Analysis of ground deformations based on parallel geostatistical computations of PSInSAR data

Jacek Strzelczyk, Stanislawa Porzycka, Andrzej Lesniak
Department of Geoinformatics and Applied Computer Science
AGH University of Science and Technology
Krakow, Poland
jacek.strzelczyk@agh.edu.pl, porzycka@agh.edu.pl

Abstract— **The ground deformations can be caused by many different factors. One of them is coal exploitation. It can cause not only strong and abrupt ground movements but also small, long period deformations which can occur even several years after exploitation is finished. In both cases these displacements are serious menace for surface and subsurface infrastructure. In this work the PSInSAR data were used to study small, long lasting ground deformations which occurred in the Upper Silesian Coal Basin (south Poland) in the years 1992-2003. In this region the intensive coal exploitation has been carried on. Additionally this area has complicated geological structure. These factors make studied region particularly threatened with terrain deformations. The subsidence phenomenon was detected in this region using PSInSAR method. PSInSAR technique is a dynamically developed branch of satellite radar interferometry. It exploits a set of dozens satellite SAR images in order to detect small ground deformations for large areas. PSInSAR technique derives information only about ground movements for stable radar targets, so called PS points. They correspond with man-made features on the ground like buildings, bridges etc. In the Upper Silesian Coal Basin the values of ground deformations were measured for more then 120000 PS points. Their location is very irregular. In order to study the origin of this ground displacements the values of deformations have to be estimated in no-observed locations. In this work this estimation was performed using kriging which is a geostatistical method of interpolation. It is very computationally expensive method because it requires the solution of a large linear system for each grid node in which the interpolation is done. Because the PSInSAR dataset is very large the kriging computations were done in parallel environment. In the designing of the parallel kriging algorithm also strong irregularity of PS points location was taken into consideration. At the beginning of the formulated parallel kriging algorithm, the number of PS points used in grid nodes interpolation was determined. Then, based on this information, data decomposition was conducted. In this process, grid points were divided between CPUs, so that each processor had the same number of calculations to perform. This solution is optimized for PSInSAR data points, which are deployed irregularly, often in local centers on the research area. It was shown in this work that geostatistical methods can be successfully used in PSInSAR data analysis, but are computationally expensive. By designing parallel kriging algorithm for PSInSAR data and using Blade infrastructure for computations, it was possible to conduct the interpolation of ground deformations in viable time. It is worth to underline the universality of this algorithm, which can be used not only for PSInSAR data, but also for other types of data with similar characteristic.**

PSInSAR; geostatistics; parallel computing

I. INTRODUCTION

Recent development of remote sensing techniques brings new possibilities of precise ground movement measurements. One of these possibilities is PSInSAR method. It allows to detect small and long period ground deformations on large areas. Despite its advantages this technique can not be the only tool to monitor unstable regions. It provides information about displacements only in *Permanent Scatterer* (*PS*) points. To analyse the whole area, deformations in non observed locations of a region has to be estimated. It can be done by geostatistical interpolation method called *kriging*. Kriging procedure can be very computationally complex and resource demanding. Estimation of displacements in a big set of points which were not measured by PSInSAR technique is very time consuming and require great amount of memory.

In this work, geostatistical analysis of PSInSAR data from north-eastern part of Upper Silesian Coal Basin in south Poland was conducted. Ordinary kriging method was used to interpolate ground displacement values. Calculations were performed in distributed computing environment. Special, parallel version of kriging algorithm was created to distribute the calculations on many processors of different computers.

This solution allowed to perform an analysis of large set of PSInSAR data. Results will help in a better and more accurate interpretation of ground deformations.

II. DESCRIPTION OF PSINSAR TECHNIQUE

PSInSAR (Permanent Scatterers Interferometry Synthetic Aperture Radar) is an advanced, dynamically developed branch of satellite radar interferometry. It was evolved by scientists from the Politecnico di Milano (POLIMI) in 1999 [1] and it is based on multi-interferogram approach. PSInSAR technique exploits set of dozen (minimum 15 to 20) satellite SAR images performed at different times in order to measure mm-scale, long period ground deformations for stable radar targets called PS points. These points can be identified on satellite images on the basis of dispersion of amplitudes of reflected radar signal recorded by SAR antenna [1]. The PS points have stable in time amplitude and phase of backscattered radar signal. PS points coincide with objects on the ground like buildings, metallic structures, bridges, viaducts, rock outcrops etc. The spacing of PS points is usually very irregular. PSInSAR technique derives information only about small ground displacements, not larger than several centimeters per year. This limitation is connected with length of radar waves used by SAR system. The maximum deformation rate that can be measured is a half of a wavelength per revisit period [2]. In the case of ESA's satellites ERS-1, ERS-2 and ENVISAT the critical deformation rate is 2.8 cm in 35 days, or 0.8 mm in a day. Accuracy of PSInSAR technique depends on the number of radar images used in processing. In order to obtain good results about 20 images are required as a minimum. Using PSInSAR technique we can measure average displacements rate with precision between 0.1 and 0.5 mm and single ground movements with precision about 2 mm. Permanent Scatterer interferometry methods uses archival radar images (dating back to 1992) giving us possibility to reconstruct the dynamics of deformations during the space of last several years. PSInSAR technique is a useful tool to monitor the ground deformations for urban areas where a lot of PS points (even more than 100 PS/ km^2) can be identified. In rural areas ground displacements monitoring performed by described technique does not give good results. To apply PSInSAR technique, at least 5 PS/km^2 have to be identified. It is also necessary to remove atmospheric effect from data.

The PSInSAR technique is used for the monitoring of subsidence or uplift, caused whether by natural failure or man-made activities. It has been applied to measure displacements connected with landslides [3], coal exploitation [4], tectonic movements [5] or water drainage [6]. PSInSAR technique can be also used to study the stability of individual buildings what is very important in urban areas monitoring.

Result of images processing using PSInSAR technique is a map of PS points (like network of GPS stations), for which the ground deformations were measured with monthly (or higher) frequency. In order to study the stability of whole area it is necessary to interpolate the values of ground displacements in points which are not stable radar targets. The values of ground movements are usually correlated in space and time therefore this interpolation can be done using geostatistical methods [2].

III. GEOSTATISTICS

Geostatistics is a branch of applied mathematics used in many fields of science and industry where a need for evaluating spatially or temporally correlated data occurs [7]. The basic function in geostatistics is semivariogram, which represents semivariance between particular parameter as a function of distance. The experimental semivariograms can be estimated by (1) where N_h is a number of sample data pairs $z(s_i)$ and the $z(s_i+h)$ that can be linked by a vector h belonging to h_j.

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

The experimental semivariogram has to be fitted by one of the theoretical models (e.g. spherical, gaussian or linear). Choice of appropriate model and its matching to the empirical semivariogram have significant influence on the correctness and accuracy of the forward interpolation results. The most common used model is a spherical model, which can be described by (2)

$$\gamma(h) = \begin{cases} C(\frac{3h}{2a} - \frac{h^3}{2a^3}) + C_0, 0 \le h \le a \\ C + C_0, h > a \end{cases} \qquad (2)$$

where a is a range of influence (beyond this range no correlation exists between two values of parameter), $C0$ is a nugget effect (represents variations for pairwise distances near zero) and C is a partial sill of semivariogram model.

The geostatistical method of interpolation is kriging. It can be used to estimate values on a regular grid using irregularly spaced data [7]. The most widely used type of kriging is an ordinary kriging (3) in which the predictions are made from N neighbouring sample points of primary locations $z(s_i)$.

$$\hat{z}_{OK} = \sum_{i=1}^{n} w_i(s_0) z(s_i) \qquad (3)$$

The values of kriging weights are given by ordinary kriging equation system (4) where $\gamma(s_i,s_j)$ is a semivariance between data points s_i and s_j, and $\gamma(s_0,s_j)$ is a semivariance between s_i data point and estimation point s_0 and φ is the Lagrange parameter.

$$\begin{bmatrix} w_0(s_0) \\ \vdots \\ w_n(s_0) \\ \varphi \end{bmatrix} = \begin{bmatrix} \gamma(s_1,s_1) & \cdots & \gamma(s_1,s_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(s_n,s_1) & \cdots & \gamma(s_n,s_n) & 1 \\ 1 & 1 & \cdots & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma(s_0,s_1) \\ \vdots \\ \gamma(s_0,s_n) \\ 1 \end{bmatrix} \qquad (4)$$

In many cases, especially in work with large data sets, we do not have to involve all data into estimation procedure [7]. In practice, we draw the circle or ellipse around given estimation point and for interpolation we take only data points that are located within this area (neighborhood). There are no strict rules to determine the size of this neighborhood. One method in common use to perform this task is to check the values of kriging variance when we enlarge the neighborhood [7]. If we do not see the increase in the precision of estimation we do not have to increase the neighborhood.

In addition to the interpolation we can calculate the estimation variance of ordinary kriging by (4), which allows us to evaluate the precision of estimation [7].

$$\hat{\sigma}^2_{OK} = 2\sum_{i=1}^{n} w_i(s_0)\gamma(s_0,s_i) - \varphi - \gamma(s_0,s_0) \qquad (5)$$

Kriging is not a fast method for spatial interpolation because it is very computationally expensive. In case of PSInSAR data set, which can include more than several hundred thousand PS points, to perform interpolation using kriging method, it is necessary to use high performance platform or distributed architectures.

IV. ANALYSIS OF PSINSAR DATA

In this work the analysis was performed for PSInSAR data that describe the ground deformation which occurred in the Upper Silesian Coal Basin (south Poland) in years between 1992 and 2003. It is industrial region where the coal exploitation has been conducted for more than two hundred years. Despite the fact that a lot of coal mines have finished exploitation in many of them the drainage is still performed. Additionally the area of study has complicated geological structure. It is crossed by many faults with main direction form NW to SE and with vertical displacements range from tens to several hundred meters. These factors make studied region particularly threatened with terrain deformations. It needs to be stressed that area of study is in larger part covered by land development therefore the ground deformation in this region has to be monitored permanently and with high accuracy.

The PSInSAR data used in this work were obtained as a result of radar image processing performed from ESA's satellites (ERS-1, ERS-2, ENVISAT). In the studied region, which covers approximately 1200 km^2, about 120 000 PS points were identified. Their spacing is very irregular as showed on Fig. 1. There are places without PS points which correspond to rural areas, forest, rivers, and lakes and etc. or to regions for which the values of ground deformations are too high to be detected by PSInSAR technique. For studied region a subsidence phenomenon is characteristic. The average annual motion rate measured for PS points in this area is equal to -1 mm/yr and the maximum rate of subsidence is equal to -39 mm/yr. Only for several PS points the lift movements were detected.

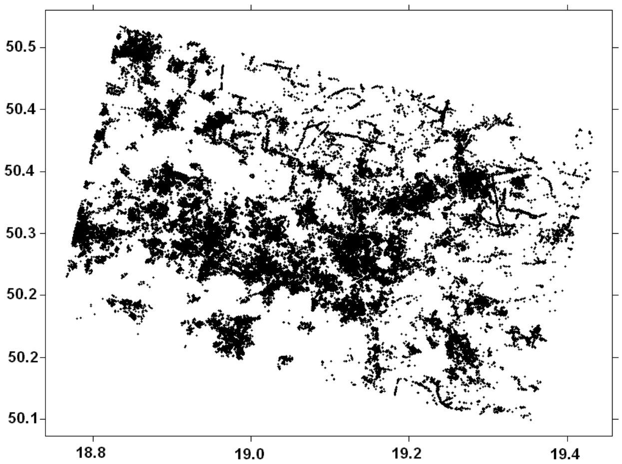

Figure 1. Example of a figure caption. (figure caption)

In order to study the stability of whole area it is necessary to interpolate the values of ground deformations in unobserved locations. In this work this estimation was performed using ordinary kriging method. In the first step of analysis the experimental semivariogram for all PSInSAR data was calculate. It was fitted by spherical model with parameters: a=10 km, C=1.7 and C_0=0.2. These calculations were performed using *R* open-source software with *gstat* package.

In case of PSInSAR data analysis the estimation of values of ground displacements has to be performed for very dense grid and with very high precision. In the kriging method it is necessary to solve the large linear equation system for each grid node. It causes that for large PSInSAR data sets kriging is too computationally expensive to be executed on a single PC. There is not enough memory to perform calculations on one computer using available software.

In this work the interpolation of the values of ground deformations based on whole available PSInSAR data set was performed in a parallel environment.

V. PROPOSED PARALLEL SOLUTION

A. *Parallel Computing*

Parallel computing is a form of computation in which many calculations are carried out simultaneously [8]. This form can significantly decrease program execution time. It is done by dividing a problem into smaller parts and solving them on different processors concurrently. Parallel algorithms run in distributed environment and often make use of a high performance machines designed to execute calculations simultaneously. A distributed system is a system in which components located at networked computers communicate and coordinate their actions only by passing messages [9]. Example of a high performance machine that can constitute a distributed environment for parallel applications is IBM Blade architecture used in this research.

Parallel computing is applied when serial algorithms cannot solve a problem due to too complicated and long lasting computations or memory constraints. Such problem can be decomposed by two procedures: domain decomposition and algorithm decomposition. Domain decomposition refers to divide and conquer techniques for solving partial differential equations. In this method the same problem defined on smaller subdomains is solved iteratively on different computational units [10]. It is used when the set of data is too big to be computed on one machine, or the calculations would last too long. In algorithm decomposition, an algorithm is divided into separate parts and executed simultaneously on different processors. Each part has limited information about what other parts are doing and communicates with them just to exchange data.

Both decompositions divide problems into separate parts that are executed in a parallel environment. Each part is executed on different computing node, different processor or different processors' core. All processes communicates and cooperates with each other in a task. One of them, called *master node* is an overriding process that controls *slave nodes* - other processes in this task. In the domain decomposition, this is usually the one that reads the input data, scatters it between slaves and gathers the results at the end.

B. Implementation

There are some parallel kriging algorithms already available. All of them base on domain decomposition but they do not consider points deployment in space. Many authors propose to decompose domain along either rows or columns ([11], [12]). In case of PSInSAR data simple dividing by rows or columns would not be effective. PS points are unevenly distributed, so decomposition by rows or columns would give unequal portions of data. That is why new parallel kriging algorithm for PSInSAR data was developed in this work.

Geostatistical computations conducted on 120,000 PSInSAR points need great amount of memory. Kriging procedure can be divided with the use of data decomposition. The examined area of Upper Silesian Coal Basin was covered by a regular grid of points. The size of a grid cell was relatively small. The aim of the algorithm was to calculate ground deformation velocity in each point. This procedure is based on the values of velocity in the PS points which are within the given radius.

First step of the algorithm was to read the PS points as an input data and send them to all nodes. After that, master node created a rectangular grid that covered the whole examined area and cut it down to the smaller rim that still covers all PS points. This procedure chose the points where the kriging must be done. Knowing which grid points to consider, master node calculated number of neighbouring PS points in given radius for each grid point. The bigger number of neighbours, the longer calculations needs to be perform. To divide computations equally between processors grid points were grouped into classes with similar number of neighbours. Complexity of computations in each grid point depends on the number of neighbouring PS points. The more points, the bigger set of equations and the longer computation time. To divide calculations equally each class was scattered between slave nodes. In result, each processor had almost the same number of grid points to compute with similar amount of neighbours.

This approach to domain decomposition makes kriging algorithm for PSInSAR data reasonable, since PS points are usually unevenly distributed on the surface. With the normal domain decomposition, the portions of computations for each processor would be different and the algorithm would be inefficient. At the end of the procedure, the results from all computer nodes were gathered and saved by the master node. They contain values of latitude, longitude and calculated velocity in each grid point.

The most resource demanding part of the algorithm was solving a set of equations for each grid point. It was a big, dense and unsymmetrical set of equations without a dominant diagonal. The size of a set depended on the number of neighbours. Different algorithms were tested, but only one of them solved this type of set of equations. Chosen algorithm - Gauss elimination with pivoting - had complexity of n^3 since it requires approximately $2n^3/3$ operations.

Parallel environment used in this experiment consisted of three IBM Blade Servers with Linux Fedora Core 10 system. Each Blade Server had 16 GB of memory and 4 processors with two cores. This gives 8 virtual processors on each server and 24 total. Communication between processes was provided by Message Passing Interface (MPI). It is a library specification for message-passing, proposed as a standard by a broadly based committee of vendors, implementors, and users. It was designed for high performance on both massively parallel machines and on workstation clusters [13].

VI. RESULTS

A. Kriging results

The interpolation described in this paper was performed for 13,200 grid nodes. The estimated values of ground deformations are presented on Fig. 2. It can be seen that subsidence with average annual rate higher than 5 mm/yr occur in the SW part of area. The NE part of the region is rather stable with small local regions with average values of deformations. The interpolated values indicate that in NW part of the area the insignificant ground lifting appear. It can be also observed that SW part of the studied region, for which the highest values of subsidence were interpolated, is divided into two parts by a stable region.

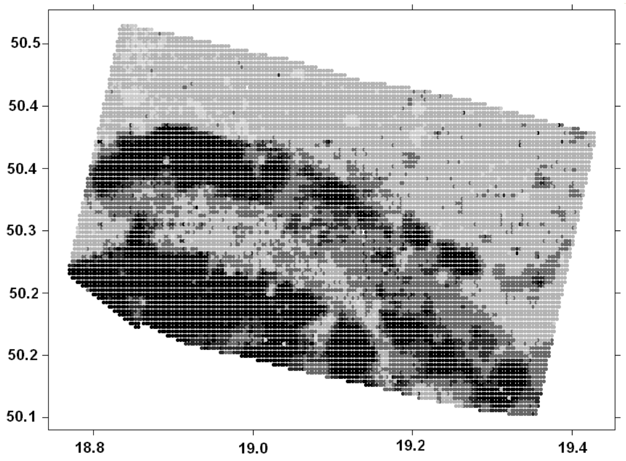

Figure 2. Map of ordinary kriging results

In order to find the origin of ground displacements which occurred in NE part of Upper Silesian Coal Basin the estimated values of deformations have to be complement by different kinds of data. In studied region probably two factors had an effect on ground movements. First one is a contemporary tectonic activity of this area and second one is an intensive underground exploitation performed there for more than 200 years. In order to estimate an influence of these factors on values of ground deformations the obtained map of kriging results should be analyzed in relation to geological maps and mining data.

B. Analyses of proposed algorithm

Processors' load tests showed that proposed algorithm evenly distribute calculations on all nodes. Acceleration described by (6) and effectiveness described by (7) were tested on smaller set of PS points

$$S_p = T_1 / T_p \tag{6}$$

$$E_p = S_p / P = T_1 / pT_p \tag{7}$$

where T_1 is time of calculations for one process and T_p is time of calculations for p processes.

Execution times for 10,000 PS points and 13,200 grid points were measured on different number of processors. Fig. 3 shows calculation times for up to 15 processors. Fig. 4 presents relation between acceleration and number of CPUs. Fig. 5 shows algorithm's effectiveness.

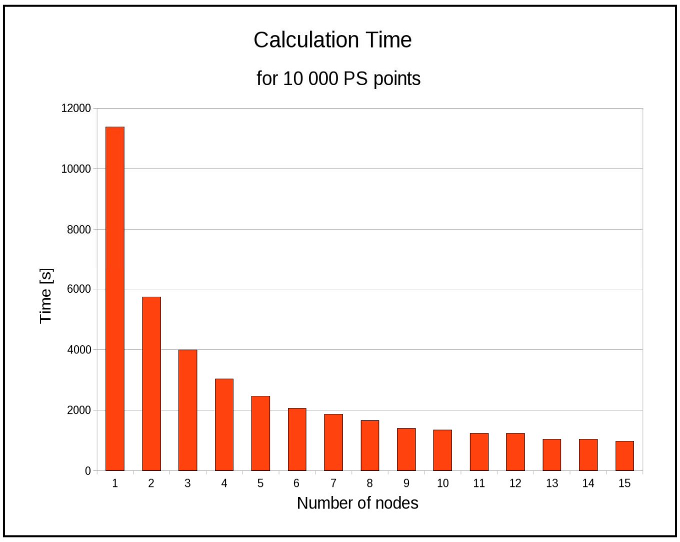

Figure 3. Calculation times for different numbers of processors

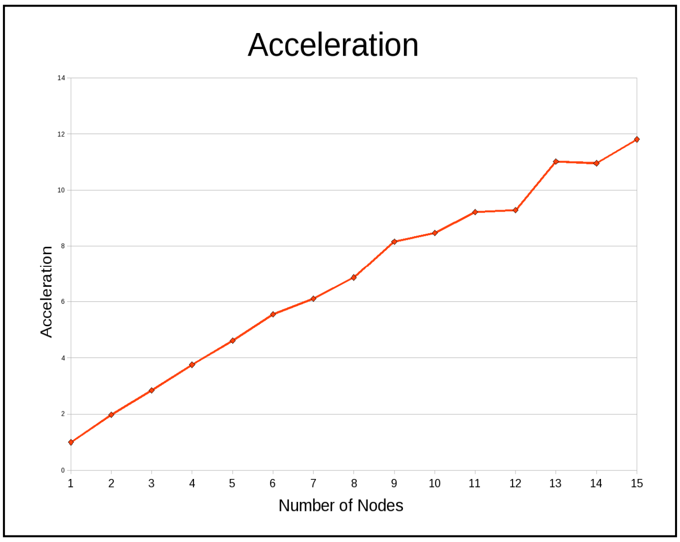

Figure 4. Acceleration for different numbers of processors

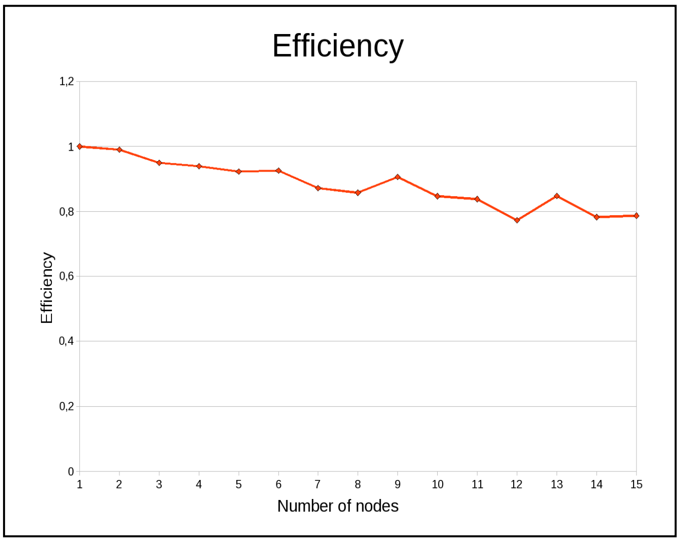

Figure 5. Efficiency for different numbers of processors

According to Amdahl's law [14] the maximum speed up of a parallel program using multiple processors is limited by sequential fraction of the program. In case of created parallel kriging algorithm, only part of it was parallelized. The preliminary operations like creating a rim or calculating number of neighbours for each grid cell were sequential. According to Fig. 3 calculation time decreased with the number of processors up to 13 CPUs. Adding next processors did not considerably shorten execution time.

On the acceleration showed on Fig. 4 some fluctuations appeared after adding 7th process. It can be caused by dividing grid points' classes to unequal portions for specific number of processors. Efficiency Fig. 5 shows similar situation. Fluctuations appeared around 6th-7th node. Further investigation is planned to clarify this phenomenon.

Complete calculations of full model of 120,000 PS points and 13,200 grid cells on three IBM Blade machines took 51 hours. They were executed on 24 processors' cores. Calculation time mostly depends on the model parameters, especially on the radius of an area in which PS points are taken into account in calculations of a grid point velocity. If the radius is bigger, the set of equations is larger and the calculation time is longer. For the experiment radius of 3.0 km was chosen.

VII. CONCLUSIONS

The obtained results showed that geostatistical methods for PSInSAR data allow to use information derived by this technique in more optimal way. The map of estimated displacement obtained during the kriging procedure gave a possibility to analyse ground deformations in relation with other types of data e.g. geological, mining or topography. Such analyse helps to find the origins of ground displacements.

In this work the parallel kriging algorithm for unevenly distributed data was created. It allowed to interpolate velocity values in grid points on the basis of PSInSAR data. The parallelization of kriging procedure allowed to perform the calculations for a large set of PS points. Tests of the proposed algorithm showed that it sufficiently accelerates the calculations.

ACKNOWLEDGMENT

This work was supported by project number 11.11.140.561 AGH University of Science and Technology (Faculty of Geology, Geophysics and Environmental Protection, Department of Geoinformatics and Applied Computer Science).

S. Porzycka is a holder of a scholarship of Anna Pasek Foundation.

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] Kim J, Kim D., Kim S., Won J., Moon W.M.: Monitoringu of Urban land surface subsidence using PSInSAR. Geosciences Journal, vol.11, no. 1, p. 59-73 (2007).

[3] Farina P. Integration of permanent scatterers analysis and high resolution optical images within landslide risk analysis. Proc. Of FRINGE 2003 Workshop, Frascati (Italy), 1-5 December (2003).

[4] Lesniak A., Porzycka S.: Environment monitoring using satellite radar interferometry technique (PSInSAR). Polish Journal of Environmental Studies, vol. 17 no. 3A, 382--387b (2008).

[5] Dehls J.F., Basilico M., Colesanti C.: Ground deformations monitoring in the Ranafjord area of Norway by means of the permanent scatterer technique. 23-th International Geoscience and Remote Sensing Symposium, Toronto, June 24-28, p.203-207 (2002).

[6] Declercq P.Y., Devleeschouwer X., Pouriel F. Subsidence revealed by PSInSAR technique in the Ottignies-Wavre area (Belgium) related to water pumping in urban area. Proc. Of FRINGE 2005 Workshop, Frascati (Italy), 28 November-2 December (2005).

[7] Weckernagel H.: Multivariate Geostatistics. Springer-Verlag, Berlin Heidenberg (1995).

[8] Almasi G.S., Gottlieb A. (1989). Highly Parallel Computing. Benjamin-Cummings publishers, Redwood City, CA.

[9] Coulouris G., Dollimore J., Kindberg T. (2005), Distributed Systems. Concepts and Design, Fourth Edition, Addison Wesley.

[10] Chan T.F., Mathew T.P. Domain decomposition algorithms, Acta Numerica (1994), Cambridge University Press.

[11] Lu K., Goddard S., GRASS-based High Performance Spatial Interpolation Component for Spatial Decision Support Systems, Proceedings of the FOSS/GRASS Users Conference, Bangkok, Thailand, 12-14 September 2004.

[12] Gebhardt A., PVM Kriging with R, Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20-22, Vienna, Austria.

[13] Message Passing Interface web page, http://www.mcs.anl.gov/research/projects/mpi, Mathematics and Computer Science Division, Argonne National Laboratory.

[14] Amdahl G., Validity of the Single Processor Approach to Achieving Large-Scale Computing Capabilities, AFIPS Conference Proceedings, (30), pp. 483-485, 1967.