Stochastic Simulation of Snow Cover

Markéta Průšová
Institute of Geoinformatics
VSB - Technical University of Ostrava
Ostrava – Poruba, Czech Republic
Email: marketa.prusova@vsb.cz

Lucie Juřikovská
Institute of Geoinformatics
VSB - Technical University of Ostrava
Ostrava – Poruba, Czech Republic
Email: lucie.jurikovska@vsb.cz

***Abstract*—The presented paper deals with a stochastic simulation of snow cover. This study aims to find the best settings of a stochastic simulation to be able to determine the parameters of the snow cover for any point of a given territory. Next, basic statistical analyses of parameters are documented, including an analysis of relationships between the snow parameters and altitude, slope and aspect. Most current methods of spatial interpolation and multifactor evaluation are based on the weighted regression relationships. That leads to smooth results and degrades our ability to properly evaluate the existence and the probability of extreme situations and their impact on the research problem. Neither alternative techniques use neural networks to bring major improvements. This research is exploring the possibility of stochastic simulation to assess the development of values, evaluating the occurrence of extreme events, monitoring the probability of exceeding the set limits, compared with application kriging errors, the use of additional qualitative information. The variants of conditional stochastic simulation were tested in particular. The application area was chosen on data of snow cover, many land-bound factors and the results are the regular mapping of forest damage. The aim is to compare and determine the best method of interpolation of snow cover, which was succeeded.**

Keywords—interpolation; stochastic simulation

I. INTRODUCTION

Geostatistical simulation is a well-known approach for modeling spatial uncertainty of a regionalized variable [3][7][8], by generating a high number of plausible realizations of a random function, conditional to the experimental data. Most interpolation methods (including kriging) give smooth images of the spatial variable while a simulation tries to mimic the true variability described by second order functions like the covariance or the variogram [11].

Snow is a dynamic natural element, the distribution of which is largely controlled by latitude and altitude. The regional extent of snow cover is an important variable in hydrology. In the hydrological cycle, snow represents seasonal water storage from where water is rapidly released during the melting period [12]. Prediction of snow cover is important for agriculture and flood prevention.

One of the main factors governing the distribution of snow properties is topography [2][10]. Although snow property data such as snow-water equivalent are often available in considerable temporal detail from a single point, the spatial resolution of snow property data is poor [13]. Often, only a few point measurements are available in the catchment of interest. Because of the extreme spatial variability of snow properties, small samples of these point data may not be representative of spatial patterns and spatial averages [5][6].

II. METHODOLOGY

Simulation is broadly defined as the process of replicating reality using a model. In geostatistics, simulation is the realization of a random function that has the same statistical features as the sample data used to generate it (measured by the mean, variance, and semivariogram). Gaussian geostatistical simulation (GGS), more specifically, is suitable for continuous data and assumes that the data, or a transformation of the data, has a normal (Gaussian) distribution. The main assumption behind GGS is that the data is stationary—the mean, variance, and spatial structure (semivariogram) do not change over the spatial domain of the data. Another key assumption of GGS is that the random function being modeled is a multivariate Gaussian random function [1].

Stochastic simulation differs from kriging in two ways, as follows:

- Kriging provides the “best”, that is, minimum variance, local estimates without regard to the resulting statistics of those estimates. In simulation, however, the aim is to reproduce the global statistics and maintain the texture of the variation, and these take precedence over local accuracy.
- A kriged estimate at any place has associated with it a variance, and hence an uncertainty, that is independent of estimates at all other places. Confidence about it is usually based on an assumed Gaussian distribution with the mean equal to the estimate and a cumulative distribution function [14].

Increased use of GGS follows a trend in geostatistical practice that emphasizes the characterization of uncertainty for decision and risk analysis, rather than producing the best unbiased prediction for each unsampled location (as is done with kriging), which is more suited to showing global trends in the data [4][5]. Simulation also overcomes the problem of conditional bias in kriged estimates (high-value areas are

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-169-4

typically underpredicted, while low-value areas are usually overpredicted).

Geostatistical simulation (GS) generates multiple, equally probable representations of the spatial distribution of the attribute under study. These representations provide a way to measure uncertainty for the unsampled locations taken all together in space, rather than one by one (as measured by the kriging variance). Moreover, the kriging variance is usually independent of the data values and generally cannot be used as a measure of estimation accuracy. On the other hand, estimation accuracy can be measured by building distributions of estimated values for unsampled locations using multiple simulated realizations that are built from a Simple Kriging model using input data that is normally distributed (that is, data that either is normally distributed or has been transformed using a normal score or other type of transformation). These distributions of uncertainty are key to risk assessment and decision analysis that uses the estimated data values [1].

GS assumes that the data is normally distributed, which rarely occurs in practice. A normal score transformation is performed on the data so that it will follow a standard normal distribution. Simulations are then run on this normally distributed data, and the results are back-transformed to obtain simulated output in the original units. When Simple kriging is performed on normally distributed data, it provides a kriging estimate and variance that fully define the conditional distribution at each location in the study area. This allows one to draw simulated realizations of the random function (the unknown, sampled surface) knowing only these two parameters at every location, and is the reason that GGS is based on a Simple kriging model and normally distributed data.

Results from simulation studies should not depend on the number of realizations that were generated. One way to determine how many realizations to generate is to compare the statistics for different numbers of realizations in a small portion of the data domain (a subset is used to save time). The statistics tend toward a fixed value as the number of realizations increases.

In conditional simulation, however, the generator must return the data values at places where we know them in addition to creating plausible values of $Z(x)$ elsewhere. We condition the simulation on the sampled data, $z(x_i)$, $i=1,2,...,N$. Denote the conditionally simulated values by $z_C^*(x_j)$, $j=1,2,...,T$. Where we have data we want the simulated values to be the same:

$$z_C^*(x_i) = z(x_i) \text{ for all } i=1,2,...,N. \qquad (1)$$

Elsewhere, $z_C^*(x)$ may depart from true but unknown values in accord with the model of spatial dependence adopted [14].

Consider what happens when we krige Z at x_0 where we have no measurement. The true value, $z(x_0)$, is estimated by $\hat{Z}(x_0)$ with an error $z(x_0)- \hat{Z}(x_0)$, which is unknown:

$$z(x_0) = \hat{Z}(x_0) + \{ z(x_0)- \hat{Z}(x_0) \} \qquad (2)$$

A characteristic of kriging is that the error is independent of the estimate, that is

$$E[\hat{Z}(y)\{z(x)- \hat{Z}(x)\}] = 0 \text{ for all } x,y \qquad (3)$$

This feature is used to condition the simulation.

We create a simulated field from the same covariance function or variogram as that of the conditioning data to give values $z_S^*(x_j)$, $j=1,2,...,T$, that include the sampling points, x_i, $i=1,2,...,N$. When we krige at x_0 from the simulated values at the sampling points to give an estimate $\hat{Z}_S^*(x_0)$. Its error, $z_S^*(x_0)- \hat{Z}_S^*(x_0)$, comes from the same distribution as the kriging error in equation (2), yet the two are independent. We can use it to replace the kriging error to give our conditionally simulated value as

$$z_C^*(x_0) = \hat{Z}(x_0) + \{ z_S^*(x_0)- \hat{Z}_S^*(x_0)\} \qquad (4)$$

The result has the properties we desire, as below [14].

1. The simulated values are realizations of a random process with the same expectation as original:

$$E[\hat{Z}_S^*(x)] = E[Z(x)] = \mu \text{ for all } x \qquad (5)$$

   where μ is the mean.
2. The simulated value should have the same variogram as the original.
3. At the data points the kriging errors $z(x_0)- \hat{Z}(x_0)$ and $z_S^*(x_0)- \hat{Z}_S^*(x_0)$ are 0, and $z_C^*(x_0) = z(x_0)$

Conditional simulation is more appropriate than kriging where our interest is in the local variability of the property and too much information would be lost by the smoothing effect of kriging. A suite of conditional simulations also provides a measure of uncertainty about the spatial distribution of the property of interest [14].

III. PILOT AREA

The pilot area is defined by the Šance catchment. It is located in the Frýdek-Mistek district. Šance reservoir was constructed on the upper course of the river Ostravice. All water drains to the river Odra.

The catchment network is defined by a regular grid of 2 x 2 km oriented along the axes of the coordinate system S-JTSK, which in total contains 52 squares. The grid is used for the systematic schema of sampling. Representative places are established inside each square cell. At least one place represents an open land and another place represents a forested land. A set of measurements (52) is performed in each representative place (around the point). The final sample data represents an average of all measured values (excluding outliers) in the location. A two kilometer step was selected. Treeless (open) and forested areas are carried out in one measurement in squares.

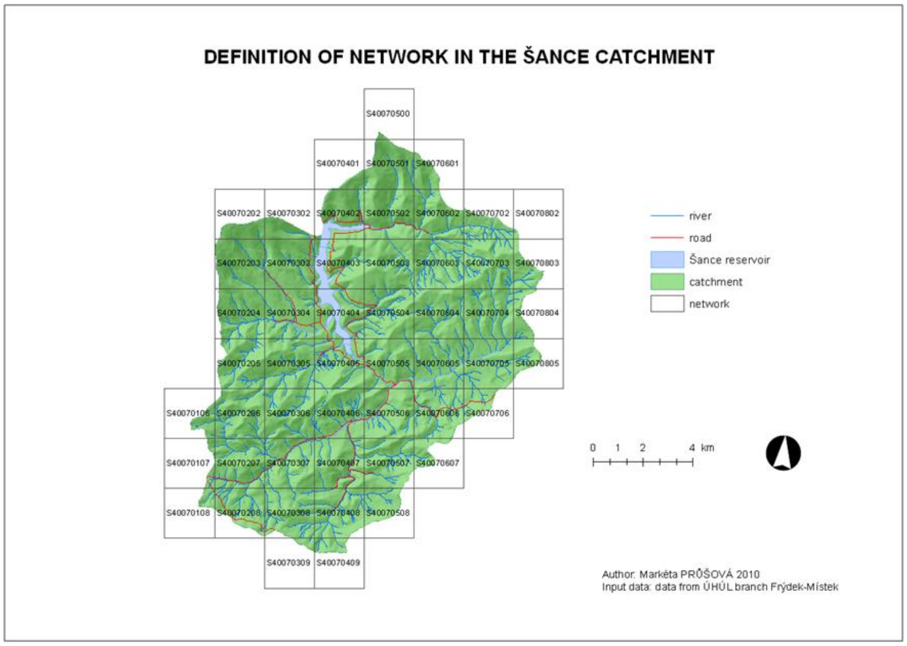

Fig. 1. Definition of network.

IV. EXPLORATORY DATA ANALYSIS

It was decided on the basis of differences in the data (in the homogeneity file), and test consensus of the medians to obtain the corresponding results. It will be necessary to perform a statistical analysis for each area (open, forested) separately. Figure 2 demonstrates the systematic difference between average snow heights for two types of lands. The figure also depicts large differences in snow cover among years. It is therefore advisable to separate the processing of data from different measurement campaigns.

Result of exploratory analysis is the statistical analysis of snow cover parameters. All results are largely influenced by time (a year and campaign measurements). Appraise: the snow covers were abundant in 2006; therefore parameters for each snow have good statistical significance. On the contrary, the snow covers were poor in 2008; therefore statistics of measured data do not have a very high predictive ability. Most data has a moderately leftward distribution. In four cases, the snow density parameter has of rightward distribution. The box plot is part of processing. It shows both the progress of time and also the count of extreme outliers during the campaign. Progress was the highest for height of snow in 2006, and then decreased in 2007. Variability of water parameters is similar to the snow height. It can also be seen at the extreme parameter values, whose occurrence is due to the fact that large amounts of snow and the current weather conditions, that especially the melting and recrystallization processes have a significant effect on the local density of snow.

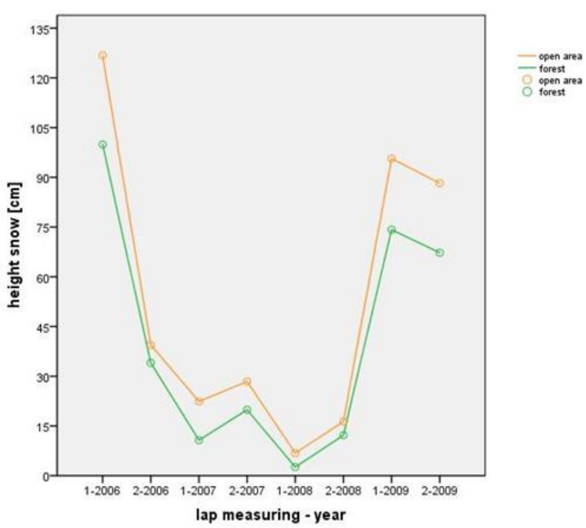

Fig. 2. Average of snow cover in time series.

A further part of EDA investigates normality of data. Most interpolation methods are based on linear estimates and require a normal distribution of sample data. If data fails in normality testing it is necessary to make an appropriate data transformation to reach the normal distribution. The following methods of transformation were tested: natural logarithm transformation, the transformation of the square and power transformation using a linear interpolation coefficient of skewness, which approximates the optimal value of the constant transformation estimate based on linear interpolation [9]. The last method is chosen as the best transformation.

During testing of the parameters of snow, it was found that altitude played the main role. This is particularly the existence of extreme outliers in Lysá Mountain and Smrk Mountain. The terrain factor in comparison with the individual characteristics of snow cover constitutes an important element in studying and evaluating the results.

Furthermore correlation and regression analysis were performed of relations between local morphological characteristics (altitude, slope and orientation) and snow

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-169-4

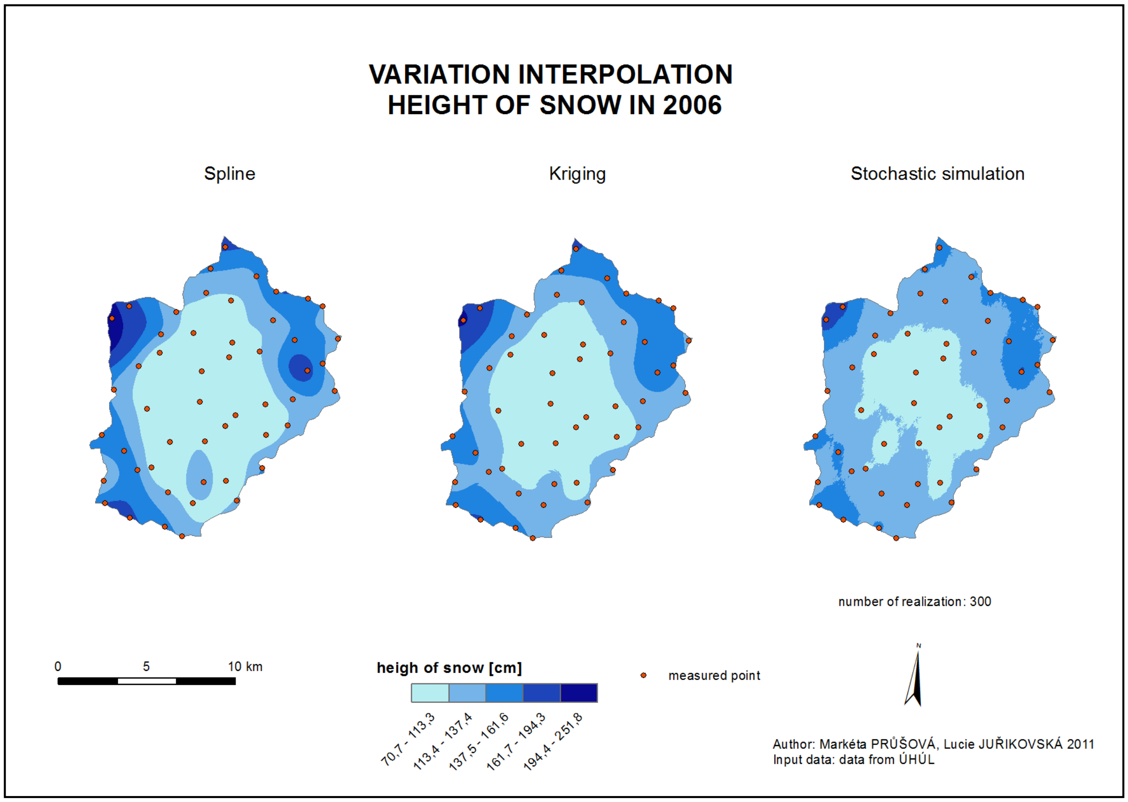

Fig. 3. Interpolation methods.

parameters. The results confirmed a clear dependence of the amount of snow on the altitude, and showed a partial dependence on slope and orientation.

V. Stochastic Simulation

Interpolation is carried out separately for the open and forested areas, according to the results of exploratory statistics. We compared these methods: simple kriging, ordinary kriging, universal kriging, simple cokriging, ordinary cokriging and universal cokriging. Interpolation results visually compared the isolines. We examine development of the shapes of contour lines and their credibility, especially in border areas. Furthermore, the methods explored relationship of known (measured) values in the basin.

We selected the stochastic simulation for height of snow. It was chosen due to its better local estimation ability than classical interpolation methods. This claim can prove the following figure (Fig.3), from which it is evident. Simulation provides better and more exact results than the methods of approximation, which leads to smoothing values even with a known value.

The simulation was carried out with software products such as ArcGIS 10 and SGeMS.

The stochastic simulation was performed with software ArcGIS 10. Gaussian geostatistical simulation was chosen as the stochastic simulation. These parameters (Table I.) configured for simple kriging. Simple kriging is a necessary condition for simulation.

TABLE I. PARAMETERS OF VARIOGRAM

Parameter	Value
Lag	542 m (according to the minimal distance Among locations)
Nugget	0
Number of lags	8
Angle tolerance	22.5 °
Minimal Range	2500
Maximal range	4200
Direction of maximum range	22°
Direction of minimum range	112°

The next step was another parameter directly for simulation.

Settings:

1. Simple kriging
2. Number of realization: 300 (1000)
3. Input feature: snow cover
4. Conditional field: height of snow
5. Cell size: 10

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-169-4

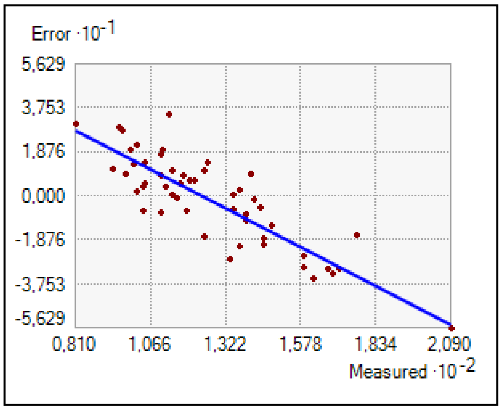

Fig. 4. Error plot.

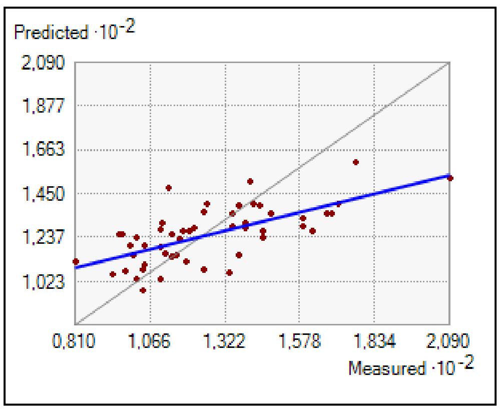

Fig. 5. Graph of prediction.

TABLE II. PREDICTION ERRORS OF SIMPLE KRIGING

Parameter	Value
ME	-1,646
RMS	19,302
ASE	18,961
MS	-0,065
RMSS	1,021

A stochastic simulation result follows in picture (Fig.6).

Of course, the higher number of realizations is simulation results more accuracy and better reflects the trend in the area. The example simulations are compared with the number 300 and 1000 implementation. The simulation with the number 1000 is seen repeating the beginning smoothing interpolation, but in comparison with the spline or kriging method it is a negligible problem.

We solve the stochastic simulation because of its good explanatory power at the measurement point.

Stochastic simulation generally gives better results than the conventional interpolation method. This assertion is based on the comparison of interpolation methods to stochastic simulations.

Stochastic simulation does not produce better results for the determination of the mean on the ground, but it provides the necessary opportunity to determine the probability of exceeding certain limits. These limits are important to the application area (for example, height of snow or water supply, causing a significant increase in crown fractures and fallen trees).

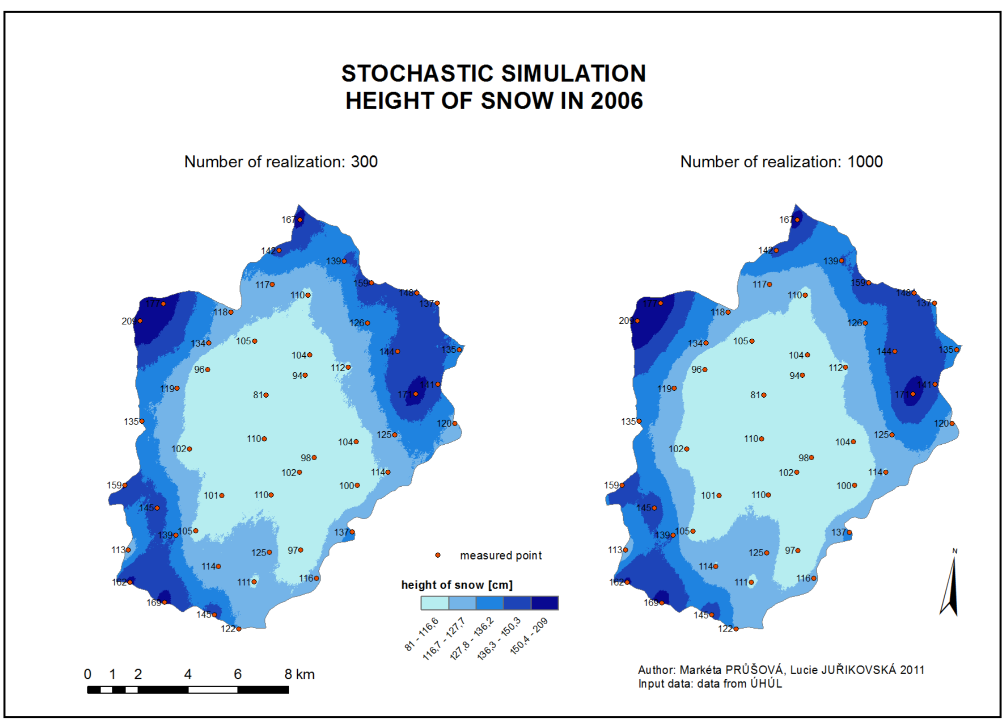

Fig. 6. Stochastic simulation in ArcGIS.

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-169-4

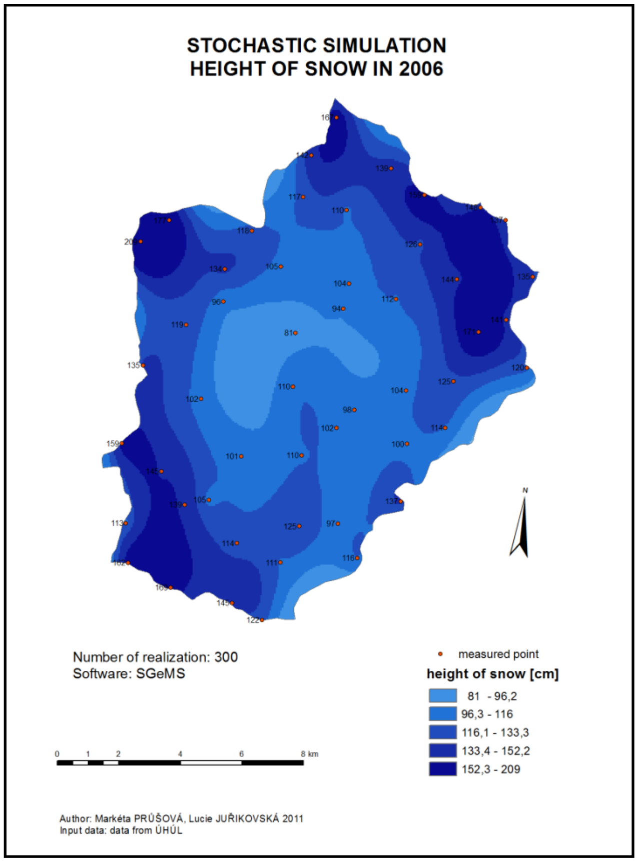

Fig. 7. Stochastic simulation in SGeMS.

VI. Conclusion and Future Work

Majority users use standard interpolation methods and takes at face value and we want to show that there are other methods that can provide better results. Of course, depends on the user which method he chooses.

When processing data, we should not forget the basic statistics of data. Some of these statistics are either completely omitted or performed it without important aspects such as testing the normality of data. This approach can completely distort the results and regardless of the choice of the best interpolation methods.

Stochastic simulation claimed his good properties in compared with others interpolation methods.

Future work will include the following enhancements to our approach: Creation conversion between two different programs for their simulation, finding the best way for compare resultant statistic with other interpolation methods and integrate information about damage of forest.

Acknowledgment

We would like to thank VŠB - Technical University of Ostrava for providing financial resources to this conference.

References

[1] ArcGIS Resources Centres, Key concepts of geostatistical simulation, 2010.

[2] B. Balk and K. Elder, " Combining binary decision and geostatistical methods to estimate snow distribution in a moutain watershed, " Water Resources Research, vol. 36, 2000, pp. 13-26.

[3] N. Cressie, Statistics for spatial data, Wiley, New York, 1991.

[4] C.V. Deutsch and A.G. Journel, GSLIB geostatistical software library and user´s guide, 2nd edition, Oxford University Press, New York, 1998.

[5] K. Elder, J. Dozier and J. Michaelson, "Snow Accumulation and distribution in an alpine watershed," Water Resour. Res., vol. 27, 1991, pp. 1541-1552.

[6] T.A. Erickson, M.W. Williams and A. Winstral, "Persistence of topographic control on the spatial distribution of snow in rugged moutain terrain, Colorado, United States, Water Resour. Res., vol. 41, 2005, doi 10.1029/2003WR002973.

[7] P. Goovaerts, Geostatistics for natural resources evaluation, Oxford University Press, New York, 1997.

[8] E.H.Isaaks and R.M.Srivastava, Applied Geostatistics, Oxford University Press, 1989.

[9] M. Kaňok, Statistical methods in management, Czech Technical University in Prague, 1996.

[10] N.P. Molotch, M.T. Colee, R.C. Bales and J. Dozier, " Estimating the spatial distribution of snow water equivalent in an alpine basing using binary regression tree models: the impact of digital elevation data and idependent variable selection, " Hydrological Processes, vol. 19, 2005, pp. 1459-1479.

[11] E. Pardo-Iguzquiza and M. Chica-Olmo, "Geostatistical simulation when the number of experimental data is small: an alternative paradigm," Stoch. Environ. Res. Risk. Assess, vol.22, Apr.2008, pp. 325-337, doi 10.1007/s00477-007-0118-1.

[12] A. Rango, "Spaceborne remote sensing for snow hydrology applications," Hydrological Sciences Journal, vol. 41, 1996, pp. 477-494.

[13] D.G. Tarboton, G. Bloschl, K. Cooley, R. Kirnbauer and C. Luce, Spatial snow cover processes at Kűhtai and Reynolds Creek, Spatial Patterns in Catchment Hydrology: Observation and modelling, Cambridge University Press, 2000, pp.158-186.

[14] R. Webster and M.A. Oliver, Geostatistics for environmental scientists, 2nd edition, John Wiley & Sons, 2007.