L. VOKOROKOS, P. BLIŠŤAN, S. PETRÍK, N. ÁDÁM

ISSN 0543-5846
METABK 43 (4) 287-291 (2004)
UDC - UDK 551.4:004.9 GIS=111

UTILIZATION OF PARALLEL COMPUTER SYSTEM FOR MODELING OF GEOLOGICAL PHENOMENA IN GIS

Received - Primljeno: 2003-12-16
Accepted - Prihvaćeno: 2004-04-20
Preliminary Note - Prethodno priopćenje

Geographical data processing using Geographical Information Systems on single processor computer is going to be very difficult in consideration of constantly growing amount of data. In the recent years there is a trend of operation of those systems on parallel computers, which brings an effect mainly in the area of their performance. This article deals with possibilities of shortening of time demanded for mathematically exacting process of both analysis and visualization of variety of geographical data using parallel architectures. It focuses on some methods of geographical data processing as well as on types of parallel computing systems, together with description of their practical application to the environment of Geographical Information Systems.

Key words: *Geographical Information System (GIS), geographical data, parallel computer, interpolation methods*

Uporaba usporednog računalnog sustava za modeliranje geoloških fenomena u sustavu zemljopisnih informacija (GIS). Obrada zemljopisnih podataka uporabom Geographical Information System (GIS) na jednoprocesnom kompjutoru bit će vrlo teška ako se uzme u obzir stalni porast količine podataka. Zadnjih godina postoji tendencija rada tih sustava na usporednim računalima, a to ima utjecaja uglavnom na područje njihovih performansi. Ovaj članak se bavi mogućnostima skraćivanja vremena potrebnog za komplicirane (exacting) matematičke operacije i vizualizacije mnoštva zemljopisnih podataka uporabom usporednih arhitektura. Usredotočuje se na neke metode obrade zemljopisnih podataka kao i na tipove usporednih računalnih sustava, zajedno s opisom njihove praktične primjene na okoliš Grographical Information Systems.

Ključne riječi: *sustav informiranja o zemljopisnim podacima (GIS), geografski podaci, usporedni kompjuteri, interpolacijske metode*

INTRODUCTION

Present times can be defined by growing amount of data and information as well as growing complexity of data flows. In order to process this amount of data and information there are efforts to development of specialized systems to process of variety of geographical data - Geographical Information Systems (GIS). The usage of GISs has successfully spread to the area of geology, mining and metallurgy.

Geological phenomena modelling using GIS depends on correct interpretation of large amount of data acquired by measurement and observation. Both primary analysis and processing of acquired data are usually performed on single processor computer. This fact caused, that entire process is going to be highly time-consuming.

Next step in a process of computer performance increasing was to merge performance of two or more processors. High performance multiprocessor systems [1] encompass mathematically highly exacting visualization [2] of modelling methods (e.g. geological objects modelling in geology).

PARALLEL COMPUTER SYSTEMS ARCHITECTURES

Multiprocessor system (MS) [1] is a group of interconnected computers, which are either solving one complex task or independently contributing to the processing of couple of programs.

Speed of MS in generally depends on used interconnection network between processors as well as on both suitability of used parallel architecture and used parallel algorithms.

L. Vokorokos, S. Petrík, N. Ádám, Faculty of Electrical Engineering and Informatics Technical University of Košice, Košice, Slovakia, P. Blišťan, Faculty of Mining, Ecology, Management and Geotechnologies Technical University of Košice, Slovakia

Multiprocessor systems are divided in generally into [1]:

1. MS with shared memory (multiprocessors). In this architecture a group of processors shared one common memory.
2. MS with distributed memory (multicomputers) [2]. Each processor in the group uses its own local memory. The system is composed by node computers linked by static interconnection network [3].

An example of technical realization of multicomputer is cluster [4], composed by group of either single- or more-processor computers. Communication between processor elements is realized by fast intercommunication network (e.g. Fast Ethernet, Gigabit Ethernet). This technology is also used by ParallelGIS system closely described in this article, which is considered as a benefit to the research of parallel architectures applications to GIS [5, 6].

Development or modification of sequential program to its parallel form [4] is a problem of assignation of partial tasks to node computers, to make entire process as effective as possible. In generally, parallelism can be functional or data-oriented [4].

SPATIAL MODELLING IN GIS

This chapter closely describes some interpolation methods used for spatial modelling in GISs. Geoinformation systems are considered as information systems for effective storage, actualization, manipulation, analysis, modelling and presentation of geographical information. Probably the most difficult task solved in GISs is spatial modelling in geology and mining. For example, terrain modelling (Figure 1.), geological objects morphology modelling or modelling of distribution of observed phenomena [7].

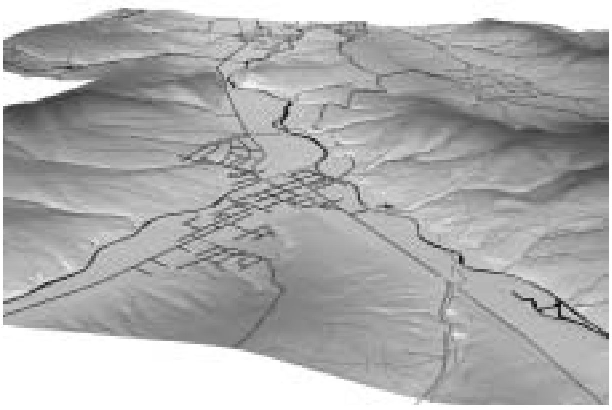

Figure 1. **Spatial terrain model created by GIS**
Slika 1. **Prostorni model terena napravljen GIS-om**

Complexity of modelling depends on both amount of input data and selected method. There are couple of reliable interpolation methods for modelling in geology. All of them are solving such situation, which can be defined as follows: inside of an area of interest (Figure 2.) are some

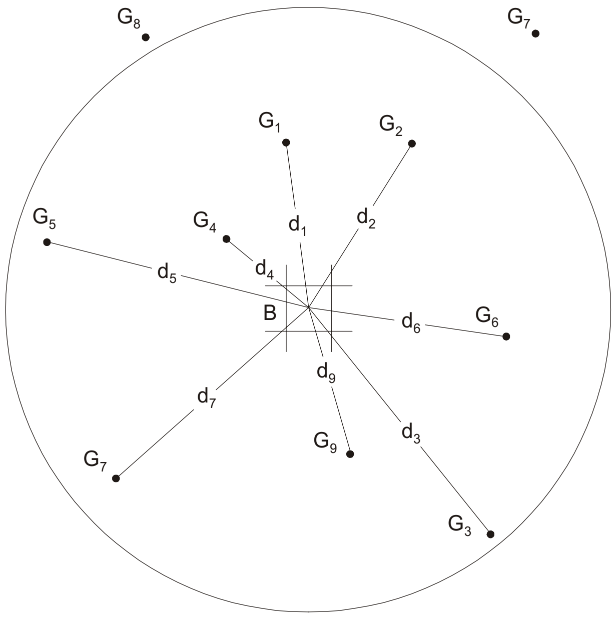

Figure 2. **Model situation**
Slika 2. **Model situacije**

irregularly placed areas G_1 - G_9 in which the values of observed variables g_1 - g_9 are known (e.g. after analysis of acquired pattern). The propose is to provide as accurate estimation of observed variable value as possible in an arbitrary point inside the area of interest (for example in area B), based on a set of known values.

Interpolation methods used in GIS

Problem given in Figure 2. is probably the most solved problem in practice. Basically, it is three-dimensional interpolation, where two dimensions (considering them as arbitrary - axe x and y) composing the surface (e.g. terrain) and the third axe is observed variable. The problems in geological practice are often considered as four-dimensional. Known points lie in the 3D area (e.g. stone block or geological object). A point position is given by the x, y and z coordinates. The fourth dimension is the value of observed variable. There are two often used methods for the estimation of observed variable in the point B (Figure 2.): Inverse Distance Square Method (IDS) and Kriging Method.

Inverse Distance Square Method (IDS)

Inverse Distance Square Method is from the group of so-called „contributional“ methods. The value of observed variable T_B in the point B (Figure 2.) can be realized as a compendium of contributions g_1 - g_9 from the known points G_1 - G_9 to B.

Hence, the value of observed variable T_B mostly depends on both known values and distances of d_1 - d_9 points from B. Logically said: the contribution of G_1 (Figure 2.) with known value g_1 to B will be surely less then the contribution of G_7 with known value g_7. This dependency can by expressed by the following equation:

$$T_B = \frac{\sum_{i=1}^{n} \frac{g_i}{d_i^k}}{\sum_{i=1}^{n} \frac{1}{d_i^k}} \quad (1)$$

IDS method is mathematically more difficult as the widely used linear interpolation. It is often used for modelling of more exacting terrain forms than usually, because of consideration of some irregularity in spatial distribution of observed phenomena. (Distribution of beneficial component in the mineral deposit as shows Figures 3. and 4.)

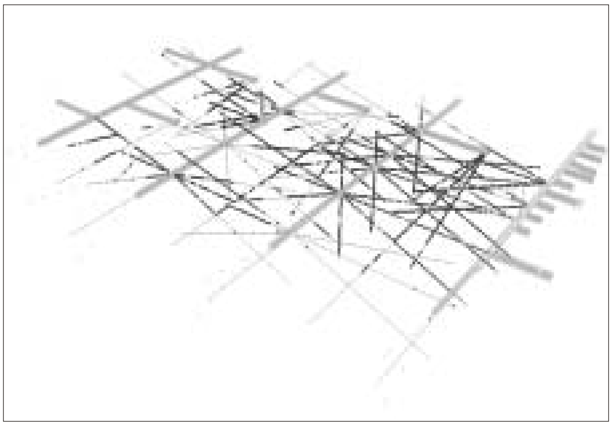

Figure 3. **Spatial position of mine and bores with values of beneficial component in patterns**
Slika 3. **Prostorni položaj rudnika i bušotina s vrijednostima pozitivne komponente u uzorcima**

Kriging Method

This method also comes under "contributional" methods. It is often used to compute both local and global estimations of observed variables [8]. There is a basic equation valuable for Kriging Method [7]:

$$T^* = \sum_{1}^{n} w_i g_i \quad (2)$$

where:

T^* - estimated value in the point B,
g_i - known value of observed variable in i-point,
w_i - weight of observation in the i-point.

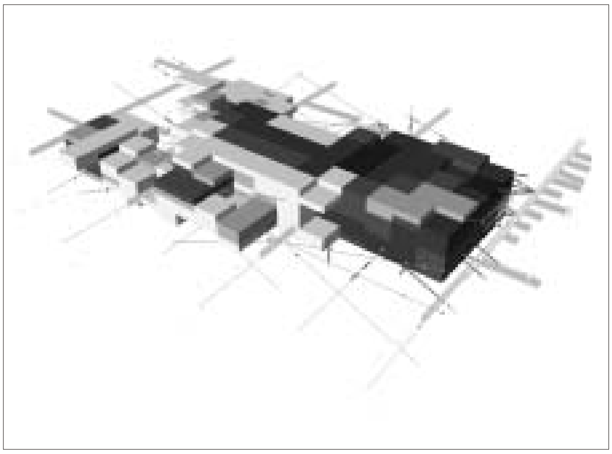

Figure 4. **Spatial model of beneficial component distribution computed by Kriging Method**
Slika 4. **Prostorni model prostiranja pozitivne komponente izračunate metodom Kriginga**

Criterion of convenience of estimation T^* against a real value T is the estimation of dispersion given by equation 3:

$$\sigma_\varepsilon^2 = 2\sum_{i=1}^{n} w_i \bar{\gamma}(g_i, A) - \\ -\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \bar{\gamma}(g_i, g_j) - \bar{\gamma}(A, A) \quad (3)$$

where:

$\bar{\gamma}(g_i, A)$ - average semivariogram between every point g_i with known value of observer variable and estimated area A,
$\bar{\gamma} = (g_i, g_j)$ - average semivariogram between the set of points with the known value of observed variable,
$\bar{\gamma} = (A, A)$ - average semivariogram between each point in an observed area A.

The concrete value of observed variable, computed by Kriging Method depends on:
- pattern geometry in an estimated area,
- semivariogram model,
- weights assigned to patterns.

There is a need of dispersion defect minimization for optimal estimation:

$$\sigma_\varepsilon^2 - \min \quad \rightarrow \quad \frac{\partial \sigma_\varepsilon^2}{\partial w_i} = 0 \quad (4)$$

and after modification of equation 4:

$$\frac{\partial\left(\sigma_{\varepsilon}^{2}-\lambda\left(\sum w_{i}-1\right)\right)}{\partial w_{i}}=0 \qquad (5)$$

If we take all partial derivatives equal to zero, we get an equation system the results of which are weight $w_i \ldots w_n$, with valuable basic condition $\sum_{i=1}^{n} w_i = 1.$

Processing of large amounts of data by Kriging Method is highly time-consuming also for computer. It takes from a few minutes to several hours to process input data (for thousands of input data), thus it is convenient to perform those calculations on high performance PCs or specialized either multiprocessor or multicomputer systems [3].

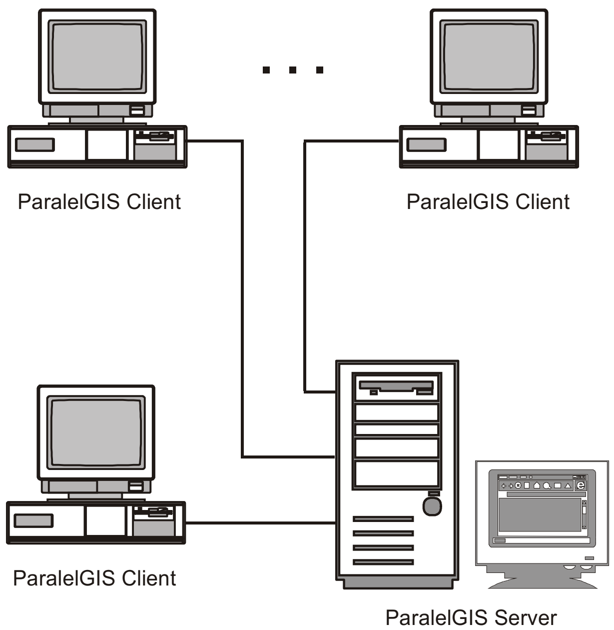

Figure 5. **ParallelGIS system architecture**
Slika 5. **Sustav arhitekture usporednog GIS-a**

It is possible to shorten processing time demanded for large sets of data using both data-oriented parallelism and clusters.

ParallelGIS SYSTEM - STRUCTURE AND COMPONENTS

Practical application of geographical data processing using interpolation methods and parallel architectures is ParallelGIS system, being developed by Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics in cooperation with Department of GIS, BERG Faculty on the Technical University of Košice, Slovakia [5, 6]. Figure 5. shows the basic scheme of ParallelGIS system.

System is composed of two basic elements: ParallelGIS Server a ParallelGIS Client.

ParallelGIS Server. Contains database of input data ordered in files. Data have been acquired by carring patterns analysis. Stored data are used for computing both beneficial (Fe) and harmful component distribution inside the deposit of iron ore. Information are provided by PGserver application installed on ParallelGIS server.

ParallelGIS Client. Provides analysis of data acquired from ParallelGIS server. The analysis is performed by ArcView system using Avenue programming language able to process sets of geographical data. Data requesting and receiving process from ParallelGIS server as well as ArcView data processing initialization procedure and processed data send-back procedure are provided by PGclient application.

ParallelGIS SYSTEM APPLICATION AND ACHIEVED RESULTS

Efficiency of data processing by ParallelGIS system was tested on a pattern of five randomly chosen mineral deposits. There was one mineral deposit used in each test, it consists of distribution of both beneficial and harmful component modelling. Mineral deposit size was rated be amount of input data (a number of patterns from of mineral deposit) and number of microblocks to which mineral deposit was divided in the time of calculation as shows Table 1. (each mineral deposit was horizontally and vertically divided to several thousand microblocks, which represent the basic mining unit (block) [9]).

Table 1. **Mineral deposits size**
Tablica 1. **Veličina mineralnih depozita**

Deposit name	Number of analysed patterns	Number of microblocks
Kobeliarovo	1500	1000
Bankov - Košice	3200	3500
Kišovce - Švábovce	150	900
Jelšava	3800	4500
Rožňava - Strieborná	117	250

There were five ParallelGIS Client computers and one Parallel GIS Server (Figure 5.) used in each test. The result of each test was rated by the time demanded to input data processing - chemical analysis of patterns from mineral deposit. The results are shown in Figure 6. It also includes time demanded for such amount of input data using single processor system (Figure 6.).

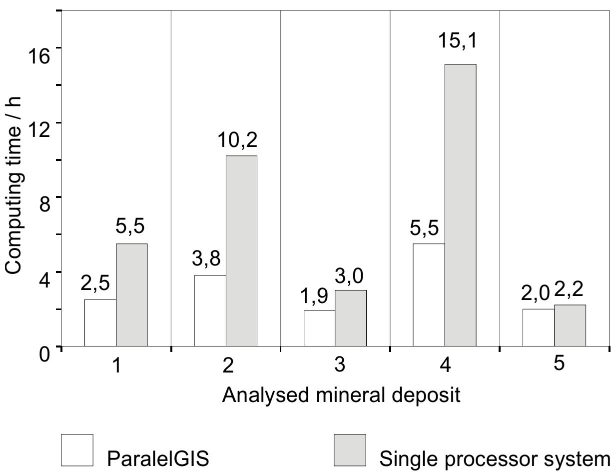

Figure 6. **Times demanded to process input data by ParallelGIS system and single processor system**
Slika 6. **Vremena potrebna za obradu ulaznih podataka sustavom usporednog GIS-a i sustavom jednog procesora**

CONCLUSION

Data analysis using Geographical Information Systems is highly time-consuming process. Hence, it was suggested to use parallel computer architectures for those computing in the form of ParallelGIS system, developed as a part of research [5, 6]. There are obvious advantages resulting from the measured times demanded for the analysis and processing of geographical data using ParallelGIS system instead of single processor computer. The test results suggest that convenience of parallel system usage depends on amount of input data and both mathematical difficulty of selected modelling method and complexity of final model.

There is a future development planned for ParallelGIS system, mostly to PGclient application functionality widening, considering multilevel data processing by additional Avenue programming language modules used in ArcView GIS.

REFERENCES

[1] K. Hwang: Advanced Computer architecture, Mc-Graw Hill, Inc., New York, 1993.
[2] B. Sobota: Počítačová grafika a jazyk C, KOPP, České Budějovice, 1995.
[3] L. Vokorokos: Principles of Data Flow Controlled Computer Architectures, Copycenter, spol. s r.o., Košice, 2002, ISBN 80-7099-824-5, (in Slovak).
[4] M. Jeřábek: Využití paralelných výpočtu v geodezi, dissertation, ČVUT, 2001.
[5] M. Jelšina: Project VEGA No.: 1/9027/2002, Výskum paralelných architektúr špecializovaných vysoko výkonných počítačových systémov: architektonické riešenie, metódy hodnotenia, simulácie a aplikácie, Košice 2002.
[6] P. Blišťan: Project VEGA No.: 1/9359/02, Analysis and modelling of geological and economical parameters influencing mining of Slovak magnesite deposits and its impact on the environment on the example of the Košice-Bankov deposit, (in Slovak), Košice 2002.
[7] P. Blišťan, J. Kondela: Geologický informačný systém pre ložisko Jelšava, Acta Montanistica Slovaca (2002) 4, 223 - 226.
[8] L. Vizi, G. M. Timčák: Význam štúdia lognormálneho rozdelenia v geológií a baníctve. Sborník vědeckých prací Vysoké školy banské - Technické univerzity Ostrava, číslo 1, 2002, Ostrava, str. 29 - 39.
[9] P. Blišťan, J. Kondela: Základy banskej geológie a výpočtu zásob, Elfa, s.r.o., Košice, 2001, pp. 103.