

# SUPPORT VECTOR REGRESSION FOR AUTOMATED ROBUST SPATIAL MAPPING OF NATURAL RADIOACTIVITY

Alexei Pozdnoukhov

*IDIAP Research Institute, Martigny, Switzerland*

*Correspondence to: Alexei.Pozdnoukhov@idiap.ch*

This paper presents an application of Support Vector Regression method for the prediction of an environmental variable such as the level of natural radioactivity. The basics of the method are described, and some practical considerations are presented, including the meaning of the method's parameters and their influence on the model. The use of the prior data is discussed. It is shown how to include the information on the variance of the measurements into the model. The use of cross-validation for tuning the parameters of the algorithm is presented. Some ideas for detecting the unusual training samples with SVR are discussed. Generally, the case study illustrates the usefulness of the considered approach for automated spatial mapping tasks in the presence of prior data.

## 1. INTRODUCTION

A number of state-of-the-art methods can be used for the task of predictive spatial mapping: deterministic interpolations, methods of geostatistics: the family of kriging estimators, see, for instance, (Deutsch and Journel, 1997), machine learning algorithms such as artificial neural networks (ANN) of different architectures, hybrid ANN-geostatistics models (Kanevski et al., 1996), etc.

Environmental empirical data is always corrupted with noise, and often with noise of unknown nature. That's one of the reasons why deterministic models can be inconsistent, since they treat the measurements as values of some unknown function that should be interpolated. Kriging estimators treat the measurements as the realization of some spatial random process. To obtain the estimation with kriging one has to model the spatial structure of the data: spatial correlation function or (semi-)variogram (Deutsch and Journel, 1997). This task can be complicated if there is not sufficient number of measurements. ANN is a powerful tool, but it is also suffer from the number of weaknesses. ANNs of a special type – multiplayer perceptrons – are often used as a de-trending tool in hybrid (ANN+geostatistics) models (Kanevski et al., 1996). So, we are interested in a method that would be robust to noise in measurements, would deal with the small empirical datasets and has solid mathematical background. The method should also allow for incorporating some prior knowledge of different types.

The present paper deals with such a model, called Support Vector Regression (SVR). It is based on Statistical Learning Theory (SLT), which is a general mathematical framework devoted to the problem of estimation of the dependencies from empirical data (Vapnik, 1998). SLT gave rise to a family of Machine Learning algorithms, based on

Support Vector approach. Concerning spatially distributed data these learning methods predict unknown mapping between inputs (spatial co-ordinates and secondary variables) and outputs (random function) from available data and a priori knowledge (Kanevski et al., 2002). The general property of the algorithms is that they aim at minimizing the empirical error (error on available training data) together with controlling the complexity of the model.

This paper is devoted to the automated application of SVR model for spatial data mapping in the presence of prior data (Pozdnoukhov et al., 2002). The complete methodology of SVR application should consist of a number of steps: monitoring network analysis, understanding of clustering, exploratory data analysis, exploratory variography, understanding of spatial continuity; data preparation: splitting of data in training, testing and validation data subsets, SVR training and testing, selection of the optimal SVR hyper-parameters, understanding the quality of the results by using exploratory analysis (statistics, variography) of the residuals, validation of the results, spatial data mapping with optimal SVR model. First stages are common for all the environmental mapping tasks and will be considered briefly. For the suggested case study, based on SIC2004 competition, this scheme has to be modified. We will present the way to tune the parameters of the SVR algorithm based on prior data, and on a standard cross-validation procedure as well. Another modification of the algorithm deals with estimating the uncertainty of the measurements and the usage of this information to refine the SVR model.

The task of the SIC2004 competition, as it was perceived by the authors, was to provide a stable noise-robust automatic tool for mapping the natural radioactivity. SVR is a robust regression approach, which performs particularly well in the presence of noise and outliers in data. Thus, it is an appropriate method for mapping the slowly changing and relatively stable spatial variables, such as natural radioactivity level.

The preliminary investigations on how to detect *anomalies* using SVR are provided.

## 2. METHODOLOGY

This section presents the basics of the Support Vector Regression method. Without going into details, we present those issues, which are important for the present case study, i.e. for automation of the mapping process, usage of prior knowledge, and anomaly detection.

### 2.1 STATISTICAL LEARNING THEORY

In Machine Learning one's aim is to find ("learn") an algorithm (modelling/mapping function) that describes training data and has good generalization abilities (that is, allows for accurate predictions at the new points, where the desired quantity is unknown). Statistical Learning Theory (SLT) is devoted to such problems of extracting knowledge from finite empirical data.

The following bounds of the generalization error were derived in SLT:

$$R(h) \leq R_{emp}(h) + R_{conf}(h), \quad (1)$$

where  $R$  is a bound of testing error,  $R_{emp}$  is an empirical risk on the training data (training error), and  $R_{conf}$  is a confidence term which depends on the "complexity" of the modelling functions.

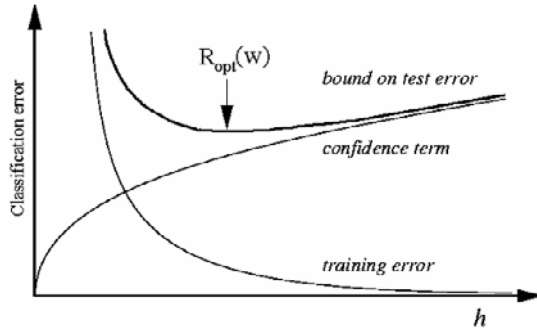


Figure 1  
Bound on test error derived in SLT. The minimum corresponds to some optimal complexity of the model for a given dataset.

The parameter that characterizes the «complexity» is called the VC-dimension  $h$  of the set of modelling functions. This parameter is introduced in SLT to describe the general notion of “complexity”, independently on the particular functions used to model the data. VC-dimension is plotted on the x-axes in Fig. 1. According to the bounds, derived in SLT, the relevant strategy for constructing a learning machine is to minimize the training error while maintaining  $h$  small (see Fig. 1). This principle is referred to as Structural Risk Minimization. This idea is realized for the specific learning tasks and results in a family of Support Vector algorithms.

## 2.2 SUPPORT VECTOR REGRESSION

Given the training data  $\{(x_i, y_i), i=1, \dots, L\}$ , where  $x_i$  are input coordinates and  $y_i$  are the measured values, the quality of how the model  $f(\cdot)$  fits the data can be expressed in terms of the *empirical risk*:

$$R_{emp} = \frac{1}{N} \sum_{i=1}^N Q(y_i - f(x_i), x_i), \quad (2)$$

where function  $Q$  is a *loss function*. It provides the penalty, given to the model in a case of discrepancy between the training data and its prediction by the model. Support Vector Regression is based on a new type of so-called  $\varepsilon$ -insensitive loss functions. For example linear  $\varepsilon$ -insensitive loss is defined as:

$$Q(y - f(x), x) = \begin{cases} |y - f(x)| - \varepsilon & \text{if } |y - f(x)| > \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

This loss function is  $\varepsilon$ -insensitive in the sense that no penalty is given to the model, if its discrepancy with the training data point is less than some non-negative constant  $\varepsilon$ . Otherwise, the penalty increases linearly with the value of the corresponding difference.

Next, one has to select some class of functions  $f(x)$ , and formulate the criterion which minimizes both the complexity of the functions and the empirical risk. Please refer to (Vapnik, 1998) for detailed explanations. Here we present the final formulation of the model. SVR predictions are calculated using the following function:

$$f(x) = \sum_{i=1}^N (\alpha_i^* - \alpha_i) K(x_i, x) + b, \quad (4)$$

where  $\alpha_i$  and  $\alpha_i^*$  are the weights,  $K(x_i, x)$  is a user-defined kernel function, and  $b$  is a constant threshold. Kernel function is some symmetric positive-definite function which may depend on some parameters. The commonly used kernel function is an isotropic Gaussian RBF:

$$K(x', x'') = e^{-\frac{\|x' - x''\|^2}{2\sigma^2}}, \quad (5)$$

with the bandwidth parameter  $\sigma$ . The weights  $\alpha_i$  and  $\alpha_i^*$  are obtained from the solution of the following Quadratic Programming optimisation problem:

$$\begin{aligned} & \text{maximise} \quad -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(x_i, x_j) \\ & \quad - \varepsilon \sum_{i=1}^N (\alpha_i^* + \alpha_i) + \sum_{i=1}^N y_i (\alpha_i^* - \alpha_i) \\ & \text{subject to} \quad \begin{cases} \sum_{i=1}^N (\alpha_i^* - \alpha_i) = 0 \\ 0 \leq \alpha_i^*, \alpha_i \leq C \quad \text{for } i = 1, \dots, N \end{cases} \end{aligned} \quad (6)$$

One of the general properties of the algorithm is linked with the bounds for the weights  $\alpha_i$  and  $\alpha_i^*$ . It appears that some (sometimes, the major part) of the training points obtain zero weights. It means that they exert no influence on the model. The training samples with non-zero weights are called the *Support Vectors* (SV). The amount of SVs is linked with the model complexity, which increases with the number of SVs.

There are several parameters that influence the SVR solution and have to be provided by user:  $\sigma$ ,  $C$  and  $\varepsilon$ . Next we will consider the properties of each parameter.

### 2.3 PARAMETERS OF THE MODEL

$\sigma$  – kernel bandwidth. The most evident properties depending on this parameter are: at small  $\sigma$  values (much less than some characteristic scale of the region) the model is close to overfitting, at large  $\sigma$  values (of the order of the region's diameter) to oversmoothing. The optimal value of this parameter mainly depends on two characteristics of the data: correlation radius and data variability. In general, the task of optimal sigma choice is connected with the complicated problem of monitoring network analysis: whether one can describe given/unknown phenomenon with the measurements from given monitoring network.

$C$  – the parameter that defines the trade-off between training error and model complexity.  $C$  defines the upper bound for the multipliers  $\alpha_i$  and  $\alpha_i^*$  (see Eq. 6), hence defines the maximal influence the point can exert on the solution. These considerations allow to conclude: 1) the more noisy the data the less should be the value of  $C$ , at the same time the lower the value of  $C$ , the smoother the results of prediction mapping 2)  $C$  should be

not much less than the maximum value of the observed training data in order to fit the high values well.

$\varepsilon$  – the width of the insensitive region of the loss function. This is the parameter that defines the sparseness of the SVR solution - the points that lie inside the  $\varepsilon$ -tube have zero weights. It is the main parameter that incorporates the information of the measurements' quality. It should be of the same order as the measurements' accuracy or as the square root of the nugget variance in model variogram. It also influences the smoothness of the mapping: the larger its value, the smoother the result. Let us remind, that nugget incorporates measurement errors and small scale variability. Next, let's note that this parameter can be selected individually for every training data sample. It would mean that we allow different discrepancy to be out of penalty at different locations. This may be based on the variations in measurements precision over the considered region.

The main parameter that controls the impact of anomalies is  $C$ . Since it defines the upper bound for the weights and kernel function is bounded, extremely high or low values will be ignored by the SVR. These values will be treated as outliers. The corresponding training residuals will be unusually large.

## 2.4 PRIOR DATA FOR TUNING THE ALGORITHMS

The prior datasets are described in details in the paper (Dubois & Galmarini, 2005). There were 10 datasets available. This information was used to tune the parameters of the SVR algorithm.

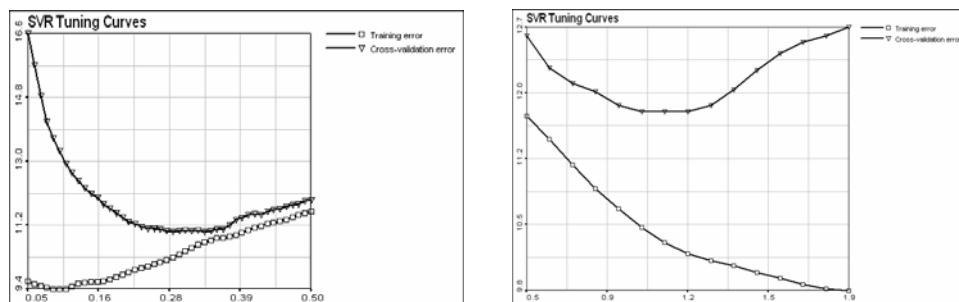


Figure 2  
SVR training and cross-validation RMS error curves. Left:  $C = 30$ , x axis – kernel bandwidth  $\sigma$ ; Right: x axis – logarithm of parameter  $C$ ,  $\sigma=0.3$ . Parameter  $\varepsilon$  is fixed at  $\varepsilon = 10$  at both graphs.

The general approach, which is commonly used in Machine Learning, is an analysis of cross-validation error. Cross-validation error is proven to be an unbiased estimate of the validation error (Luntz & Brailovskiy, 1969). So far, given a dataset for analysis and SVR model with three parameters, one has to calculate the cross-validation error for all possible values of the parameters. The ranges of parameter variation can be chosen taking into account the properties we described above. As an example, let us present the cross-validation error for the first prior dataset. Fig. 2 (previous page) presents training RMSE and cross-validation RMSE as a function of different values of the kernel bandwidth  $\sigma$  (left) and logarithm of the parameter  $C$  (right). The training error is calculated using model predictions of the training data. It illustrates how well the model fits the training data. Let's note that training error provides a way to control the model complexity (see

also Fig. 1). The training error is usually nonzero due to the presence of noise in data and limited complexity of the model.

Similar results were obtained for the other prior datasets. Finally, the parameters for predictive mapping were fixed at  $\sigma = 0.3$ ,  $\varepsilon = 10$  (see the details below),  $C = 30$ .

To incorporate the knowledge about the quality of the measurements and their preciseness, it is possible to set parameter  $\varepsilon$  individually for every training data location. Fig. 3 presents the mean and the standard deviation of the measurements calculated over 10 prior datasets. Voronoi polygons are used for the visualization in Fig. 3.

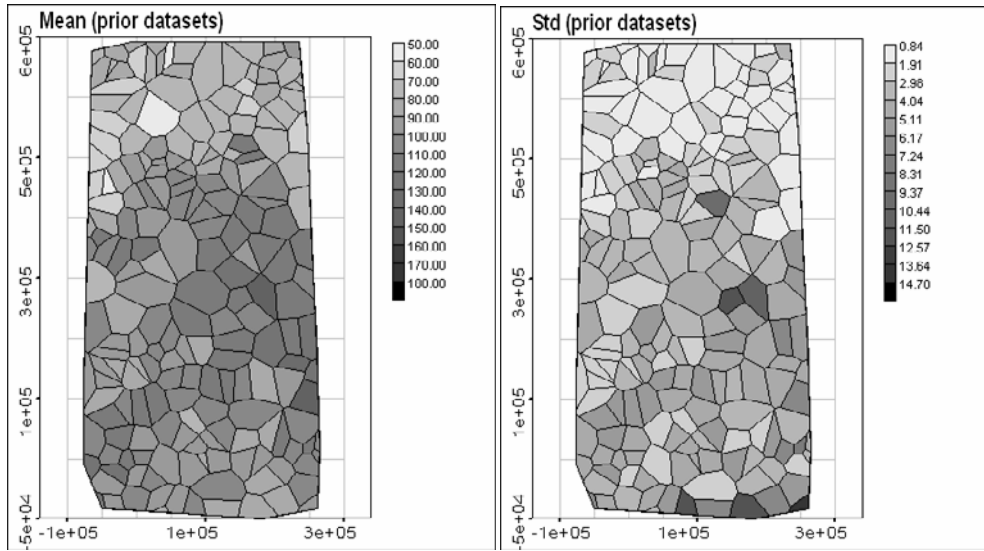


Figure 3  
The mean (left) and the standard deviation (right) of the measurements calculated over 10 prior datasets.

Parameters  $\varepsilon_i$  were set as follows:  $\varepsilon_i = \varepsilon \cdot \gamma_i$ , where the common multiplier  $\varepsilon$  was tuned by cross-validation and set to  $\varepsilon=10$ , and  $\gamma_i$  are the values of standard deviation for every measurement (see Fig. 3, right). The value of  $\varepsilon$  was found by the direct search over the set of pre-defined values; the optimal value  $\varepsilon=10$  corresponds to the lowest cross-validation error.

Cross-validation procedure can be made fully automatic. Since the cross-validation error curves are smooth enough, it is possible to find reasonable parameters according to its minima either by direct search or more sophisticated procedures. For the present case study, however, the parameters were fixed at the values obtained after the preliminary analysis of the prior data sets, as it was described before.

### 3. RESULTS

The algorithm was applied for the estimation of the validation dataset. The predicted values were compared with the real ones, provided in the final phase of the SIC2004 competition. The results are described below.

### 3.1 OVERALL RESULTS

The table below gives the minimum, maximum, mean, median and standard deviation of the 808 estimated values and observed values, for the two datasets.

N = 808	Min.	Max.	mean	median	std. dev.
Observed (first data set)	57.0	180.0	98.0	98.8	20.0
Estimates (first data set)	68.9	125.2	98.0	100.7	14.1
Observed (second data set)	57.0	1528.2	105.4	99.0	83.7
Estimates (second data set)	69.0	125.6	98.7	101.5	14.6

Table 1  
Comparison of the estimated and measured values (nSv/h).

Data sets:	MAE	ME	Pearson's <i>r</i>	RMSE
First data set	9.22	-0.037	0.79	12.47
Second data set	16.25	-6.70	0.28	81.0

Table 2  
Comparison of the errors.

Fig. 4 presents the omnidirectional variogram of the first validation data set and the omnidirectional variogram of the validation residuals: difference between SVR predictions and true values. As one can see, the variogram of the residuals demonstrate the behaviour close to almost pure nugget effect. Therefore, almost all spatially structured information was extracted from data by SVR model. The presence of minor correlation at small scales may suggest that the selected model represents slight over-smoothing.

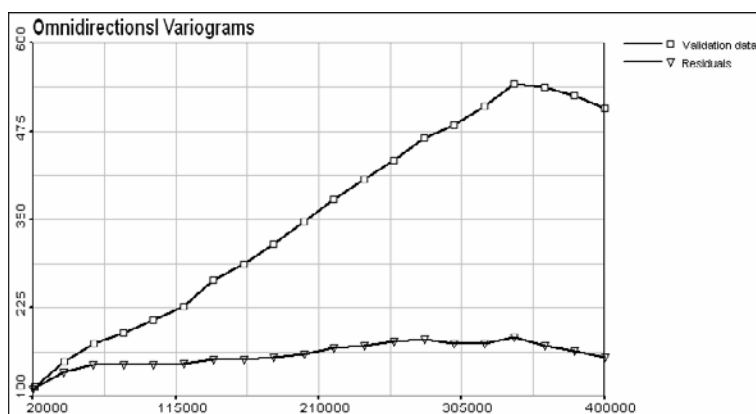


Figure 4  
Omnidirectional variogram of the validation data and the residuals obtained with SVR model predictions. Almost all spatially structured information was extracted by SVR model.

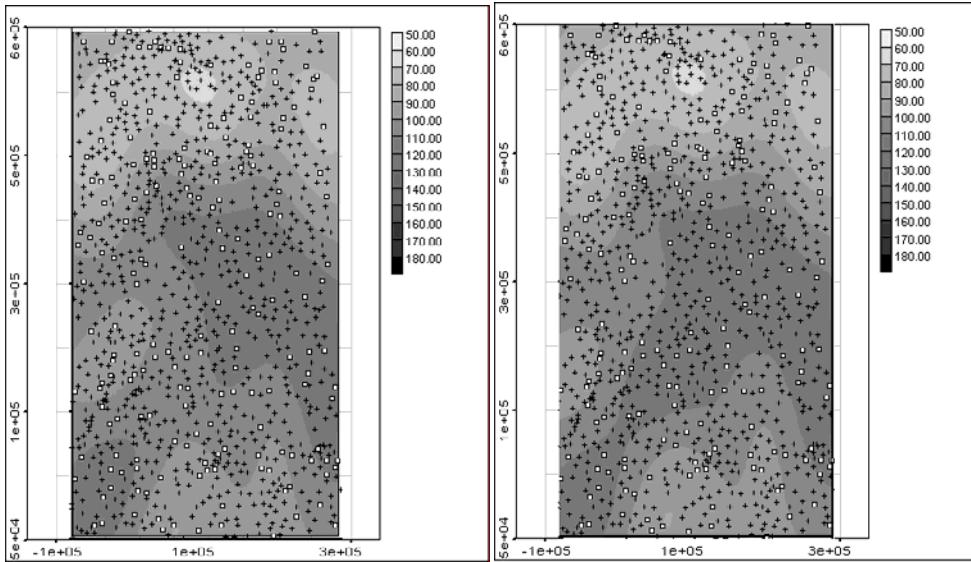


Figure 5

Predictive mapping for the 1<sup>st</sup> set (let) and the 2<sup>nd</sup> set (right). The predicted values are shown in 10 nSv/h. Crosses point to the locations of the estimated values; empty squares indicate the locations of the input values.

Predictive mapping results are presented in Fig. 5 for both datasets. As one can see, there is only a minor difference between two datasets. This is the consequence of the selected approach for data modelling, which aimed to provide a robust approach for spatial predictions using noisy data. The model basically ignored the unusually high values of the measurements. They were treated as outliers.

Uncertainty assessment is still an open issue in Support Vector Regression method. There's not a well-founded approach to quantify the uncertainty attached to the estimation.

### 3.2 DETECTING ANOMALIES AND OUTLIERS

As it was mentioned before, the model was built to provide robust and noise insensitive tool for predicting the natural radioactivity level. The approach was not aimed at modelling the anomalies. SVR model deals particularly well with the outliers by “neglecting” these unusual values. This is an advantage of the model if the outliers are due to noise. But if the appearance of these anomaly values is due to some different spatial process, the robustness of the classic SVR can be considered as a drawback.

However, there is an evident way to assess the presence of such abnormal data points in the dataset by considering the residuals of the training data and the corresponding weights. For the present case study, the training residuals for the locations of modelled emission were abnormally (10 times) large. This is a clear indication that these locations need a special treatment, since they indicate outliers (probably noise) or come from a different phenomenon from the one, which was modelled (artificial emission and not a natural radioactivity).



## 4. DISCUSSION

Support Vector Regression is a flexible model that allows modelling environmental variables. For the present case study we observed promising performance for predicting the natural radioactivity level.

The average computational time for model training and predictions was around 10 seconds. Generally, for  $N$  input data samples, it takes between  $O(N^2)$  and  $O(N^3)$  operations to train the model (depending on implementation, particular case study, and parameters) and  $O(N \cdot L)$  operations to make predictions in  $L$  locations.

Concerning the anomaly detection, the model is able to detect the abnormal data samples by registering the unusual training residuals. However, if one needs to model these situations, an advanced tool has to be used. It could be a multi-scale SVR model, which provides predictions as a mixture of models for several spatial scales. It is possible to define these scales according to the minima of cross-validation error.

An important open issue is also a problem of how to provide confidence measure for SVR predictions.

## 5. CONCLUSIONS

The paper considered the application of Support Vector Regression model for the prediction of environmental variable such as the level of natural radioactivity. The basics of the SVR method are described, and some practical considerations were presented, including the meaning of parameters and their influence on the model.

Prior data were used for tuning the parameters of the model. This data also allowed estimating the variance of the measurements. This information was included into the SVR by the precise choice of the parameters of insensitivity. At the same time, SVR model provides limited capabilities for modelling unusual phenomena due to its robustness. The standard SVR algorithm considers extreme values as outliers.

It was mentioned how to use the SVR algorithm for detecting the unusual training samples.

## CODES

The analysis and modelling of the prior data sets were carried out using the GeoStat Office software. The GeoSVM module was used for Support Vector Regression algorithm. Please consult the book (Kanevski and Maignan, 2004) or contact the author for the details.

---

## REFERENCES

- Deutsch C. V. and A. G. Journel. *GSLIB. Geostatistical Software Library and User's Guide*, New York: Oxford University Press, 1997.
- Dubois, G., Galmarini, S. 'Introduction to the Spatial Interpolation Comparison (SIC) 2004 exercise and presentation of the data sets'. *Applied GIS* 2005; 1 (2): 09-01 to 09-11.
- Kanevski M., R. Arutyunyan, L. Bolshov, V. Demyanov, M. Maignan. 'Artificial neural networks and spatial estimations of Chernobyl fallout'. *Geoinformatics* 1996; 7 (1): 5-11.
- Kanevski M., Pozdnukhov A., Canu S., Maignan M. 'Advanced spatial data analysis and modeling with support vector machines'. *International Journal of Fuzzy Systems* 2002 Mar; 4 (1): 606-616.

- Kanevski M. and Maignan M. *Analysis and Modelling of Spatial Environmental Data*. EPFL Press, Lausanne, 2004.
- Luntz A., Brailovskiy V. 'On estimation of characters obtained in statistical procedure of recognition'. *Technicheskaya Kibernetika* 1969; 3.
- Pozdnoukhov A., Kanevski M., Maignan M., Canu S. 'Robust mapping of spatial data with Support Vector Regression'. Preprint IBRAE-2002. Nuclear Safety Institute RAS, 2002, 15p.
- Vapnik V. *Statistical Learning Theory*. New York: John Wiley & Sons, 1998.

Cite this article as: Pozdnoukhov, A. 'Support vector regression for automated robust spatial mapping of natural radioactivity'. *Applied GIS*, Vol 1, No 2, 2005. pp. 21-01 to 21-10. DOI: 10.2104/ag050021

Copyright © 2005 Alexei Pozdnoukhov