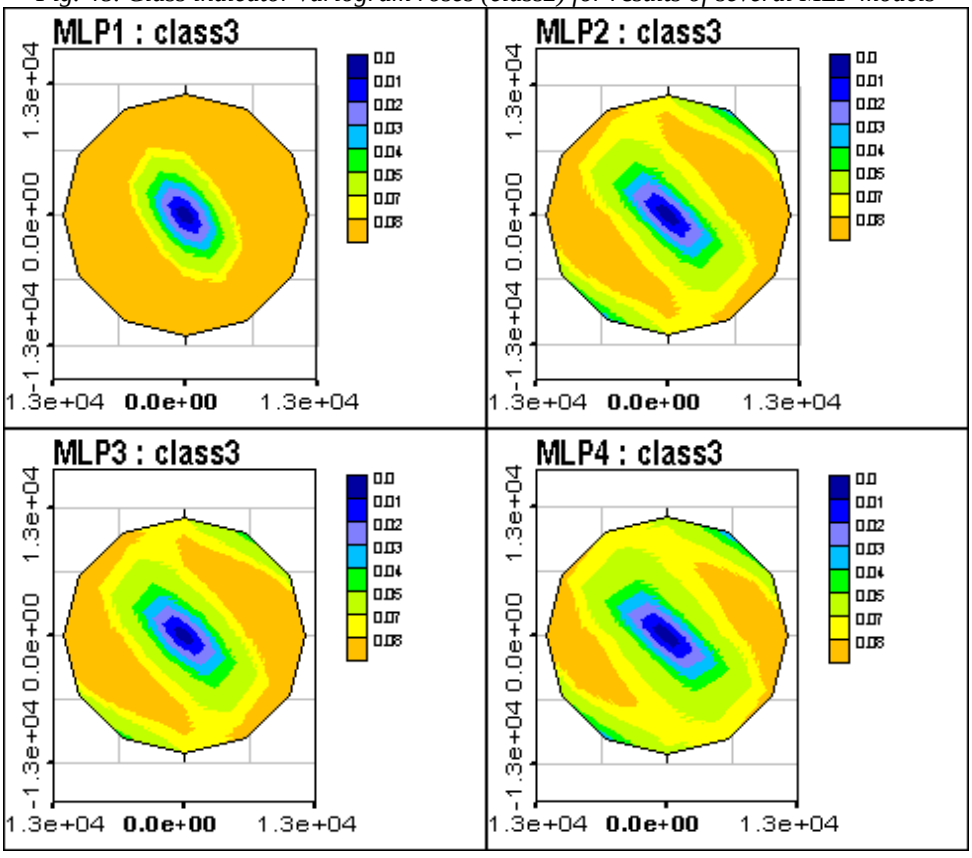


Fig. 49. Class indicator variogram roses (class3) for results of several MLP models

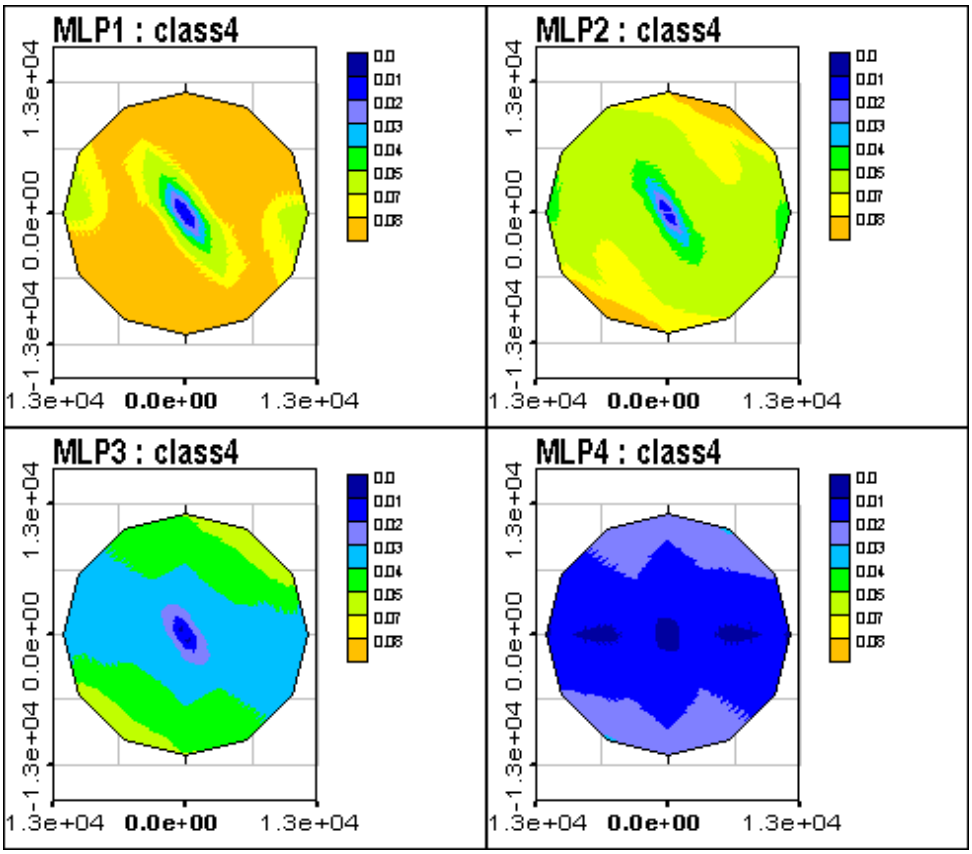


Fig. 50. Class indicator variogram roses (class4) for results of several MLP Models

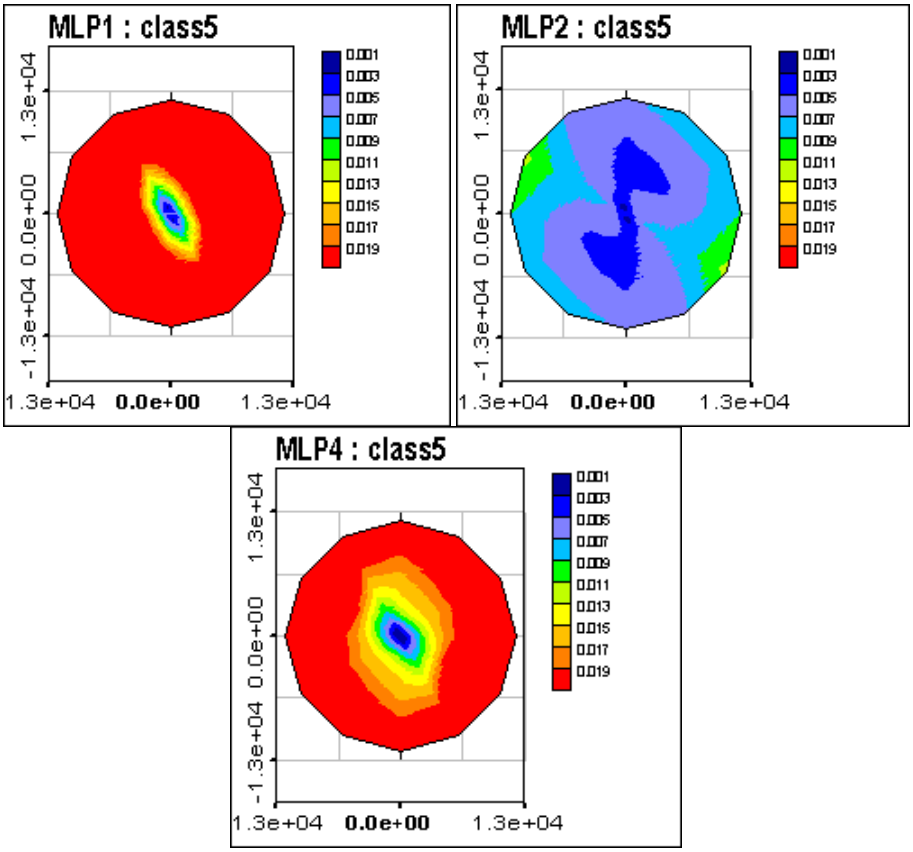


Fig. 51. Class indicator variogram roses (class5) for results of several MLP models

5. Support vector machine for U1 zonation

The U1 zonation task is a 5-class classification problem. The original SVM was developed for binary classification problems. We examined several approaches to generalize the Support Vector Machines approach for multi-class problems.

As SVMs are originally constructed as binary classifiers, most of the multi-class extensions are based on some scheme that combines several binary classifiers into one multi-class classifier. By using this approach, SVMs lose some of their theoretical foundations; nevertheless such schemes are both flexible and easy-to-use. The most widely used are:

- "One-against-rest": to obtain M-class classifier we train M binary classifiers to classify each class against all other data, and then combine.
- "One-against-one": to obtain M-class classifier we train M(M-1)/2 binary classifiers, one for each pair of classes, and then combine.

The usual way of combining the binary classifiers is the following. Let the output of every *m* classifier be

$$y^{(m)} = sign\ (\sum_i \alpha_i^{(m)} y_i K^{(m)}(x, x_i) + b^{(m)})$$

Then for "one-against-rest" scheme we can take the classifier that has maximal decision function:

$$y = \arg \max_m \sum_i \alpha_i^{(m)} y_i K^{(m)}(x, x_i) + b^{(m)}$$

and for "one-against-one" scheme we have to use some voting procedure to choose a classifier from the pair-wise competition.

5.1. Classical SVM

We start from the "pure" SVM model, without taking into account the data confidence or expert's map as auxiliary data. Three approaches to the M-class problem will be presented:

generalization of the binary classification prolem to the multiclass classification by

- "one-against-rest" model,
- "one-against-one" model,

and the generalization of the binary classification to the current specific multi-class problem.

5.1.1. One-against-rest.

As the number of points in classes 2, 4, and 5 is not sufficient to obtain the representative training and testing subsets, the parameters for these classes can be taken from cross-validation error analysis (that is also quite complicated) or heuristically. The heuristics are: the parameters should correspond (be similar) to the parameters for the other classes (classes1 and 3).

Parameters for classes 1 and 3 were selected based on the standard procedure, with several test subsets selected.

The class 1 testing error surface demonstrates typical behavior for the SVM model when optimal parameters are "proportional" to each other. The class 3 error surface is different: there is a large region of low (almost zero) testing error. This is because of the class 3 spatial distribution and the absence of regions of overlap with other classes. However these surfaces can give just a rough estimate of the parameters, since the final result mostly depends on the combination of binary classifiers.

For this "one-against-rest" method it is evident that parameter C has very significant influence: it defines the limits for the coefficients/weights, hence it influences the amplitude of the decision functions being compared. Generally, the outputs of the classifiers should be comparable and the parameters should not be very different.

The parameters suitable for the "one-against-rest" method are gathered in Table 13. The final U1 zonation by an SVM trained by the one-against-rest approach using those parameters is presented in Fig. 52.

Table 13. Parameters of SVM "one-against-rest"

	Class 1	Class 2	Class 3	Class 4	Class 5
RBF kernel sigma	5920	5920	5550	5920	5920
C parameter	100	300	300	100	300

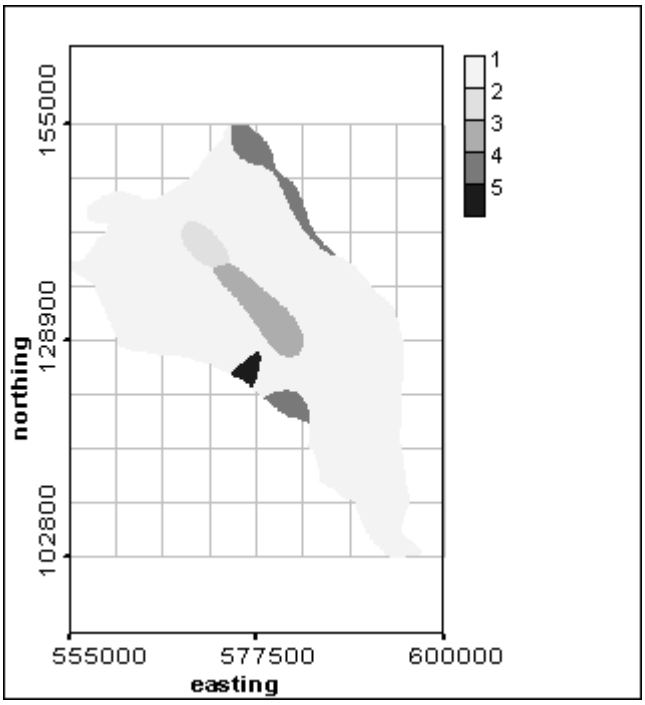


Fig. 52. U1 zonation by SVM one-against-rest classification

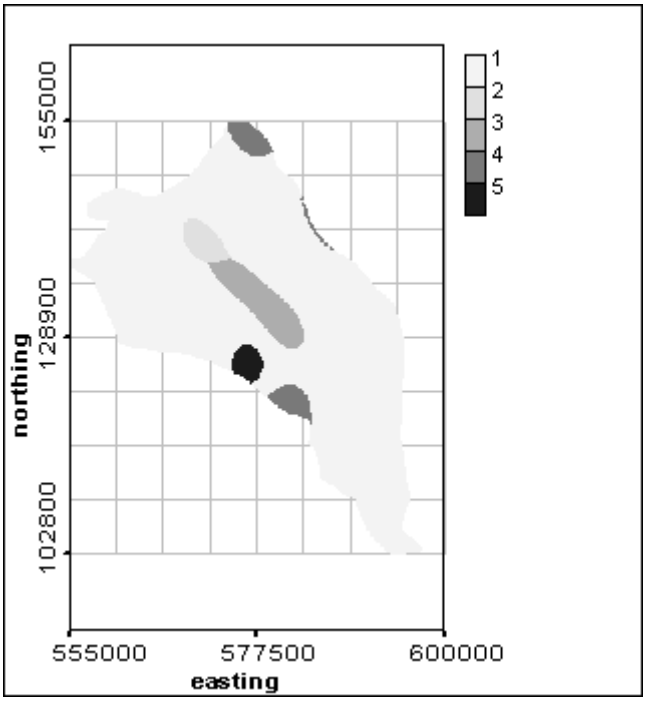


Fig. 53. U1 zonation by SVM one-against-one classification

5.1.2. One-against-one.

"One-against-one" scheme is the most flexible and the hardest to tune among the schemes being presented. However, it was noticed that it gives a very "ordinary" result, similar to the other methods. The same observation was made in some other multi-class classification tasks.

The lack of data for classes 2, 4, and 5 is also a problem for the "one-against-one" generalization scheme. Nevertheless, it is possible to obtain reasonable parameters for each pair of classes. Some pairs of classes allow building test error surfaces, the standard approach, while some classes must be tuned by heuristics. As the RBF kernel parameter (bandwidth) corresponds to some average range of influence of the class members, its values

should be: 1) comparable with the size of the region "occupied" by the corresponding class, and 2) comparable with the kernel bandwidth of the neighbor class. Another useful heuristic is the expected shape of the decision boundary: the larger the RBF kernel bandwidth, the smoother the boundary. Some classes (for example, class 5 and class 2) probably don't have a common boundary at all and can be tuned quite roughly.

We will not stress the parameter tuning procedure, and mainly present the final classification result – Fig. 53.

5.1.3. "Pseudo-binary" approach.

As previously mentioned, the disadvantage of the generalization schemes described above is that SVM loses some of its nice properties, including some of its generalization ability. Binary SVM builds a separating hyper-plane with maximal margin in a feature space, giving rise to a good generalization. By combining several SVMs we lose this property. Although there are some theoretical extensions to "multi-class margins", their applications to real tasks are still too complicated to be well understood.

An easy way to overcome these difficulties is to use an approach based on the evident structure of the data. In the case study of interest, it is possible to solve the multi-class task by mainly dealing with one binary problem: by classifying class 1 against the rest data. Then there are two binary problems left: to classify class 4 against class 5, and class 2 against class 3. One point belonging to class 5 (near class 2 boundary) can be considered as an outlier and omitted. The rest of the classification boundaries are already obtained. The main advantage of this scheme is that it is mostly binary, and hence keeps the good (theoretically founded) generalization abilities of SVM.

The main step, class1-against-rest classification is already performed during "one-against-rest" study. The other necessary steps, 4-aganst-5, and 2-against-3, were performed during the "one-against-one" study. The figure 54 demonstrates prediction classification mapping for these preliminary tasks.

Because the classification boundaries for the 2-against-3 task depend significantly on the kernel in the "extrapolation" region, they are almost identical in the region of interest. Classification 4-against-5 is almost a degenerate task since class 5 consists of only 1 data point. This particular classification is not presented on the figures. The combination of the binary classifiers in Fig. 54 resulted in the final classification presented on Fig. 55.

5.1.4. Preliminary discussions

Three multi-class SVM methods that can be used for the given task were considered: general and wide-used "one-against-rest", "one-against-one" and specific for this data "pseudo-binary". The advantages and disadvantages of all the methods were mentioned before. Concerning the prediction classification we can conclude that there is no significant difference in the result. The only significant difference between the results of the 3 methods is in class 5 classification, since it consists of only one point.

5.2. Incorporation of the confidence data

Some confidence values were provided for every measurement by Hanford Site geologists (see Fig. 1b). To incorporate this knowledge into the SVM model, the C parameter is taken individually for every point; the measurement with confidence value k has C/k as a parameter. Since the value of this parameter is an upper bound of the weight that the corresponding point has, the more confidence we have in the measurement, the more influence it will have on the solution. The C value itself can be tuned by standard methods.

Since we significantly change the model by incorporating the confidence data, all the parameters have to be re-tuned. The results are two-fold: confidence information incorporation leads to a more stable region of minimal testing error, but the maximal testing error increased significantly. Here we consider only the results of the "one-against-all" SVM classification. The "pseudo-binary" classification gave almost the same result.

The results are presented in the Figure 56. As a rule, points that were misclassified tend to have a low confidence level. Incorporation of the confidence data resulted in the existence of a larger region occupied by class 5.

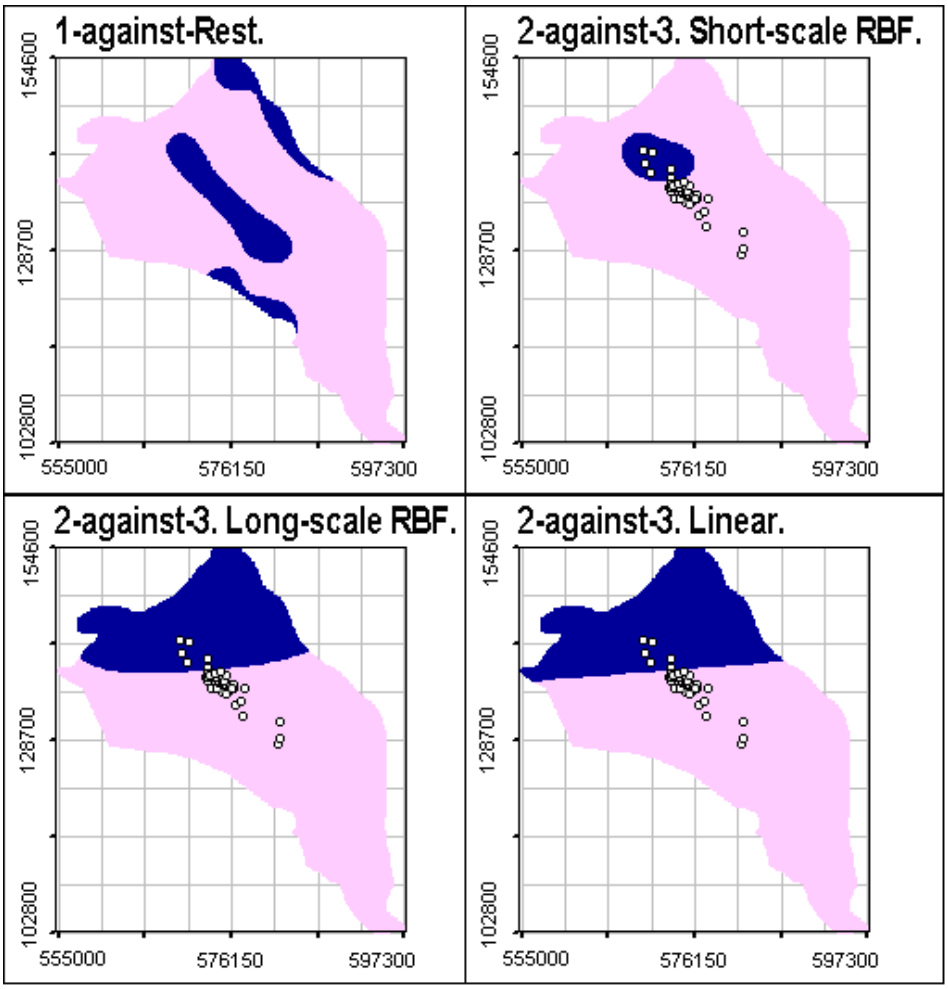


Fig. 54. Sub-tasks for "pseudo-binary" approach. Top-left: class 1(light)-against-rest(dark). Other: different classification of 2(dark)-against-3(light). Small empty circles correspond to data

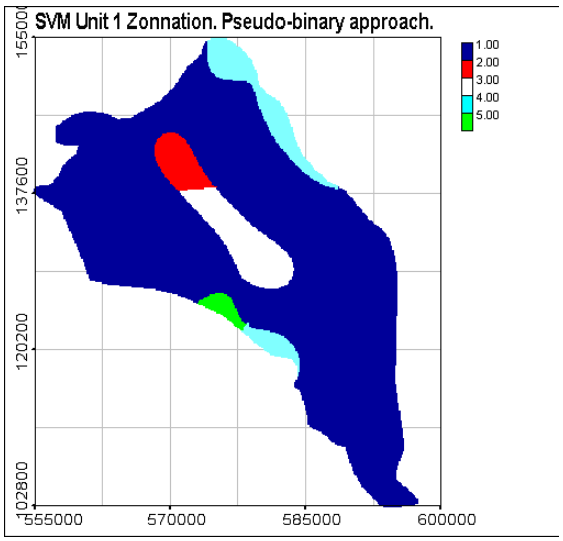


Fig. 55.U1 zonation by SVM "Pseudo-binary" classification

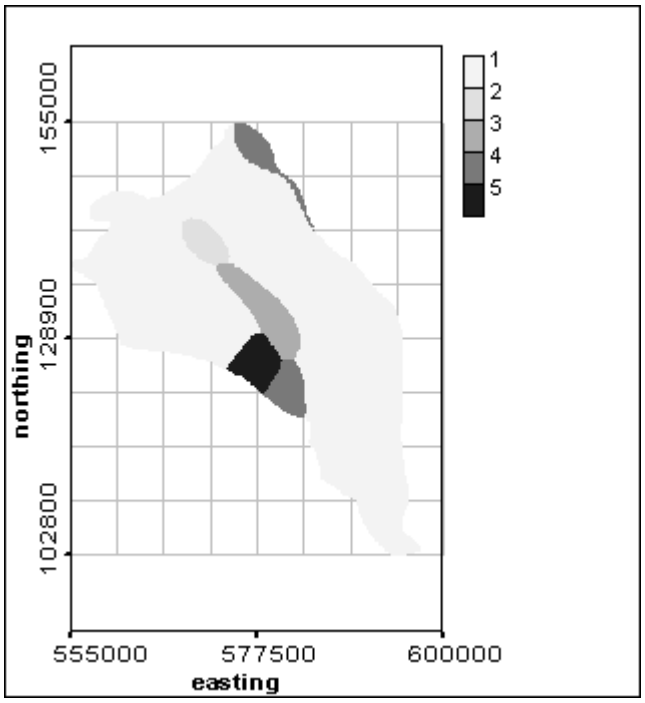


Fig. 56. U1 zonation by SVM one-against-rest classification using the confidence information

5.3. Comparison of different SVM models

The following conclusions can be made: the incorporation of the confidence information has led to a more extensive region of low testing error in the space of model parameters, but did not lead to a decrease in the value of the errors. At the same time the value of maximal testing error increased. The mapping results are almost unchanged for most of the classes (2, 3, upper part of 4), and changed for the classes with a lower number of data points: class 5 and the lower region occupied by class 4. These changes also led to changes in the occurrence of class 1, which occupies most of the study area.

According to the Hamming distance between the SVM classification and the a priori geologic map (see Tables 14 and 15) the confidence information used by SVM did not improve the results. This may be because the incorporation of the confidence data leads to a more complicated model.

Zones of difference between the results of the SVM predictions and the geologic map are presented in Fig. 57.

There appears to be no significant difference between the 2 types of SVM generalization to yield a multiclass classifier. Currently, the main problem with SVM is the absence of a probabilistic interpretation for the method.

Table 14. Hamming distance between a priori geologic map and SVM predictions

Type of generalization	Weighting	Number of differences (in pixels)	Measure of difference (in %)
ONE-AGAINST-REST	NONE	2327	5.61
ONE-AGAINST-ONE	NONE	2620	6.32
ONE-AGAINST-REST	confidence	3539	8.54

Table 15. Hamming distance between a priori geologic map and SVM predictions per class

Type of generalization	Weighting	Class1	Class2	Class3	Class4	Class5
ONE-AGAINST-REST	NONE	857	785	133	295	257
ONE-AGAINST-ONE	NONE	801	814	74	757	174
ONE-AGAINST-REST	confidence	1882	723	224	612	98

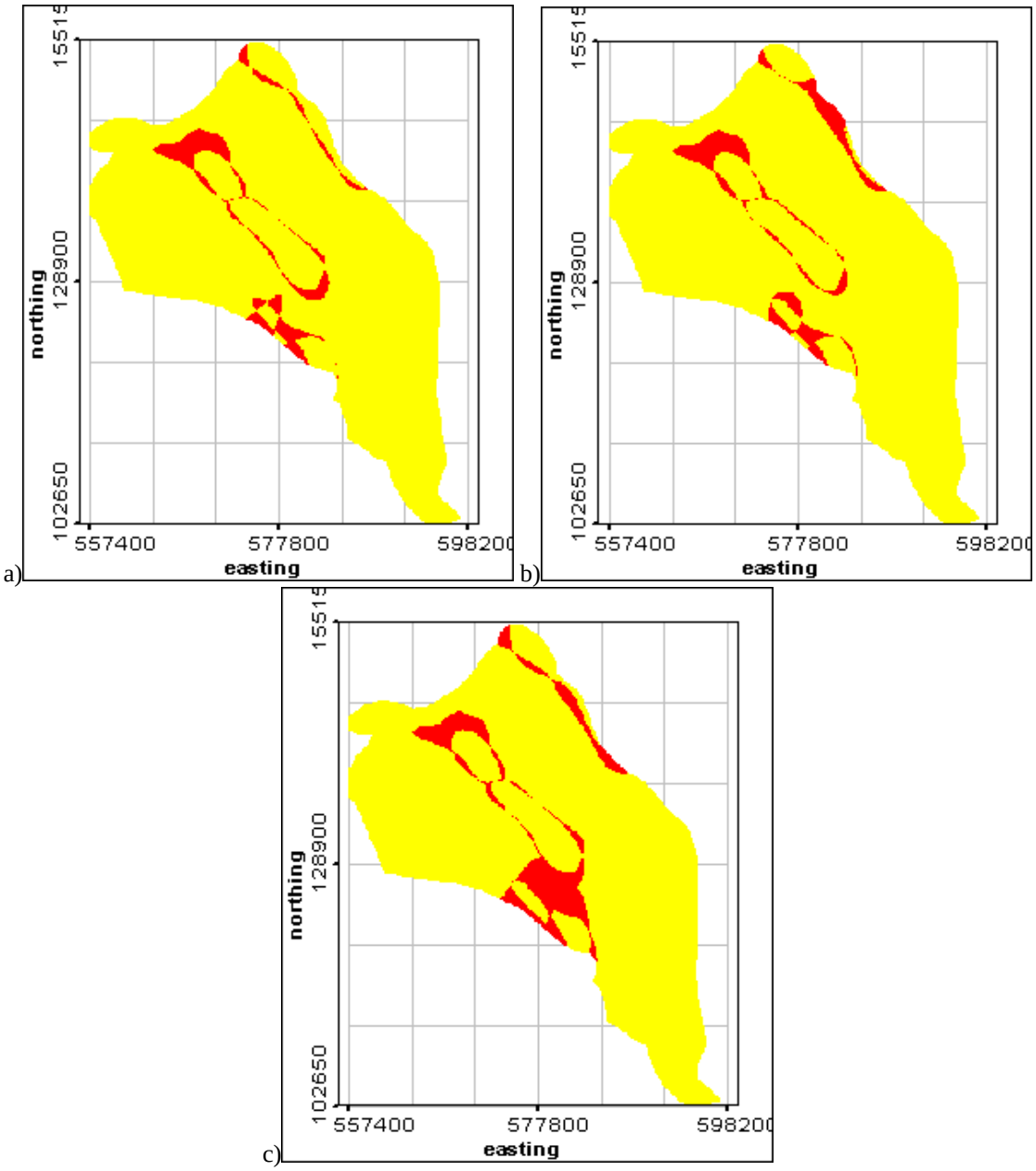


Fig. 57. Zones of difference between a priori geologic map and SVM with different generalization to multi class classification: one-against-rest (a), one-against-one (b) and one-against-rest with confidence (c). Areas of difference are dark

The three models provided by the SVM approach were analyzed using class indicator variograms. They were called SVM1 – one against rest multiclass generalization, SVM2 – one against one multiclass generalization, and SVM3 – one against rest multiclass generalization using confidence data. All three types of SVM show very good correspondence with class variograms generated from the results of that method and with the class indicator variograms generated from the a priori geologic classification. The main problem is with class 5 where the anisotropy orientation for small lags is shifted. Corresponding class indicator variogram roses are presented in Figs. 58 – 62. Class indicator variogram RMSE indexes for the SVM prediction models are presented in Table 16.

Table 16. RMSE for class indicator variograms of SVM predictions

Method	Class1	Class2	Class3	Class4	Class5
SVM1	0.017	0.015	0.004	0.001	0.0015
SVM2	0.016	0.016	0.006	0.0096	0.004

SVM3	0.02	0.013	0.003	0.005	0.019

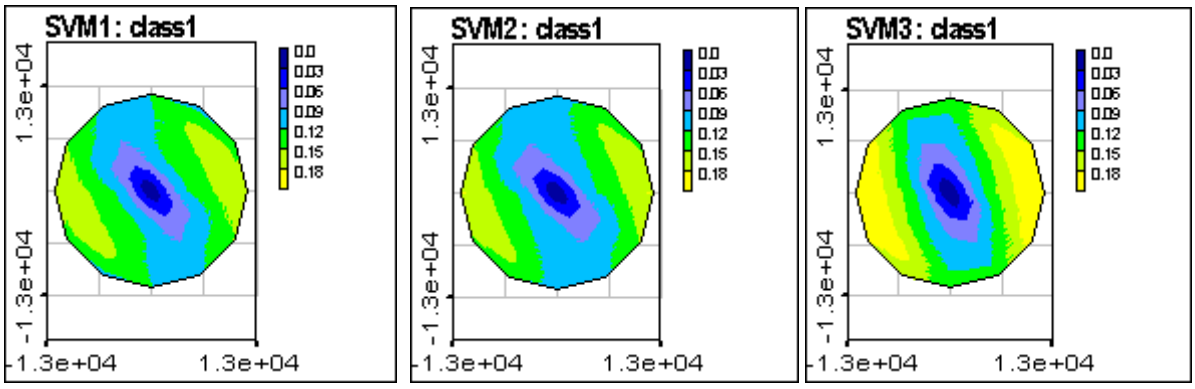


Fig. 58. Class indicator variogram roses (class1) for results of several SVM models

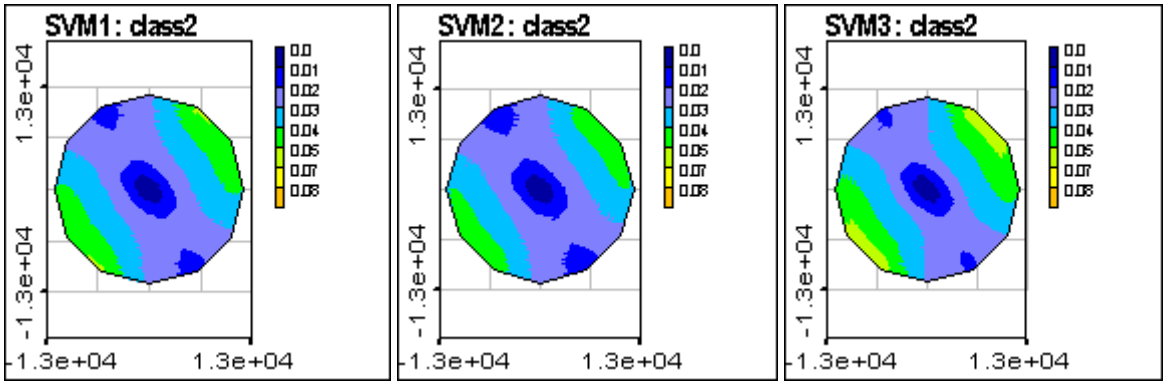


Fig. 59. Class indicator variogram roses (class2) for results of several SVM models

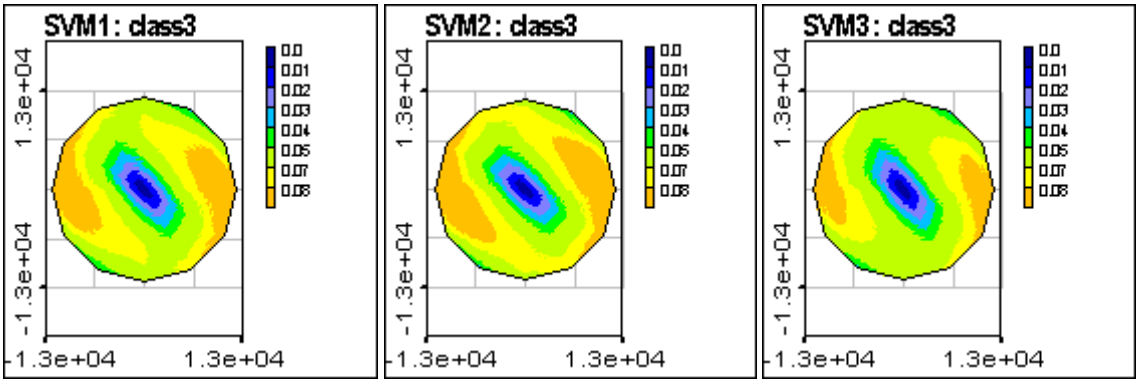


Fig. 60. Class indicator variogram roses (class3) for results of several SVM models

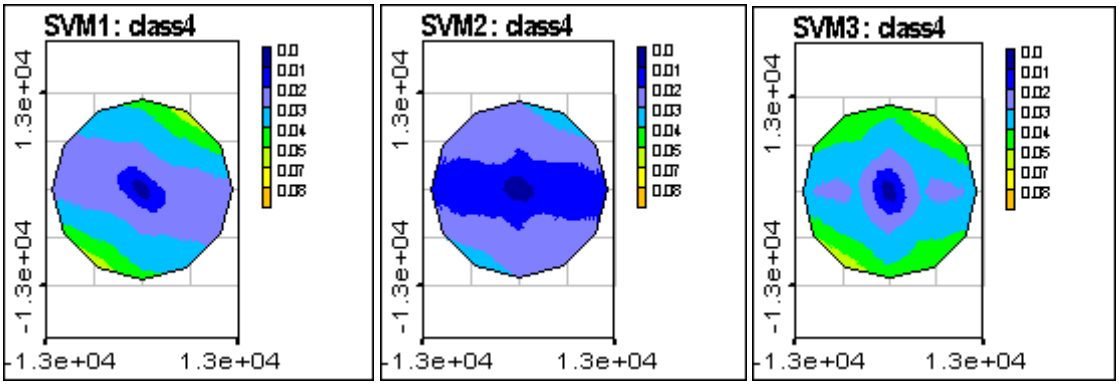


Fig. 61. Class indicator variogram roses (class4) for results of several SVM models

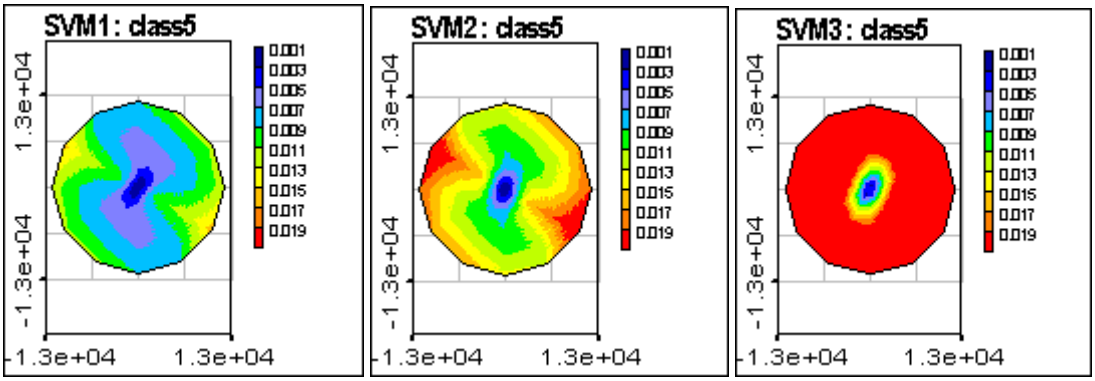


Fig. 62. Class indicator variogram roses (class5) for results of several SVM models

6. Indicator approach for U1 zonation

The first step of any indicator method's application is an indicator transform of the data. The indicator transform for categorical data is defined as:

$$I(x) = \begin{cases} 1, & x \in class \\ 0, & otherwise \end{cases}.$$

Results of the indicator transform of the hard U1 zonation data data are presented in Fig. 63. The standard indicator transform cannot be used for class 5, as this class has only one member. Class 5 can't be analyzed by geostatistical indicator methods because the spatial correlation structure of the class is impossible to estimate with only one sample, and modelling of the spatial correlation structure is the key step in all geostatistical analysis. To be able to include class 5 in the indicator based geostatistical analysis the a priori geologic information was used as a proxy. The class 5 members from the geologic classification map were inserted into the raw database (see Fig. 64a) and the indicator transform for class 5 was performed using that amended dataset (see Fig. 64b).

6.1. Indicator kriging for U1 zonation

Results of class indicator variogram modeling are presented in Figures 65 – 69. We present the raw variogram roses, model variogram roses (to compare the reproduction of spatial anisotropy by the model, if any) and directional 1D graphs (to see the reproduction of directional variograms).

Parameters of the fitted variogram models are summarized in Table 17. In this table, range 1 orresponds to the range in the main anisotropy direction and range 2 is the range in the direction perpendicular to the main direction. The orientation is measured in degrees from the X axis in the counterclockwise direction. The indicator variogram for class 5 was constructed based on class 5 members from a priori geologic classification result, as described above. Thus, "soft" geologic data was used for estimating the range and anisotropy direction of class5.

The results of U1 zonation by indicator kriging are presented in Fig. 70. The accuracy test of the indicator kriging results indicates no misclassifications, which results from the exactitude property of kriging. The lack of any misclassifications could lead to suspicion that overfitting of the data had occurred, but the rather wide zones of uncertainty cause us to reject that proposition (see Fig. 70b).

Table 17. Parameters of class indicator variogram models

	Model	Nugget	Sill	Main direction	Range 1	Range 2
Class 1	SPHER	0.	0.27	150	21870	8023
Class 2	SPHER	0.	0.023	130	8050	2289
Class 3	SPHER	0.	0.16	140	12080	2906
Class 4	SPHER	0.	0.052	120	20860	8870
Class 5	GAUSS	0.	0.38	150	48520	15240

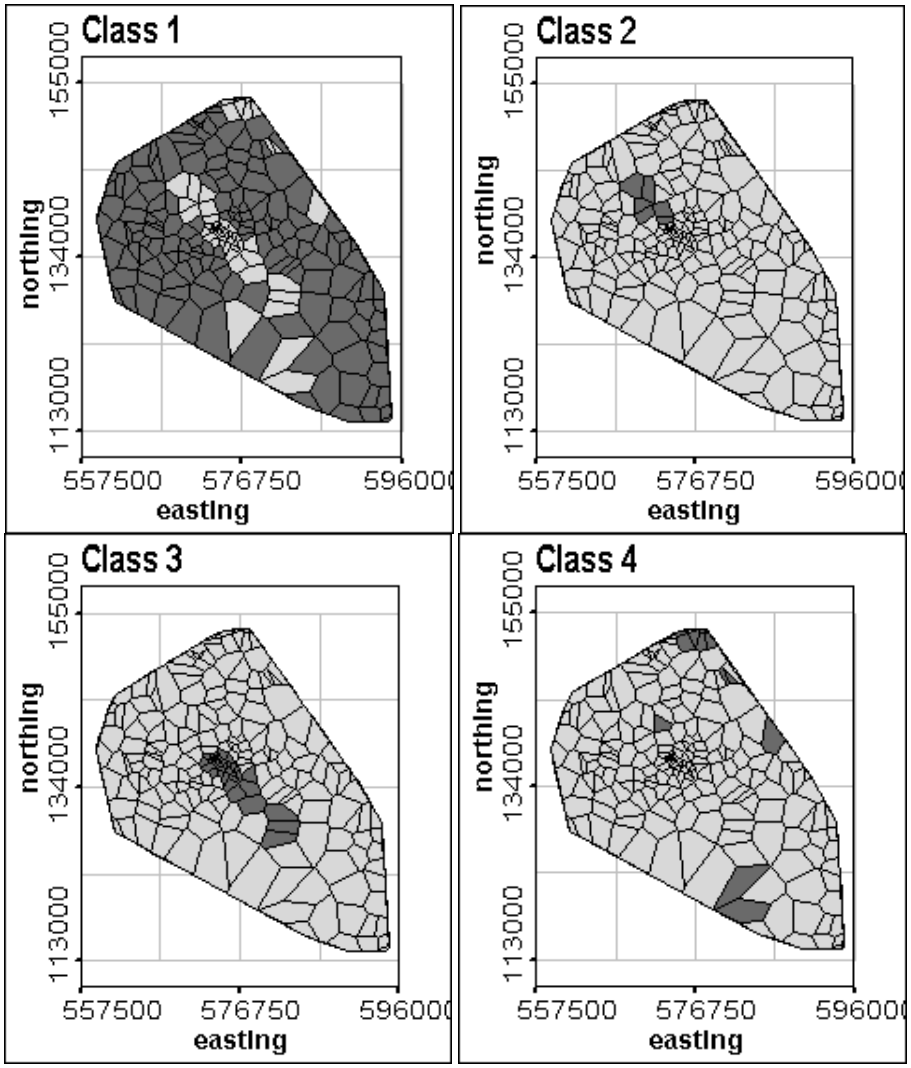


Fig. 63. Indicator transform for U1 zonation. Dark cells indicate class presence

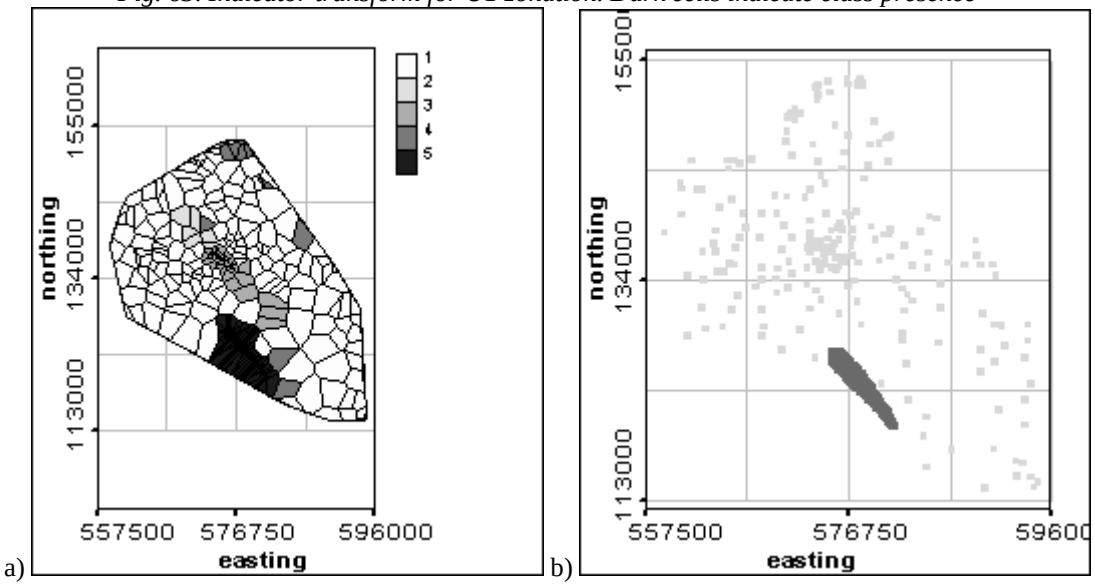


Fig. 64. Hard information on U1 zonation + soft information on class5 positioning (a) and indicator transform of amended dataset for class 5 (b)

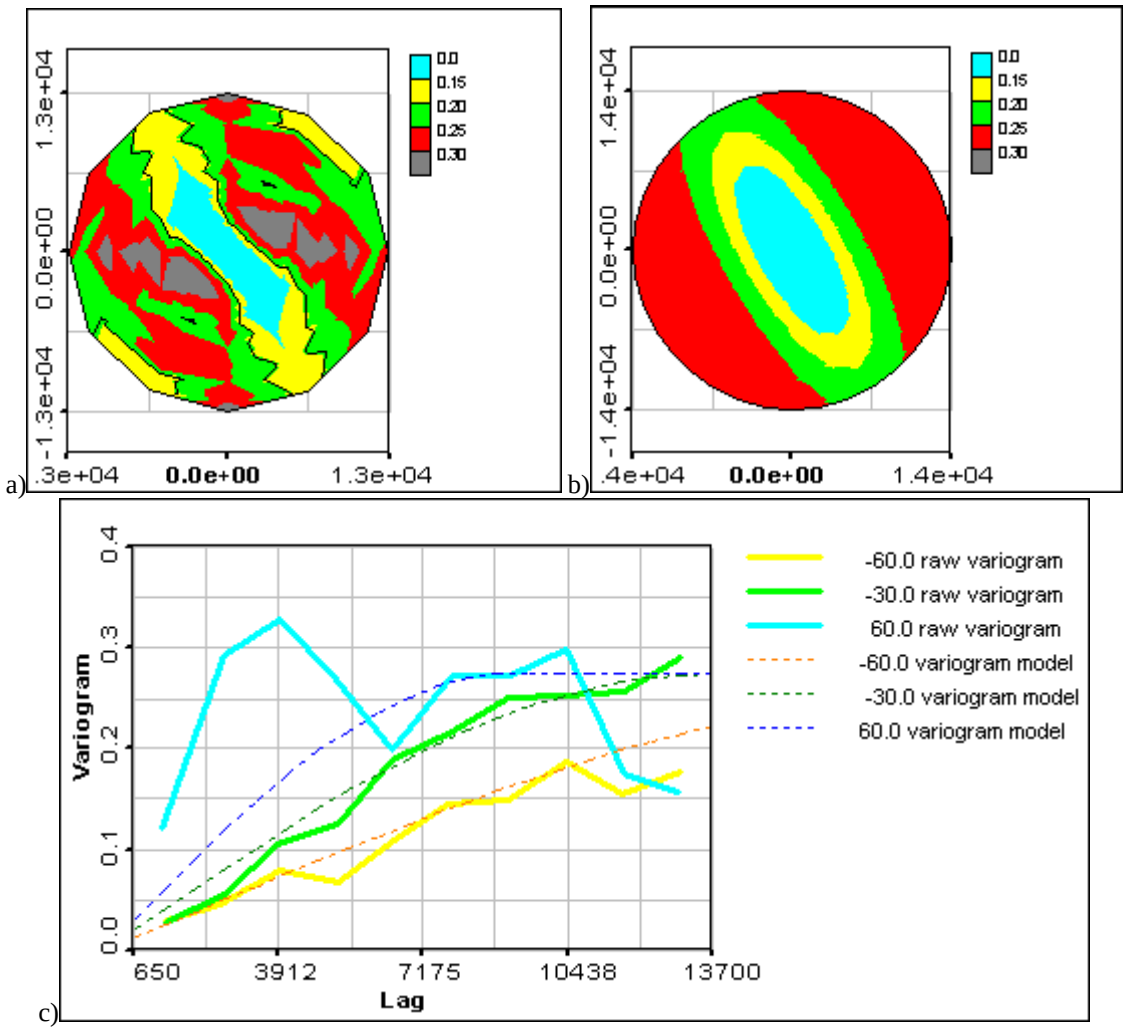
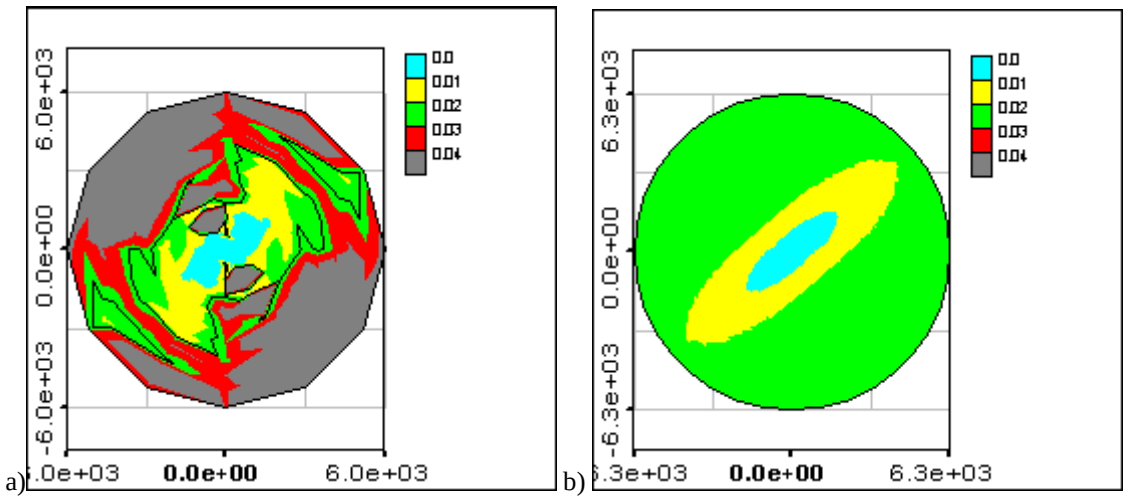
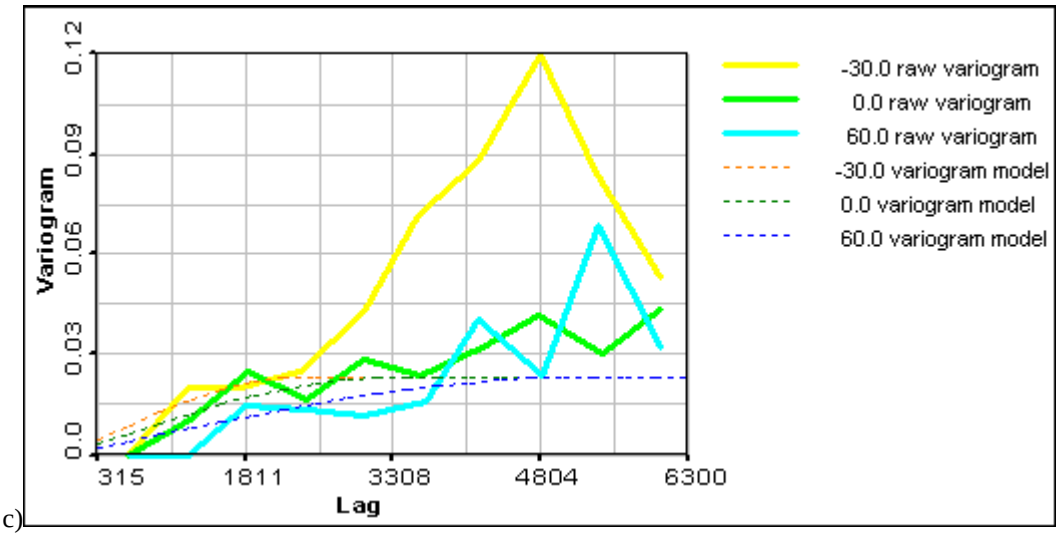


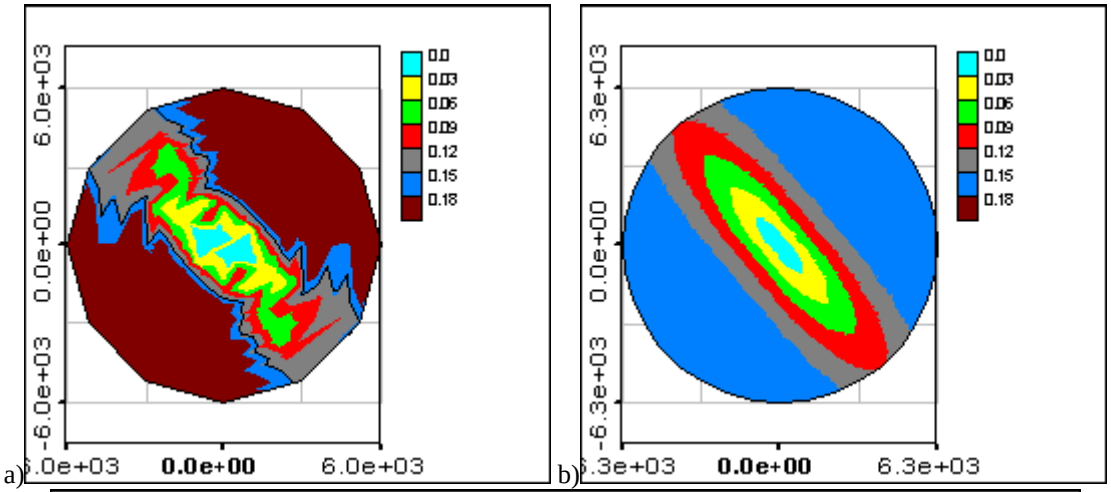
Fig. 65. Class 1 indicator variogram modeling: raw variogram rose (a); variogram rose model (b) and 1D variogram modeling (c)



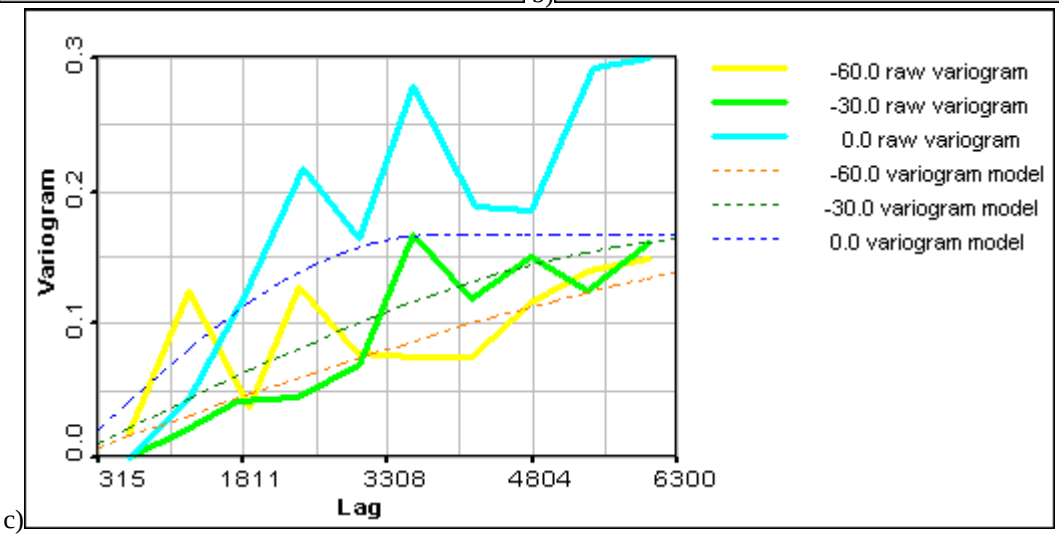


c)

Fig. 66. Class 2 indicator variogram modeling: raw variogram rose (a), variogram rose model (b) and 1D variogram modeling (c)



a) b)



c)

Fig. 67. Class 3 indicator variogram modeling: raw variogram rose (a), variogram rose model (b) and 1D variogram modeling (c)

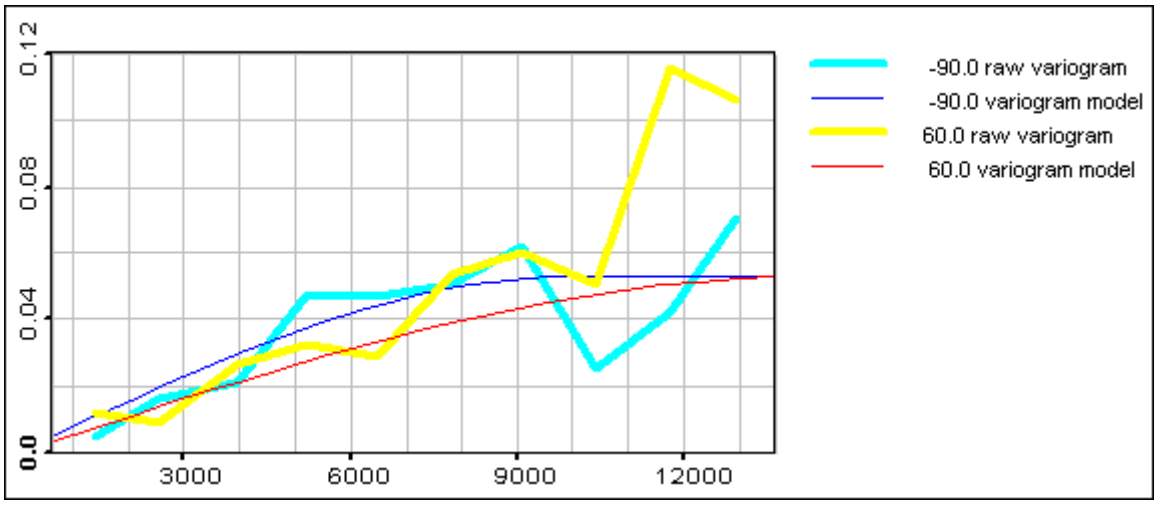


Fig. 68. Class 4 indicator variogram modeling

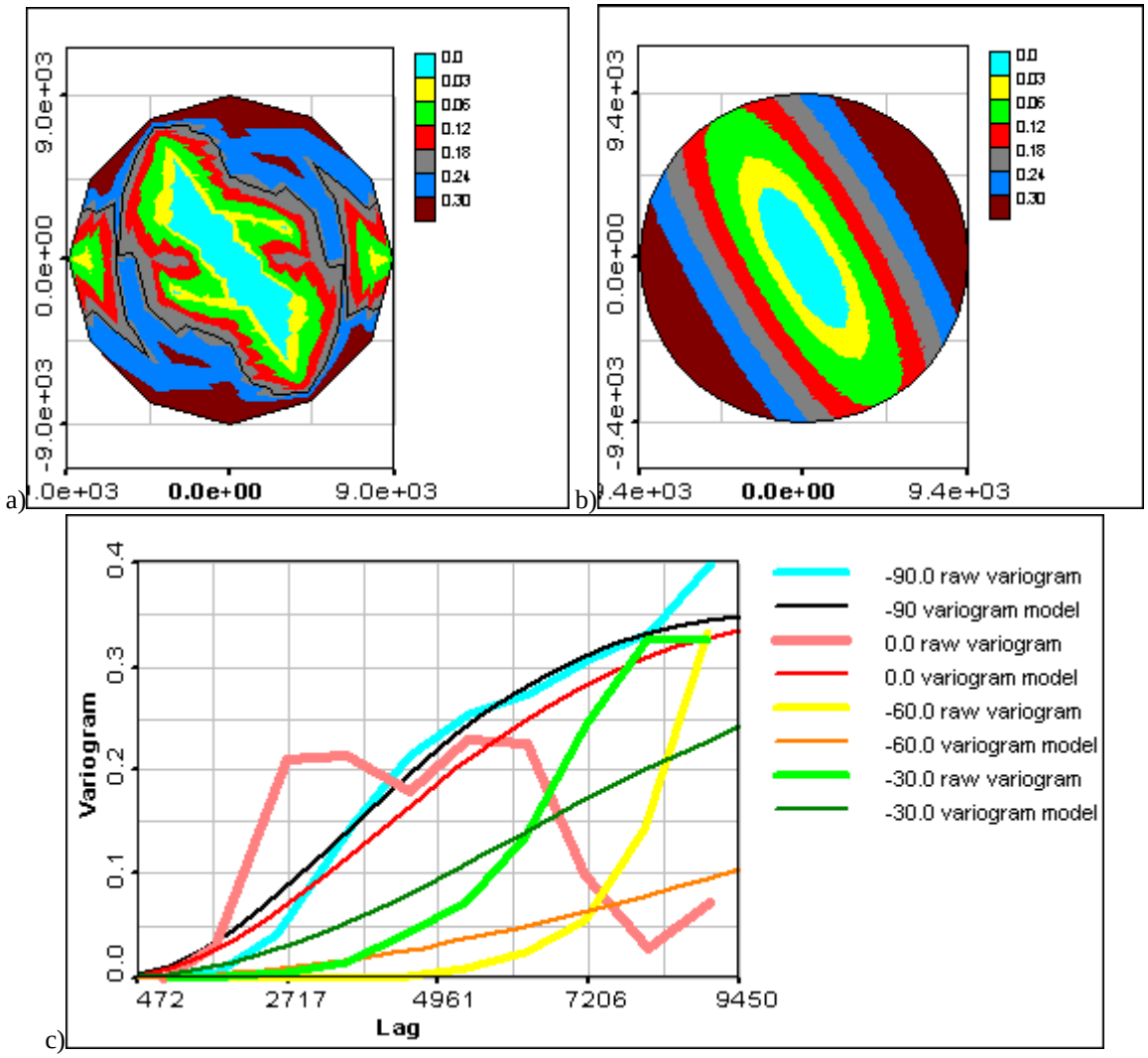


Fig. 69. Class 5 indicator variogram modeling: raw variogram rose (a), variogram rose model (b) and 1D variogram modeling (c)

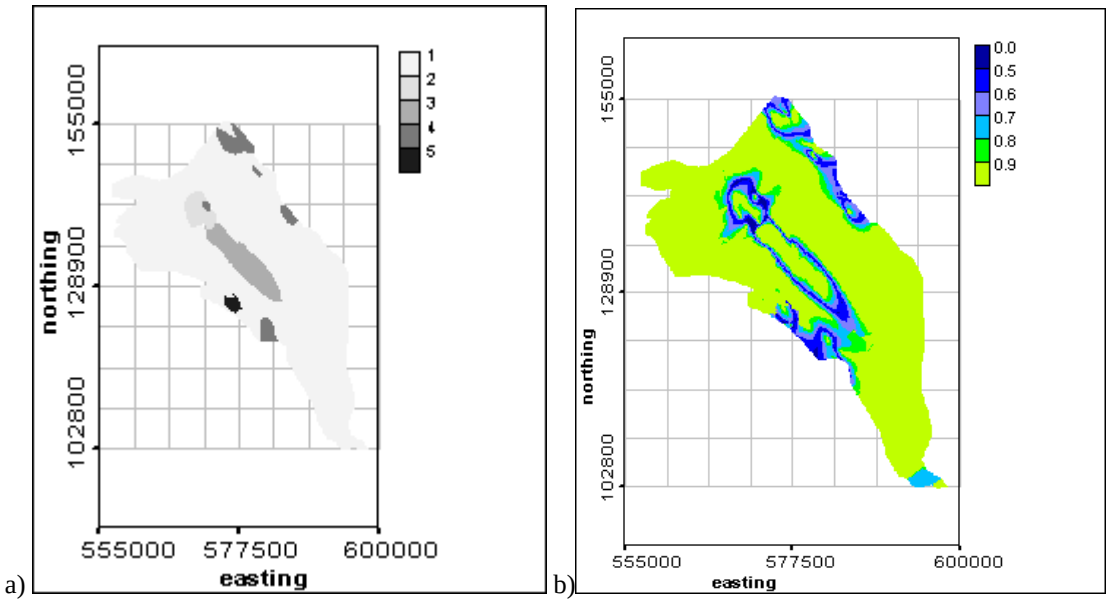


Fig. 70. U1 zonation by indicator kriging: classification (a) and probability of the class winner (b)

6.2. Indicator simulations for U1 zonation

Indicator simulations were performed using the class variogram models constructed in section 6.1. Fifty sequential categorical indicator simulations were generated. The most probable class at each location was calculated from the indicatior simulations by the following procedure:

1. On the basis of N realizations the probability of each class was estimated. The maximum probability is presented in a Fig. 71b.
2. The class is estimated as the class winner according to the class that has the maximum probability.

Such "maximum-probability" estimates of the class types over 50 realizations provides results very similar to the results provided by indicator kriging (compare Figs. 70 and 71). The main difference is in the probabilistic interpretation – the uncertainty map generated from the simulations has wider uncertainty bands than were found for indicator kriging (compare Fig. 71b and 70b). Perhaps it is caused by non-stability of class proportions in the realizations. In the current set of realizations we found very high variability of class2 proportion in realizations – range of proportions was from 0.024 to 0.11 (mean value 0.048, standard deviation 0.02). Others classes also showed non-stability, but not of such order. The other problem with class 2 was appearance of it in realizations in unexpected locations (even where other class was measured and no signs of class2 were detected).

After different investigations we found that the main source of unstable behavior of the realizations was the choice of variogram models and search radii used for the simulations. In particular, large ranges and consequently search radii appeared to cause the major problem. Changing only variogram model for class 2 led to an increase of proportion variability of the other classes. So, the variogram models (for all classes) were significantly changed. New class indicator variogram parameters are summarized in Table 18. The difference can be observed comparing parameters from Tables 17 and 18. The main changes performed upon variogram parameters can be formulated as follows:

- Ranges were significantly decreased – some of them up to 3 times;
- For classes 2, 4 and 5 omnidirectional models were selected;
- The main direction for anisotropic classes was chosen the same, to simplify the determination of search area.

Table 18. Updated parameters of class indicator variogram models

	Model	Nugget	Sill	Main direction	Range 1	Range 2
Class 1	SPHER	0.	0.25	150	17520	3874
Class 2	SPHER	0.	0.033	omni	3762	
Class 3	SPHER	0.	0.19	150	10040	4614
Class 4	SPHER	0.	0.04	omni	7863	
Class 5	SPHER	0.	0.35	omni	16980	

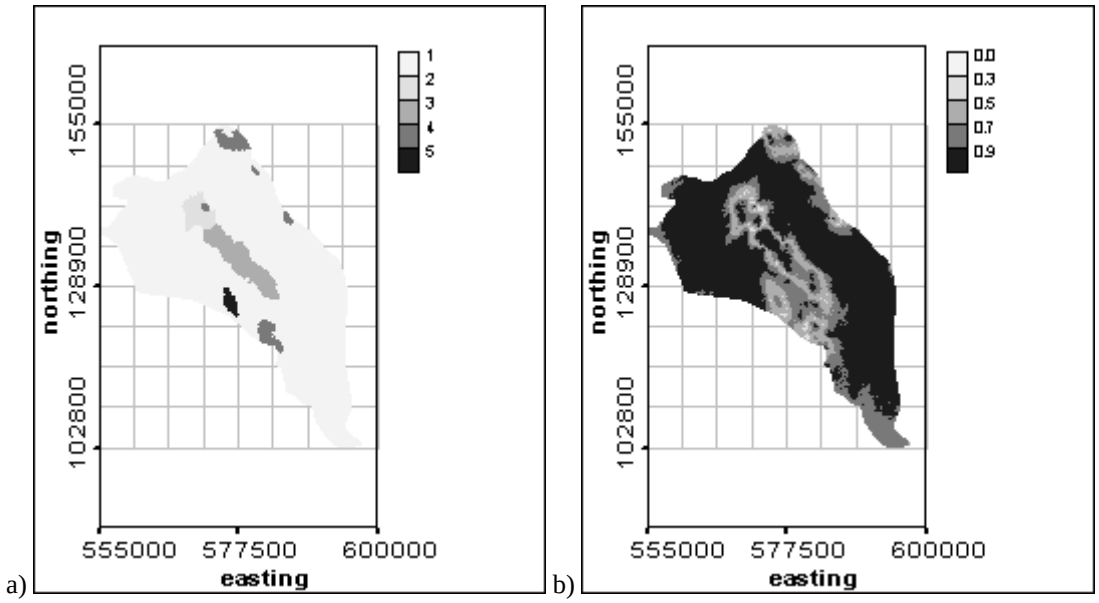


Fig. 71. U1 zonation by determining the most frequently simulated class for the suite of 50 sequential indicator simulations: classes (a), probability of the class winner (b)

New variogram models were first used to estimate the probability of each class (one class against all – binary stochastic simulations). Stochastic simulations were used instead of indicator kriging, which might be traditionally used for that purpose, because indicator kriging provided reasonable results for the original variogram models:

- no misclassifications in the accuracy test;
- very natural looking prediction on the grid.

All unexpected fluctuations appeared in single realizations. Results of probabilities calculated for classes from 2 to 5 on the base of 20 realizations of binary stochastic indicator simulations are presented in Fig. 72. The results seem to be reasonable. Spots of high probability appear in locations where real measurements indicate the presence of the same class.

So, the new class indicator variogram models were used for the generation of categorical stochastic indicator simulations. The other parameters important for simulations are the following:

- Search area is anisotrophic with main direction 150 degrees and search radius 10000 and 5000 m (reduced in comparison with preliminary set of simulations).
- The number of neighbors taken into account for estimation was reduced to 10 samples and 5 simulated points.
- Ordinary kriging was used instead of simple recommended in the theory. In reality the case is not stationary – classes 2, 3 and 5 are presented each by a spot, class4 is presented by 3 spots. The only class really distributed over the whole area is class 1. In such situation a priori pdfs used as mean for simple kriging do not reflect correctly local situations. Ordinary kriging provides more reliable local estimate.

The mean class proportions for 100 realizations, together with standard deviation and the values selected as proportions to be used in simulations using simple kriging (after some declustering analysis) are presented in Table 19. Small values of standard deviations indicate that there is no strong variability of proportions between realizations. Also obtained proportions seem to correspond more closely with the expected values.

Table 19. Mean proportions of 100 new realizations

	Mean proportion	Sigma of proportion	Expected proportion
Class 1	0.846	0.0086	0.85
Class 2	0.018	0.0015	0.04
Class 3	0.056	0.0065	0.05
Class 4	0.071	0.0036	0.048
Class 5	0.009	0.0057	0.012

Six realizations are presented as examples in Figure 72. The proportions of classes provided by these realizations are collected in Table 20. It can be seen that proportions really do not show fluctuations. But it can be remarked that mean proportion of class 2 is underestimated, while the proportion of class 4 is overestimated. This variation should be evaluated in a more detailed analysis.

Realizations present lots of single points of a class surrounded by another class. This can be caused by some kind of overfitting because of reduced variogram ranges. However, the same situation is mentioned in [4], and commonly occur in stochastic simulations of categorical indicators. The pattern generated by the realizations is unlikely to reproduce correctly the real geologic situation, because changes in geologic class would not be expected to be so sharp. So the simulation smoothing algorithm suggested by Deutsch [4] was employed – the class at each simulated location was based on the most frequent class found in the 12 surrounding nodes. The results of this smoothing (for the same 6 realizations) are presented in Figure 73. The patterns of the processed simulations are much smoother, with classes showing better connections and the class proportions are more stable (see Table 21).

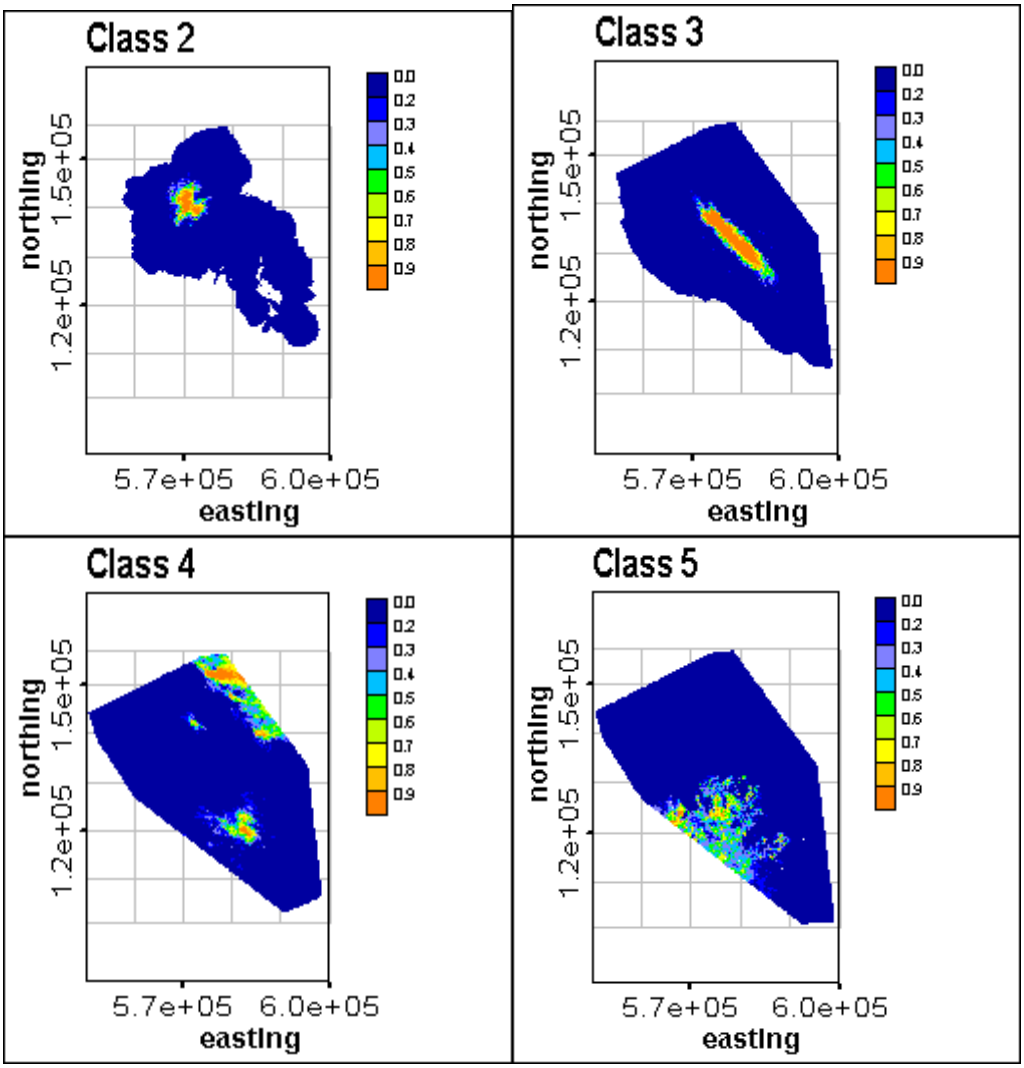


Fig. 72. Class probabilities obtained based on binary stochastic simulations

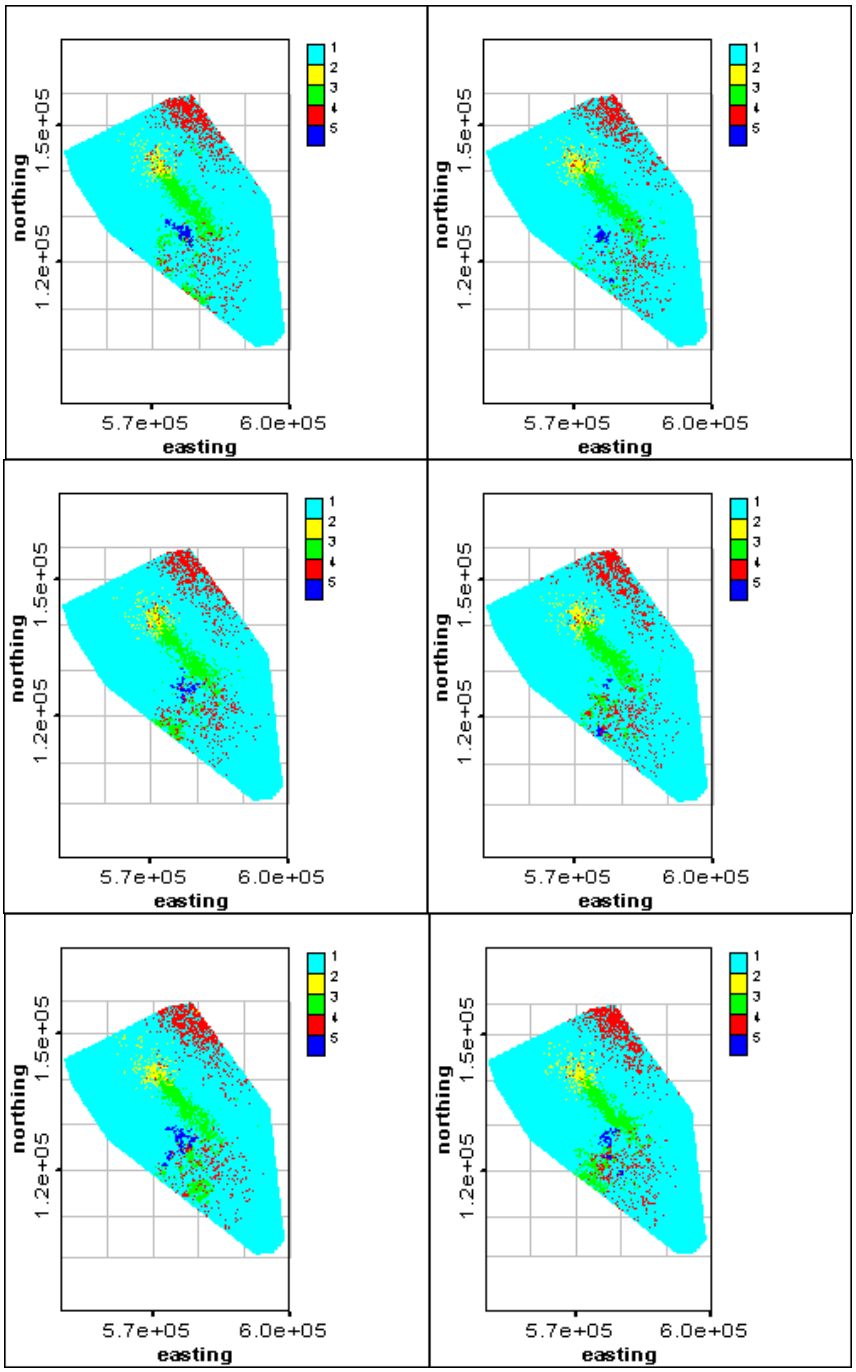


Fig. 73. First 6 realizations provided by categorical indicator simulation procedure

Table 20. Proportions of 6 first realizations

Realization number	Class 1	Class 2	Class 3	Class 4	Class 5
1	0.84	0.02	0.065	0.072	0.008
2	0.85	0.01	0.05	0.07	0.005
3	0.849	0.02	0.058	0.07	0.005
4	0.85	0.015	0.056	0.07	0.003
5	0.845	0.017	0.058	0.07	0.009
6	0.84	0.017	0.06	0.07	0.007

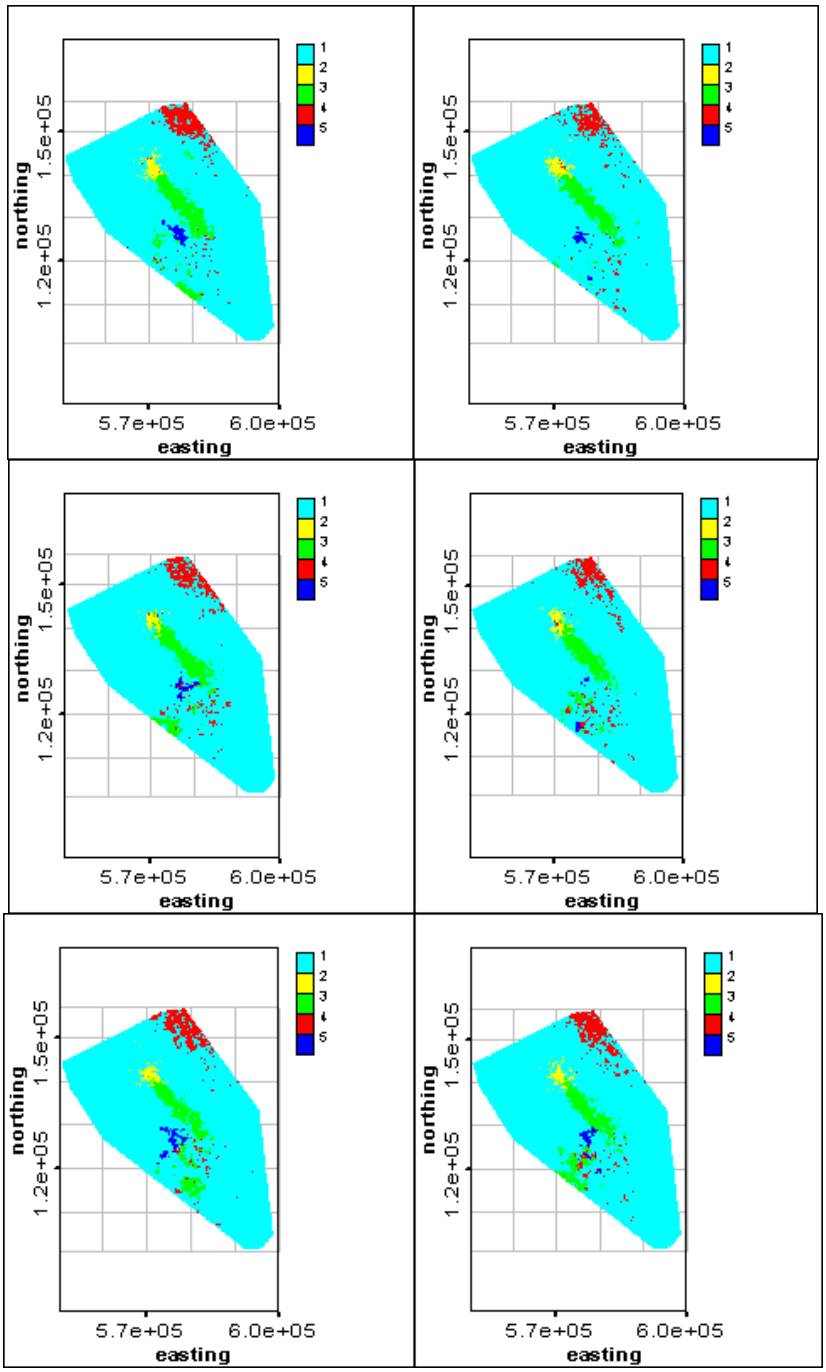


Fig. 74. First 6 realizations obtained during categorical indicator simulations after 12 neighbor smoothing procedure

Table 21. Proportions of 6 first realizations after smoothing

Realization number	Class 1	Class 2	Class 3	Class 4	Class 5
1	0.88	0.011	0.062	0.043	0.009
2	0.9	0.012	0.048	0.037	0.005
3	0.89	0.009	0.056	0.041	0.006
4	0.89	0.01	0.052	0.041	0.003
5	0.88	0.01	0.056	0.042	0.009
6	0.88	0.01	0.057	0.045	0.007

The main conclusions concerning the indicator stochastic simulations for U1 zonation problem:

- The main problem deals with real non-stationarity of data. Theoretically, geostatistical approaches were not developed for such conditions. Some localization performed during estimation, especially by the use of ordinary kriging during the simulations, allowed us to stabilize the results.
- Smoothing procedures are required and need to be performed with attention. It is possible that the smoothing process should be adaptive – different for different areas and classes.
- A priori geologic information can be important for such simulations – when there are too few members of classes (e.g., class 2 has few members, but is expected to have higher proportion).

6.3. Comparison of indicator approach predictions with an a priori geologic map

Results provided by the indicator approach (indicator kriging and ”maximum-probability” over 50 stochastic realizations) were compared the a priori geologic map. General Hamming distances and Hamming distances per class are presented in Tables 22 and 23, respectively. Zones of difference between the indicator based predictions and the a priori geologic map are presented in Fig. 75.

Both indicator kriging and indicator simulation provide very good results, as indicated by the similarity to the a priori information. The zones of difference tend to occur in zones of uncertainty, as expected.

Table 22. Hamming distance between a priori geologic map and indicator based predictions

Method	Number of differences (in pixels)	Measure of difference (in %)
Indicator kriging	2892	6.98
”Maximum-probability” for 50 realizations	3205	7.73

Table 23. Hamming distance between a priori geologic map and indicator based predictions per class

Method	Class1	Class2	Class3	Class4	Class5
Indicator kriging	983	770	129	753	257
”Maximum-probability” for 50 realizations	1064	774	266	874	227

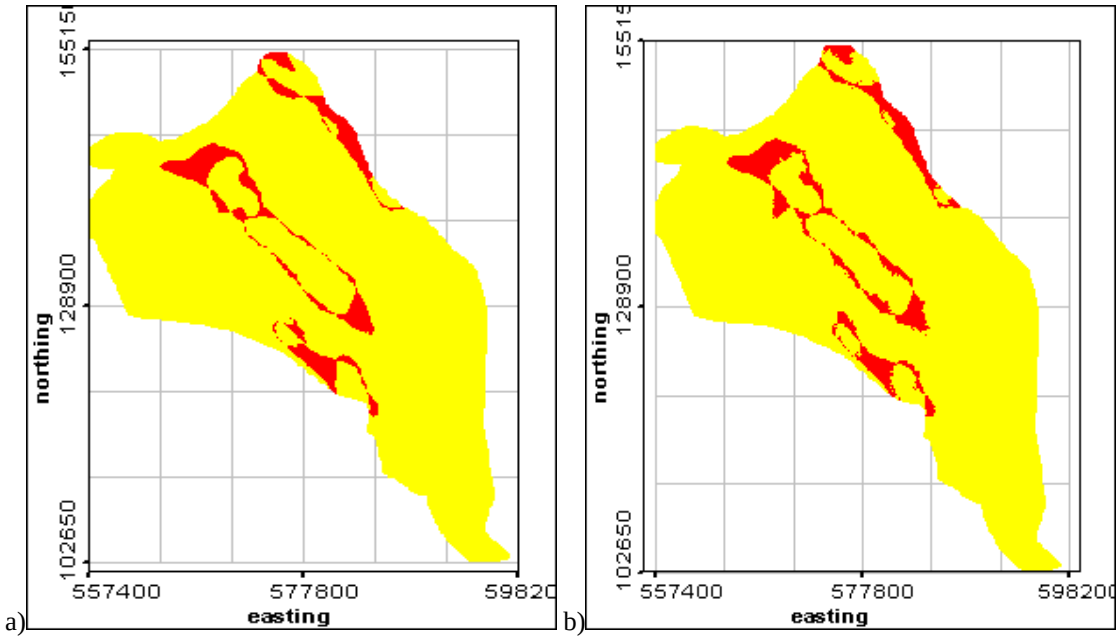


Fig. 75. Zones of difference between a priori geologic map and indicator kriging (a) and ”Maximum-probability” after 50 indicator simulation realizations (b). Areas of difference are dark

7. Discussion and Conclusions

The unit 1 zonation task was formulated as a multiclass classification problem. Multiclass modifications of the binary classification machine learning models discussed in section 1 were used to solve this problem. A set of soft information (a priori zonation, performed by Hanford Site geologists on the basis of the same samples and their expert knowledge of the region) was also used during analysis. For example, this soft data was used to estimate spatial correlation structure of the class with only 1 member among the sampled data.

Different geostatistical and machine learning methods were presented and discussed for this task. It was observed that the machine learning methods provide very narrow zones of uncertainty (zone where probability of the class winner is below 0,8). This situation can be treated as overfitting, but in the current case it seems reasonable to assume that the distribution of classes is easy for non-linear machine learning classifiers given the available distribution of samples.

One of the main problems that appeared during solving of this task was the low number of representatives in some classes (e.g., two classes had 6 or fewer sample data supporting them). A subset for validation and comparison of obtained results could not be extracted from this initial data set. This made it impossible to evaluate the results using standard methods. Several alternative evaluation approaches were proposed and used, but all of them have significant drawbacks.

- Comparison on the base of the accuracy test. The accuracy test is based on the application of the trained model to the data it was trained on. The accuracy test allows us to estimate the error of trained data reproduction by a model. The test can show which model provides better result on the training data, but it does not allow us to compare a more significant aspect of the results of the machine learning methods – whether they can be generalized to new data. One needs to be careful using just the accuracy test for comparison of models – e.g., overfitted models do very well on the accuracy test.
- Pseudo-validation can be performed on a validation subset constructed from the members of classes having sufficient number of representatives. But in the current case it is difficult to choose validation subsets even from prevailing classes, without encountering the possibility of distorting the boundaries.
- Comparison can be made on the level of correspondence with a priori information. Here the main question is the confidence in the a priori information and how exactly it is expected to be reproduced. If it is exact knowledge, the only purpose of application of any other methods is to study their abilities. In the current case we deal with an a priori map, which is not exact, but is simply one possible interpretation of the geologic data. Other experts given the same data would tend to draw different maps, with different boundaries for the classes, especially in areas where data are sparse.

In spite of these disadvantages, all possible approaches to comparison of results were used. The most attention was paid to the qualitative and quantitative comparison with the a priori soft map. Qualitative comparison was based on the visual aspects of the result, such as:

- Spatial orientation of main features (class zones);
- Spatial sizes for different class zones in special orientations (for example, too long, too narrow, etc.).

To perform reasonable qualitative comparison it is very important to carefully prepare all images: they need to be of the same size, colored according to the same scale, coordinate axes need to be the same and with same density.

Quantitative comparison can be performed using some special indices. The simplest among them is the Hamming distance – it is the quantitative measure used for comparison of the images in this study. The Hamming distance indicates the measure of difference between two images that occupy the same pixels, and it can be measured in pixels (number of pixels which differ) or in the percent of different pixels. The number of differences per class can also be estimated.

The comparison of estimation methods can also be performed using geostatistical tools, including the comparison of class indicator variograms. This method seems to be useful, but also has some problems:

- If the a priori map is not correct, it may have a distorted spatial correlation structure for more distributed classes (for example, class 4, in the current case);
- The comparison of class spatial correlation structure requires a great deal of work – each method is presented by several (5 in the current case) estimates of the spatial correlation structure – a set of directional variograms.
- Variogram rose presentation can help in visualization of directional variograms and to simplify visual comparison of directional variograms. But again all images of the variogram roses need to be carefully prepared. It is especially important for comparison of continuous values, where the scale of the variogram rose needs to be carefully selected.

- Special quantitative indices can be used for variogram comparison. For example, one can use the traditional root mean square error (RMSE), where the differences between the two variograms being compared are treated as errors.

Among the estimation methods, SVM gave good classification results. Currently the drawback of the method is the lack of an uncertainty measure for the estimate. This topic needs more advanced and basic research. Traditional approaches to derive confidence measures are usually based on i.i.d. hypotheses. Studies of special comprehensive simulations with artificial data may bring more understanding to the topic.

Application of geostatistically based methods is significantly complicated by non-stationarity of the problem. Methods that allow one to re-estimate the local mean during simulation or estimation allow one to improve the situation and to obtain reasonable results.

Nearly all methods provided reasonable results. But this comparison only allows us to check the main orientation of classes, basic forms, and the general similarity of obtained results to the geologic map. But the geologic result is itself only one of possible realizations, the main difference is that it was made using additional geologic knowledge of an expert geologist. All other obtained results can have the same interpretation. The expert with his experience can better select the best result and make the final classification solution.

Some methods can be excluded from further consideration based on the results of testing of the methods presented in the current work:

1. KNN – it provides too many misclassification errors.
2. PNN with too different values of kernel bandwidth for the different directions – it produces a very anisotropic model with long and thin classes.
3. PNN where the orientation of the kernel ellipse differs from other PNN models – they tend to change the main anisotropy of the pattern.
4. MLP with 1 hidden layer trained using class size as a weighting coefficient in the cost function.

All other results can be presented for the expertise.

In the present study the geological image was used as a reference map in order to compare models and their differences, and the Hamming distance was used for the comparison. Of course, it is not optimal because it is an aggregated index, but it gives a clear qualitative, and even quantitative, indication about the difference between models.

Aknowledgement

The work was performed under the Memorandum of Understanding between the U.S. Department of Energy and the Russian Academy of Sciences, Implementing Arrangement 2, Annex B (“Uncertainty Assessment and Advanced Modeling of Mathematical Geology in Development of Inverse Flow and Transport Models”) and supported by CRDF Grant Assistance Program (GAP) Project RM0-20201-EM50.

References

1. Specht, D., Probabilistic Neural Networks, *Neural Networks*, V. 3, pp. 109-118, 1990.
2. Parzen, E., On Estimation of a Probability Density Function and Model, *Annals of Mathematical Statistics*, V. 33, pp. 1065-1076, 1962.
3. Haykin S., *Neural networks: a comprehensive Foundation*, Macmillan College Publishing Company, Inc., Englewood Cliffs, 1994.
4. Deutsch C., *Geostatistical Reservoir Modeling*, Oxford University Press, 2002