GEOGRAPHICAL DATA PROCESSING USING INTERPOLATION METHODS AND PARALLEL COMPUTER SYSTEM

Liberios VOKOROKOS, Slavomír PETRÍK
Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, 042 00 Košice, Slovak Republic, E-mail: Liberios.Vokorokos@tuke.sk, Slavomir.Petrik@tuke.sk

SUMMARY

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 these 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 a 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.

Keywords: *parallel computing, Geographical Information System, interpolation method*

1. 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 develop specialized systems to process geographical data – Geographical Information Systems (GIS). GISs are foremost intended to archive data, but they are also solving some specific tasks, e.g. modeling beneficial component distribution of mineral deposits.

Geological phenomena modeling 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 cause, that entire process is going to be highly time-consuming.

A single processor computers passed by many changes during their development, mainly focused on increasing of their performance. Considering fact, that single processor computers have their upper performance limits, the next step in a process of computer performance increasing was to merge performance of two or more processors. This technique requires special hardware and software support, but once it is correctly used, it surpass performance of single processor computers. Multiprocessor systems encompass mathematically highly exacting visualization [7] a modeling methods (e.g. geological objects modeling in geology) in a few minutes or seconds instead of days or hours when using single processor computer.

2. PARALLEL COMPUTER SYSTEMS ARCHITECTURES

Multiprocessor system (MS) [3,4,5,8] is a group of interconnected computers, which are either solving one complex task or independently contributing to processing of couple of programs.

There is a need to solve a communication between a couple concurrently performed processes (except a parallel processes which don't need to use results (data) of other concurrently performed process). Communication between processors is provided by interconnection network (ICNW – Interconnection Network) [5,6]. Topology of these networks is one of the key parameters of overall performance and fidelity of parallel system. Examples can be de Bruin networks, tree based networks or hyper cubic networks [5,6].

Software solution of mentioned inter-processor communication strongly depends on used hardware. Multiprocessor systems are divided in generally to [4]: MS with assigned memory (multiprocessors) and MS with shared memory (multicomputers).

Clusters could be an example of technical realization of multicomputer. Clusters are composed by a group of either single- or more- processor computers. Communication between processor elements is usually realized by fast interconnection network (e.g. Fast Ethernet or Gigabit Ethernet). Cluster technology is also used by developed ParallelGIS (developed as a part of researches [9]) system closely described later in this article, because of its relatively unpretentious technical and economical realization in contrast of MS with assigned memory.

2.1 Parallel algorithms and parallelization

Development or modification of sequential program to its parallel form [6] is a problem of assignation of partial tasks to node computers, to make entire process as effective as possible. Parallelization can be done either manually or automatically. To parallelize sequential program automatically a special compiler is used, which solves apart from parallelization itself also data dependencies between concurrent program branch. Very eligible parts of program to parallelize are loops, but mentioned data dependencies need to be solved.

A great many of tasks can not be parallelized completely, because usually there are parts which must be performed sequentially. To asses an effect of parallelization Gene Amdahl envolved so called Amdahl rule [6], which expresses a ratio of time demanded to solve problem on single-processor computer and a time demanded to solve the same problem partially parallelized a and performed on multiprocessor computer.

If a task can be divided to more different subtasks, which can be done concurrently, we can divide this program to more functions which can be done concurrently. This is called functional parallelism.

When all concurrently running processes are performing the same task but with different data it is called data parallelism [6]. ParallelGIS system (developed as a part of parallel system research [9]) described later in this article is an example of data parallelism.

Next chapter closely described some methods and techniques of spatial modeling in GIS. With usage of data parallelism and aforementioned clusters a time demanded to process large sets of data could be strongly shortened.

3. SPATIAL MODELING IN GIS

Geoinformation systems are considered as information systems for effective storage, actualization, manipulation, analysis, modeling and presentation of geographical information. Probably the most difficult task solved in GISs is spatial modeling in geology and mining. For example terrain modeling (Figure 1), geological objects morphology modeling or modeling of distribution of observed phenomena [1] (distribution of beneficial component in the mineral deposit as show Figure 4).

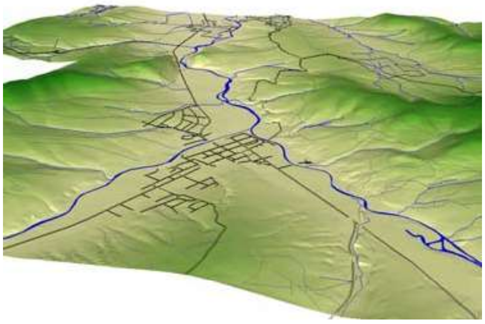


Fig. 1 Spatial terrain model created in GIS

Complexity of modeling depends on both amount of input data and selected method. There are couple of reliable interpolation methods for modeling 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 irregularly placed areas $G_1 - G_9$ in which a 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.

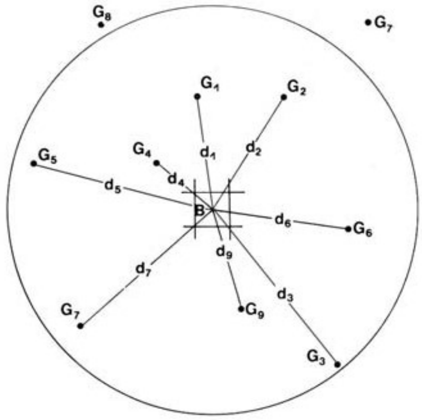


Fig. 2 Modeled situation

3.1 Interpolation methods used in GIS

Problem given by Figure 2 is probably the most solved problem in praxis. 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. Problems in geological praxis are often considered as four-dimensional. Known points lies in the 3D area (e.g. stone block or geological object). Point position is given by the x, y and z coordinates. Fourth dimension is value of observed variable. There are three oftentimes used methods for estimation of observed variable in the point B (Figure 2): Triangular method, Inverse Distance Square Method (IDS) and Kriging Method.

3.1.1 Triangular method

Triangular method provides estimation of unknown value based on linear dependency. Plane is a linear object in three dimensional space. Plane is given by an equation:

$$z = a . x + b . y + c \quad (1)$$

Equation (1) contains three coefficients. This means that there is need to known tree points to determine arbitrary plane. e.g. $G_2 = [x_1,x_1,x_1]$, $G_4 = [x_2,x_2,x_2]$, $G_9 = [x_3,x_3,x_3]$, which holds the rules:

$$\begin{aligned} z_1 &= a . x_1 + b . y_1 + c \\ z_2 &= a . x_2 + b . y_2 + c \\ z_3 &= a . x_3 + b . y_3 + c \end{aligned} \quad (2)$$

Because all x_i, y_i a z_i values are known, the result of equation system (2) are coefficients a,b and c of those plane which includes the points G_2, G_4 a G_9. Estimation T_B in point B is given by equation:

$T_B = a.x_B + b.y_B + c,$ (3)

where x_B and y_B are coordinates of B (Figure 3).

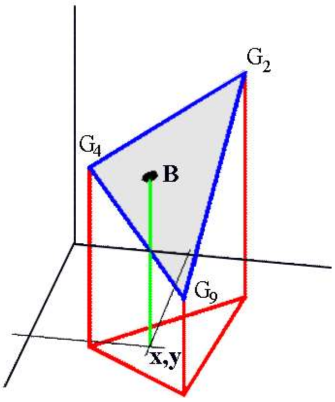


Fig. 3 Triangular method

3.1.2 Inverse Distance Square Method (IDS)

Inverse Distance Square Method is from the group of so-called „contributional“ methods. 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 known points G_1-G_9 to B. Hence, value of observed variable T_B mostly depends on both known values and distances of d_1–d_9 points from B. Logically said: contribution of G_1 (Figure 2) with known value g_1 to B will be surely less then contribution of G_7 with known value g_7. This dependency can by expressed by the following equation:

$$T_B = \frac{\sum_{i=1}^{n} \frac{g_i}{d_i^k}}{\sum_{i=1}^{n} \frac{1}{d_i^k}} \quad (4)$$

IDS method is mathematically more difficult as the widely used linear interpolation. It is often used to modeling of more exacting terrain forms as usually, because of consideration of some irregularity in spatial distribution of observed phenomena. Computer processing of large amount of data by Kriging Method is slightly more time-consuming then aforementioned linear interpolation, thus it should be realized on high performance computers.

3.1.3 Kriging Method

This method also comes under "contributional" methods. It is often used to compute both local and global estimations of observed variables. There is a basic equation valuable for Kriging Method [1]:

$$T^* = \sum_{1}^{n} w_i g_i \quad (5)$$

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.

Criterion of convenience of estimation T^* against real value T is the estimation dispersion given by equation 6.

$$\sigma_\varepsilon^2 = 2\sum_{i=1}^{n} w_i \bar{\gamma}(g_i, A) - \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \bar{\gamma}(g_i, g_j) - \bar{\gamma}(A, A) \quad (6)$$

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 known value of observed variable, $\bar{\gamma} = (A, A)$ - average semivariogram between each point in an observed area A.

Particular 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 estimation dispersion defect minimization for optimal estimation:

$$\sigma_\varepsilon^2 - \min \quad \rightarrow \quad \frac{\partial \sigma_\varepsilon^2}{\partial w_i} = 0 \quad (7)$$

and after modification of equation 7:

$$\frac{\partial\left(\sigma_\varepsilon^2 - \lambda\left(\sum w_i - 1\right)\right)}{\partial w_i} = 0 \quad (8)$$

If we take all partial derivative equal to zero, we gets an equation system which results are weight $w_i \ldots w_n$, with valuable basic condition:

$$\sum_{i=1}^{n} w_i = 1 \quad (9)$$

This method is mathematically more exacting as previous one. Kriging is proper to solve either complicated geological problems or modeling of exacting terrain form, where it gets very good results. Manual processing of data by this method is highly time-consuming, thus Kriging Method is widely used since computers take place in a processing of geological data. 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.

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

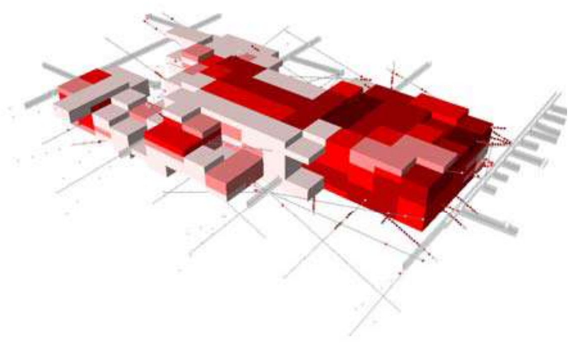


Fig. 4 Spatial model of beneficial component in a mineral deposit

4. 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 on the Technical University of Košice, Slovakia as a part of research project [9]. Figure 5 shows the basic scheme of ParallelGIS system.

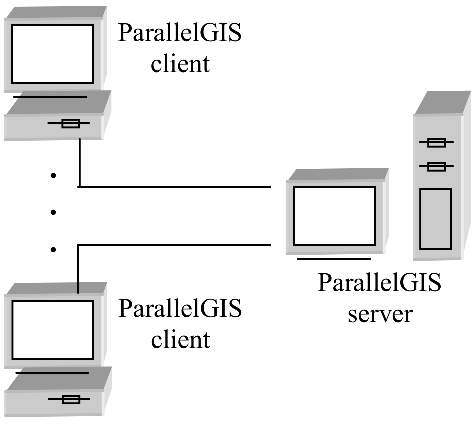


Fig. 5 ParallelGIS system structure

System is composed by three basic elements: ParallelGIS Server, ParallelGIS Transaction Monitor and ParallelGIS Client.

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

ParallelGIS Client. Provide analysis of data acquired from ParallelGIS server. Analysis is performed by ArcView system using Avenue programming language able to work over 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 Transaction Monitor. Provides monitoring of data transmissions between PGclient and Pgserver and moreover provides a system diagnostic (mainly network communication failures). ParallelGIS TransactionMonitor is realized as in/out application filter installed on ParallelGIS server as shows Figure 6.

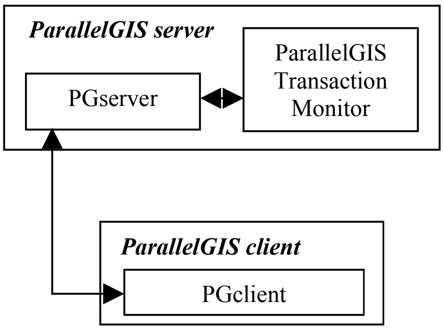


Fig. 6 Components of ParallelGIS system

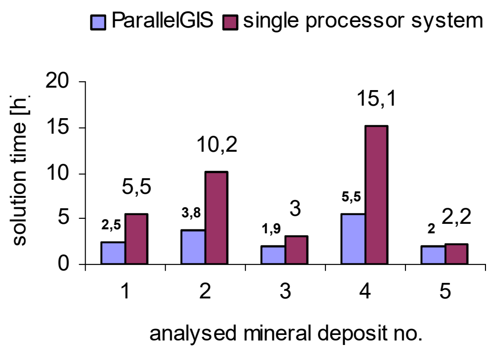
5. PARALLELGIS SYSTEM APPLICATION AND ACHIEVED RESULTS

Efficiency of data processing by ParallelGIS system was tested on pattern of five randomly chose mineral deposits. There was one mineral deposit used in each test. Calculation of distribution of both beneficial and harmful component have been performed over it. Mineral deposit size was rated be amount of input data (number of patterns from 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) [2]).

Názov ložiska	**Počet analyzovaných vzoriek**	**Počet vypočítaných mikroblokov**
Kobeliarovo	1500	1000
Bankov - Košice	3200	3500
Kišovce - Švábovce	150	900
Jelšava	3800	4500
Rožňava - Strieborná	117	250

Tab. 1 Parameters of mineral deposits used in tests

There were five ParallelGIS Client computers and one Parallel GIS Server (Figure 6) used in each test. Result of each test was rated by the time demanded to input data processing – chemical analysis of patterns from mineral deposit. Results are shown by Graph 1. Also includes time demanded for such amount of input data using single processor system.



Graph 1 Times demanded to process input data by ParallelGIS system and single processor system

6. 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. There are obvious advantage resulting from measured times demanded for analysis and processing of geographical data using ParallelGIS system instead of single processor computer. Test results suggest that eligibility of parallel system usage depends on amount of input data and both mathematical difficulty of selected modeling method and complexity of final model. Time demanded for data transmissions among ParallelGIS system elements must considered.

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] Blišťan, P., Kondela, J.: Geographical Information System for Jelšava mineral deposit, Acta Montanistica Slovaca, 4/2002, Košice, pp. 223-226, ISSN 1335 – 1788

[2] Blišťan, P., Kondela, J.: Fundamentals of Mining Geology and resources computing, Elfa, s.r.o., Košice, 2001, pp. 103, (in Slovak)

[3] Hwang, K.: Advanced Computer architecture, Mc-Graw Hill, Inc., New York, 1993

[4] Jelšina, M.: Computer Systems Architectures, Elfa, s.r.o., Košice, 2001, ISBN 80-89066-40-2, (in Slovak)

[5] Jelšina, M.: Project VEGA No. 1/9027/2002, Research of Parallel Architectures of specialized high performance computer systems: architectonic, rating methods, simulations and applications

[6] Jeřábek, M.: Utitisation of Parallel Computing in Geodesy, dissertation thesis, ČVUT, 2001, (in Czech)

[7] Sobota, B.: Computer Graphics and C Language, KOPP, České Budějovice, 1995, (in Slovak)

[8] Vokorokos, L.: Principles of Data Flow Controlled Computer Architectures, Copycenter, spol. s r.o., Košice, 2002, ISBN 80-7099-824-5, (in Slovak)

[9] Vokokos, L.: Project VEGA No. 1/1064/2004, Parallel Computer Architectures, their specification methods, developement technologies and implemntations

Supported by VEGA projects No.: 1/9027/02, 1/1064/2004.

BIOGRAPHY

Liberios Vokorokos (doc., Ing., PhD.) was born on 17.11.1966 in Greece. In 1991 he graduated (MSc.) with honours at the department of Computers and Informatics of the Faculty of Electrical Engineering and Informatics at Technical University in Košice. He defended his PhD. in the field of programming device and systems in 2000; his thesis title was "Diagnosis of compound systems using the Data Flow applications". He was appointed Associative Professor and extraordinary professor for Computers Science and Informatics in 2001 and 2003, respectively. Since 1995 he is working as an educationist at the Department of Computers and Informatics. His scientific research is focusing on parallel computers of the Data Flow type. In addition to this, he also investigates the questions related to the diagnostics of complex systems. Currently he is a member of the State Examination Committee in the field Computing engineering and Informatics. His other professional interests include the membership on the Advisory Committee for Informatization at the faculty and Advisory Board for the Development and Informatization at Technical University of Košice.

Slavomír Petrík (Ing.) was born in Spišská Nová Ves, in 1980. He received the Ing. (MSc.) degree in computers and informatics from the Faculty of Electrical Engineering and Informatics, Technical University of Košice, in 2003. He deals with topologies and simulations of interconnection networks in parallel computers.