Geostatistical Computing in PSInSAR Data Analysis

Andrzej Lesniak and Stanislawa Porzycka

AGH University of Science and Technology Krakow, Poland
lesniak@agh.edu.pl, porzycka@agh.edu.pl

Abstract. The presented paper describes the geostatistical analysis of PSInSAR data. This analysis was preceded by short description of PSInSAR technique. The geostatistical computations showed in this article were performed with the *R* open-source software containing the *gstat* package. The analysis contains variograms computing (directional variograms) and ordinary kriging interpolation. The computationally costly problems in geostatistical analysis of PSInSAR data were discussed.

Keywords: geostatistics, ground deformations, interferometry, kriging, PSInSAR.

1 Introduction

The computations in geosciences frequently require working with different types of data. Nowadays the satellite data are used very often. Among them we can distinguish radar and multispectral images. Both types derive information for each pixel of the images. Quite often, geoscientists have to analyze these remote sensing data in relation to ground measurements, which are usually performed for irregular and rare grid due to economical and physical constraints. Comparing large remotely sensed data with ground measurements is not only an extremely important task, but also a very computationally costly challenge. An example of a computationally intensive processing component in geographic data analysis is kriging, which is a geostatistical method of interpolation.

This paper presents the analysis of PSInSAR data. PSInSAR method is dynamically developed branch of satellite radar interferometry. Nowadays many institutions have implemented the PSInSAR technique to monitor ground deformations. This technique cannot constitute an independent tool to study ground displacements and PSInSAR data have to be joined with other measurements. The analysis of PSInSAR data, which are huge data sets, meets several computational problems. In this paper the analysis of PSInSAR data was performed with the use of geostatistical methods. The main goal of this analysis was to interpolate values of ground displacements at unmeasured locations to produce maps, which can be easily analyzed together with satellite images or with interferograms. Performed interpolation gives us the opportunity to do a useful comparison between different datasets. This work points the main computationally costly problems in geostatistical analysis of PSInSAR data.

G. Allen et al. (Eds.): ICCS 2009, Part I, LNCS 5544, pp. 397–405, 2009.

2 PSInSAR Data Set

Scientists from the Politecnico di Milano (POLIMI) elaborated the PSInSAR (Permanent Scatterer Interferometry SAR) technique in the nineties of the 20th century. This method exploits sets of dozens SAR images in order to detect small (not bigger than several centimeters per year), long period ground deformations [1]. SAR (Synthetic Aperture Radar) is a form of radar mounted on the satellite. It emits its own microwave radiation towards the surface of the Earth and records the amplitude and phase of the signal, which returns to the radar antenna (each pixel of the radar image contains information about the phase and amplitude of backscattered signal). PSInSAR technique derives information about ground deformations only for PS points. PS points are stable radar targets, which means that PS points have time stable amplitude and phase in all exploited radar images. These stable radar targets correspond very well with man-made features on the ground, such as buildings, bridges, viaducts and etc., therefore the density of PS points is much higher in urban areas (even more than 1000 PS/km^2) than it is in rural areas. PSInSAR method provides information about ground displacements for large areas of interest, even exceeding 10 000 km^2. The spacing of PS points is usually very irregular. PSInSAR technique uses archival images, dating back to 1992, and giving us the opportunity to reconstruct previous ground deformations. This method enables to detect displacements with average annual rate equal to 0.1 mm/yr. Despite the fact that PSInSAR technique cannot be an independent tool for ground movements monitoring, it complements considerably the conventional leveling and GPS surveying.

PSInSAR data, which have been presented in this paper, describe small, long-lasting ground displacements, which occurred in the Upper Silesian Coal Basin (Southern Poland) in the years between 1992 and 2003. In this region the intensive coal exploitation has been carried on for more than two hundred years. This exploitation and complicated geological structure (a lot of faults) makes this area particularly endangered with terrain deformations. PSInSAR data for Upper Silesian Coal Basin were obtained as a result of 79 SAR images processing. These radar images were performed by ESA's satellites (ERS-1, ERS-2 and ENVISAT). In the studied region, which covers more than 1200 km^2, about 120 000 PS points were identified (Fig. 1). For each of them the average annual motion rate (mm/year) and value of coherence were calculated. For 30 000 of these PS points the values of monthly, relative ground deformation were also determined. Locations of PS points correspond very well with the land development. In this region there are also areas without PS points. These areas represent agricultural regions, forests and areas with strong ground displacements caused directly by mining activity (in this last case areas without PS points are located usually exactly above exploitation parcels) [2]. For the Upper Silesian Coal Basin the subsidence phenomenon is characteristic. The values of average annual motion rates in this region range from -39 mm/yr to 25 mm/yr. In order to explain the origin and mechanism of ground deformations in this studied region the PSInSAR data have to be analyzed together with geological, hydrogeological and mining data. This analysis has to be preceded by exploratory PSInSAR data study and interpolation of displacements at unobserved locations.

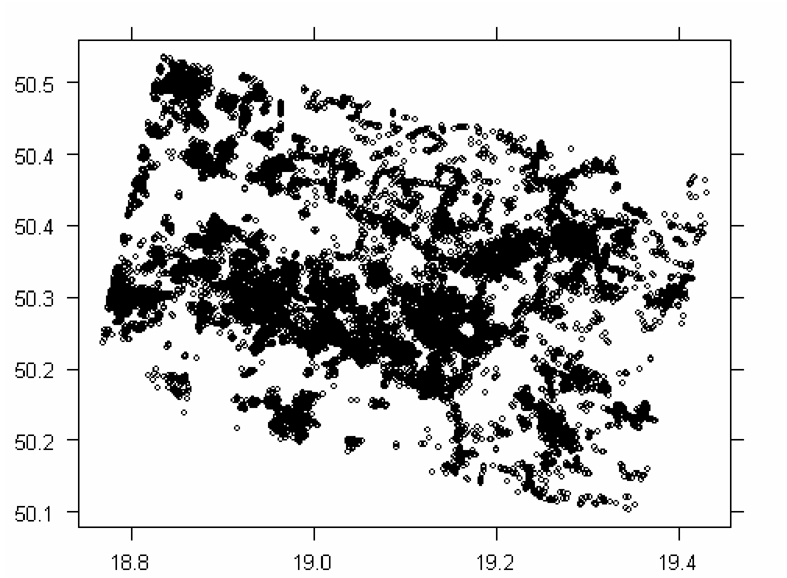

Fig. 1. Location of 30 000 PS points in the area of Upper Silesian Coal Basin (southern Poland)

3 Geostatistical Analysis of PSInSAR Data

The PSInSAR data analysis was performed with the use of geostatistical methods. Geostatistics is a subset of statistics specializing in analysis and interpretations of geographically referenced data [3]. In geostatistics a spatial autocorrelation among sample data is described. This autocorrelation is modeled by a semivariogram, which plots the semivariance as a function of distance. The semivariogram (empirical, experimental) can be estimated from *Nh* sample data pairs *z(si)* and $z(s_i+h)$ for a number of distances (or distance intervals) h_j by Eq.(1).

$$\gamma(\tilde{h}_j) = \frac{1}{2N_h}\sum_{i=1}^{N_h}(z(s_i) - z(s_i + h))^2, \forall h \in \tilde{h}_j \tag{1}$$

The semivariograms provide insight into the spatial structure of a random process. One of the main goals of geostatistics is to predict values of variable at unobserved locations (in space or in time). Kriging is the geostatistical method of prediction. It is based on the theory of the regionalized variables [4]. Interpolation of value of variable at an unmeasured location is based on observations of its value at nearby locations. A standard version of kriging is called ordinary kriging. The predictions are made as in Eq.(2).

$$\hat{z}_{OK} = \sum_{i=1}^{n} w_i(s_0)z(s_i) = \lambda_0^T z \tag{2}$$

where the λ_0^T is a vector of kriging weights (*wi*) and *z* is the vector of *n* observations of primary locations. The values of kriging weights should reflect the true spatial autocorrelation structure and they are given by ordinary kriging equation system Eq.(3):

$$\begin{bmatrix} w_0(s_0) \\ \vdots \\ w_n(s_0) \\ \varphi \end{bmatrix} = \begin{bmatrix} \gamma(s_1,s_1) & \cdots & \gamma(s_1,s_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(s_n,s_1) & \cdots & \gamma(s_n,s_n) & 1 \\ 1 & 1 & \cdots & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma(s_0,s_1) \\ \vdots \\ \gamma(s_0,s_n) \\ 1 \end{bmatrix} \quad (3)$$

where φ is called *Langrange* multiplier. In addition to the estimation we can also calculate the prediction variance (variance of the prediction error) Eq.(4):

$$\sigma_{OK}^2 = C - \sum_{i=1}^{n} w_i(s_0)C(s_0,s_i) + \varphi \quad (4)$$

where $C(s_0,s_i)$ is the covariance between the new location and the sampled point pair and *C* is a sill (upper bound) of semivariogram. Values of prediction variance derive information about quality of used kriging model.

Geostatistical analysis of PSInSAR data for the area of Upper Silesian Coal Basin was performed using *R* with *gstat* package. *R* is a language and environment for statistical computing and graphics. It provides methods for advanced statistical analysis. *Gstat* is a package, which derives functions for geostatistical analysis. The PSInSAR data analysis was done for 1240 randomly selected PS points and includes four main steps (excluding the exploratory data analysis): semivariogram estimation, kriging, calculation of prediction variance and cross-validation. All calculations were performed on a PC.

In the first part of the analysis the values of empirical (experimental) semivariograms were computed. In order to check how the data's variation depends on the relative orientation of data locations the four directional semivariograms were calculated. They were performed for directions: 0, 45, 90 and 135 (where 0 is North and 90 is East) (Fig. 2). In the next step of analysis the isotropic semivariogram was created (Fig. 3). The obtained isotropic semivariogram has a shape suggesting a spherical model. This model was adjusted by weighted least-squares (Fig. 3). In case of selected PSInSAR data the distance at which the semivariogram reaches the sill (range) is equal 12.2 km and beyond this range no correlation exists between two values of ground displacements at PS points.

The isotropic semivariogram estimation is a computationally time-consuming task but it has to be executed only once per data set. The same situation holds also for semivariogram theoretical model fitting. The semivariograms calculation can cause more problems in case of anisotropy that can be modeled by defining range ellipse [5]. In this case several directional semivariograms have to be calculated and then for each of them the theoretical models have to be fitted. Figure 4 shows the relation between numbers of used PS points and semivariogram computational times. The relation is distinctly non-linear. In case of large PSInSAR dataset necessity of calculation more than one semivariogram causes meaningful increase of computational time.

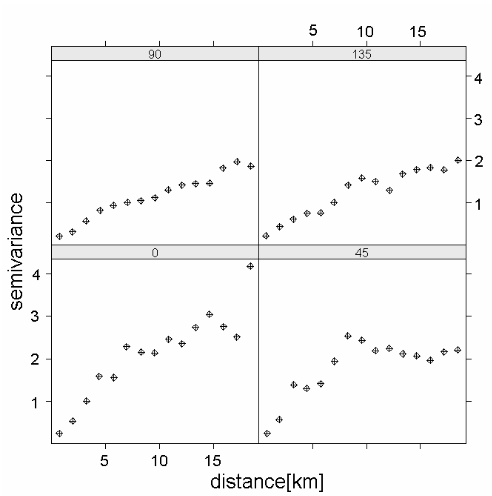

Fig. 2. Directional semivariograms for four directions (0 is North, 90 is East)

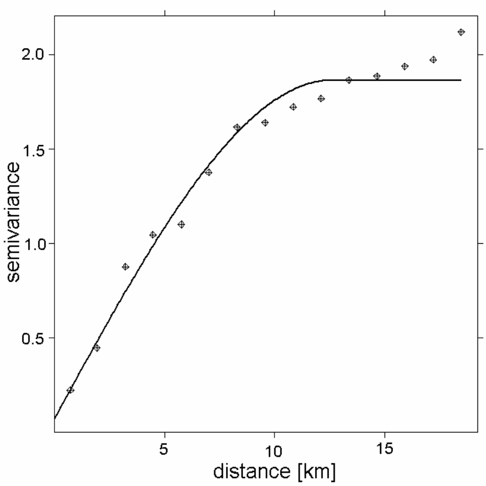

Fig. 3. Experimental isotropic semivariogram with fitted theoretical model

The goal of the next step in the geostatistical analysis of PSInSAR data was to predict values of ground deformations at unmeasured locations. This task was performed using ordinary kriging method. For PSInSAR data it is important to predict the values of variable for very dense grid. It is essential when the results of interpolation are used to study the stability of individual buildings. In case of PSInSAR data, kriging computations are hindered because of location of PS points, which is very irregular. In this work the values of subsidence were predicted for the grid with only 20000 nodes. The results of kriging are presented in the Fig. 5.

Kriging is an example of computationally intensive method because it requires the solution of a large linear system for each grid node. In the case of large datasets analysis (like PSInSAR dataset) kriging it is too computationally demanding to be run on a PC or low performance computing platform. The kriging produces the best results when the largest possible number of known points is used to predict values in no measured location [6]. This is the most expensive option. In case of PSInSAR data analysis even several thousand points can be used to estimate the variable for one grid node.

The computational time of kriging proceeding also increases when the interpolation is done for the dense grid. In this work the computational times for kriging proceeding were measured in the *R*. In the first case the elapsed time was measured in relation to the number of grid nodes where the values of ground deformations were predicted. In this case the number of PSInSAR data was constant

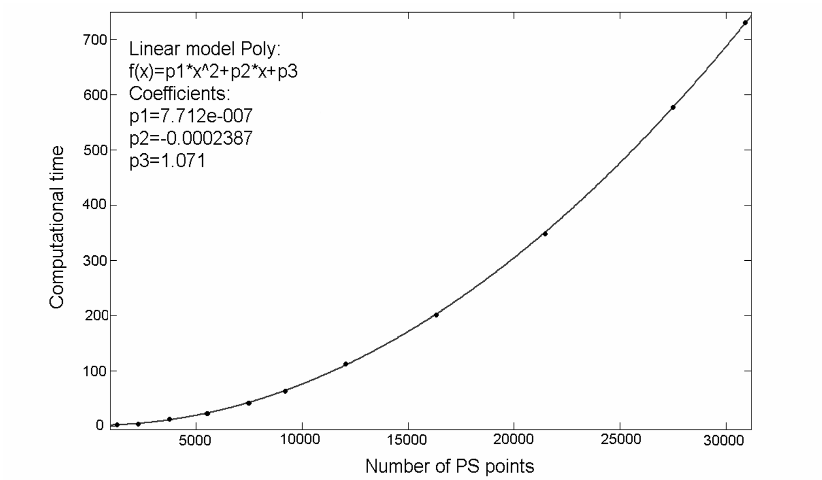

Fig. 4. Timing results for isotropic semivariogram algorithm for different numbers of PS points

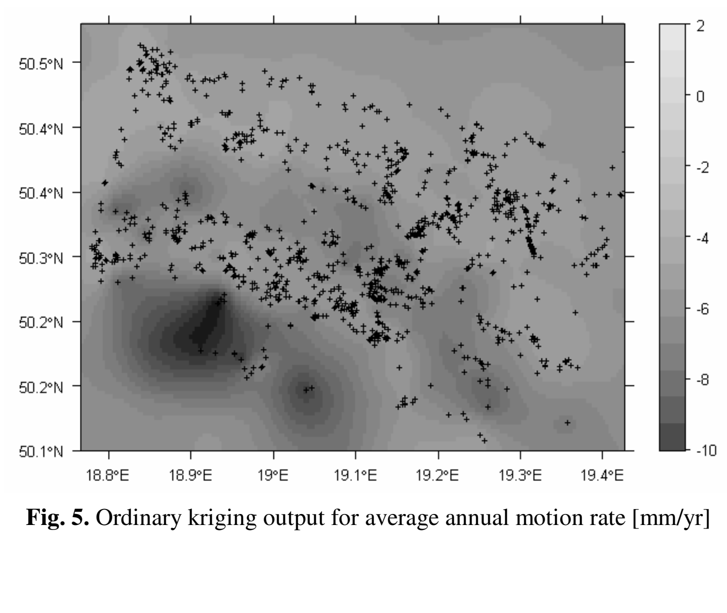

Fig. 5. Ordinary kriging output for average annual motion rate [mm/yr]

and equal to 1240 PS points. For the studied area five regular grids were prepared. The numbers of nods of these grids were corresponding to: 1250, 5000, 20000 and 80000 nodes. As it can be seen in the Fig. 6 the relation between the computational time and the number of grid nodes is linear e.g. double increase of grid nodes causes double increase of kriging computational time.

In the second case the number of grid nodes was constant and equal to 5000 nodes. In this part of work the relation between kriging computational times and numbers of PS points (used to interpolate values at grid nodes) was studied. It should be underlined that the spacing of PS points is irregular and the subsets of PS points were selected randomly. The number of PS points was changed from 83 to 2395 points. In this case the relation between the computational time and the number of PS points is non-linear (Fig. 7). It can be evaluated that for 120 000 PS points the computational time equals about 5 days and 13 hours. In order to reduce this time the kriging algorithm can be performed in parallel environment [7].

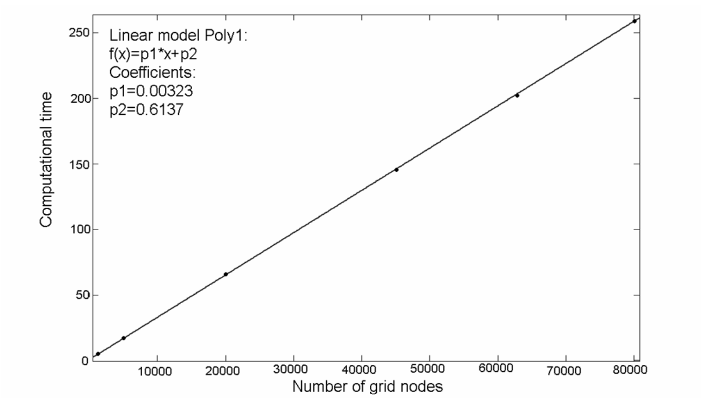

Fig. 6. Timing results for kriging algorithm for different numbers of grid nodes

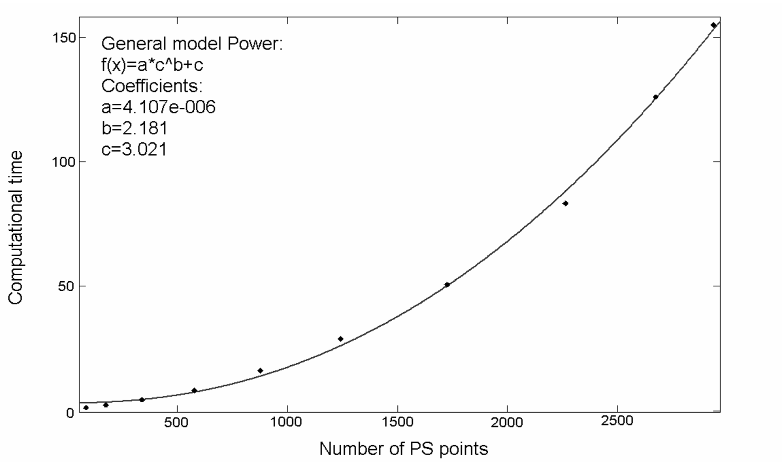

Fig. 7. Timing results for kriging algorithm for different numbers of PS points

In the third part of analysis the map of prediction variance was performed (Fig. 8). The ability of kriging to produce this kind of map is what separates it from other spatial interpolation methods.

In the last part of the geostatistical analysis the leave-one-out cross validation method was used to pinpoint the most problematic PS points. This algorithm of leave-one-out cross validation is a very computationally costly procedure. In this method the value of variable for each individual point is assessed against the whole data set. Each data point is visited and the prediction is done with kriging method, but without using the observed value. Fig. 9 presents the timing results for leave-one-out cross validation algorithm for different numbers of PS points. The number of PS points was changed from 83 to 1241 points. The relation between the computational time and the number of PS points is non-linear.

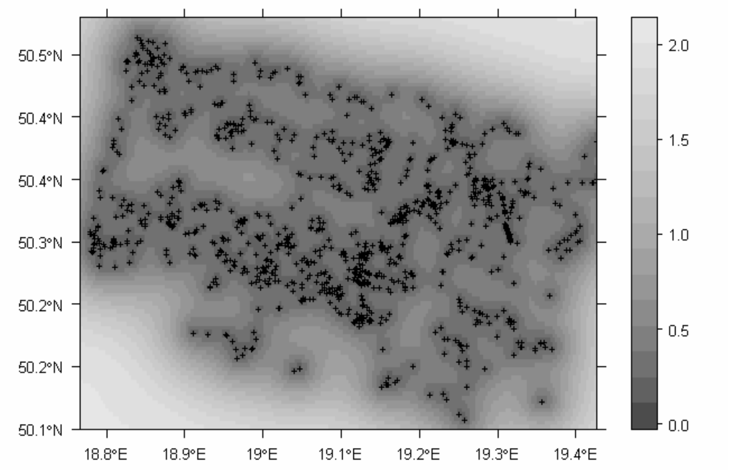

Fig. 8. Map of prediction variance

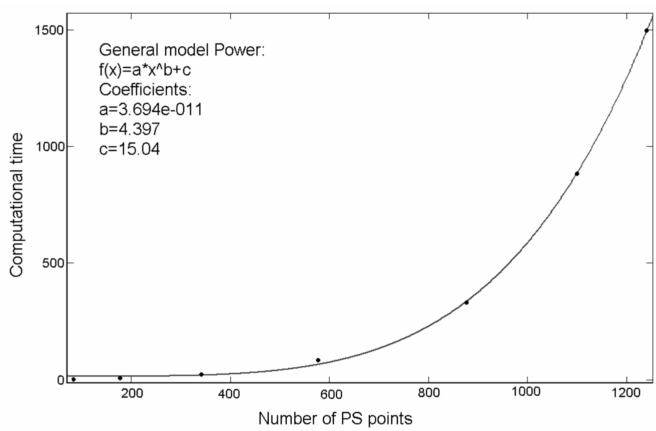

Fig. 9. Timing results for cross validation algorithm for different numbers of PS points

4 Conclusions

The geostatistical analysis of PSInSAR data gives good results, yet it is a very computationally costly procedure. In this work the analysis was performed only for 1240 PS points selected randomly from the dataset which includes 120 000 PS points. Kriging is the most computationally costly task in geostatistical analysis of ground deformation. This task is also crucial because kriging results constitute the base for the geological interpretations. The maps of kriging output and prediction variance obtained in this work and complemented by different kind of data (geological, hydrogeological, mining) enable to determine the relations between values of subsidence and mining activity and geological structure of studied region. The interpolation of ground displacements for very dense grid give us also opportunity to monitor stability of individual objects on the ground e.g. buildings.

For small datasets the geostatistical analysis can be run on a PC but in case of the whole PSInSAR dataset it is necessary to use high performance computing platform or distributed architectures. In order to perform the geostatistical analysis for all 120 000 PS points the parallel kriging algorithm has to be prepared. Designing this algorithm the very strong irregularity in data locations has to be taken into consideration.

References

1. Ferretti, A., Prati, C., Rocca, F.: Permanent Scatterers in SAR Interferometry. IEEE Transaction on Geoscience and Remote Sensing 39(1), 8–20 (2001)
2. Lesniak, A., Porzycka, S.: Environment monitoring using satellite radar interferometry technique (PSInSAR). Polish Journal of Environmental Studies 17(3A), 382–387 (2008)
3. Goovaerts, P.: Geostatistics for Natural Resources Evaluation. Oxford University Press, New York (1997)
4. Weckernagel, H.: Multivariate Geoststistics. Springer, Heidelberg (1995)
5. Bivand, R.S., Pebesma, E.J., Gomez-Rubio, V.: Applied Spatial Data Analysis with R. Springer, New York (2008)
6. Border, S.: The use of indicator kriging as a biased estimator to discriminate between ore and waste. Applications of Computers in the Mineral Industry, University of Wollongong, N.S.W (1993)
7. Kerry, K.E., Hawick, K.A.: Kriging Interpolation on High-Performance Computers. In: Bubak, M., Hertzberger, B., Sloot, P.M.A. (eds.) HPCN-Europe 1998. LNCS, vol. 1401, pp. 429–438. Springer, Heidelberg (1998)