UDC 004.67:514.83

RECOVERY OF TWO-DIMENSIONAL GEOFIELDS: PROBLEMS AND SOLUTIONS

M.V. Kopnov, R.V. Kovin

Institute «Cybernetic centre» of TPU
E-mail: kopnovmv@tpu.ru

Complexities of two-dimensional geofields recovery have been analyzed. The results of investigations and developments in the field of geoinformation systems and technologies for analyzing two-dimensional geofields were given.

Introduction

One of the most urgent and dynamically developing directions in geoinformatics for today is tridimensional analysis of objects representing extensive continuous surfaces. Such surfaces may show surface propagation of temperature and pressure, lay of land heights above sea level, chemical elements distribution in soil etc. In geoinformatics surfaces described definitely by scalar function of two spatial values x and y are called geofield [1].

The most widespread problem when one works with spatial distributed data is to obtain geofield values in the range where measures were not carried out. This problem solution is complicated by the following features of initial spatial data:

- information about investigated phenomenon is known to a certain degree of confidence only in some regions of geofield;
- these regions represent more often testing points (points of area, points of field measuring in a certain medium where the investigation was carried out and as a result of which the geofield value was determined in them etc.);
- as a rule, testing points represent irregular point network.

Requirements to recovery methods of two-dimensional geofields

The most part of researches thinks that [1, 2] recovering geofields the method should:

- be stable to measurement errors, supports efficient filtration of these errors;
- be stable relative to the structure of initial data network;
- support obtaining approximation function rather smooth both in the neighborhood of experimental point and within the whole geofield excluding occurrence of false (unstipulated by existing information) anomalies that is great deviation from regional constituent of geofield which can occur owing to the features of the applied approximation methods;
- support the possibility of carrying out variant calculations at geofield recovery;
- be rather manufacturable that is rather accurate and efficient in consuming machine resources.

Among the most profitable products allowing solving the problems of geofield recovery the software products Spatial Analyst and Geostatistical Analyst for GIS ArcGIS of ESRI Company (USA), Vertical Mapper from MapInfo Corporation (USA) for MapInfo Professional, Encom Discover from Encom (Australia) for MapInfo Professional may be noted. Among the domestic products the system SurfMapper for MapInfo Professional [3] may be noted. The listed products are not independent systems but they represent complementary modules to the specified universal geoinformation systems. The product Surfer from Golden Software (USA) should be singled out among the systems being not geoinformation one.

Investigations of software implementation of many products showed [3] that determinative methods of two-dimensional geofield recovery became the most widespread. However, the suggested methods do not fully satisfy the described requirements in some products.

Determinative methods

Using determinative methods it is supposed that the analyzed data are described by a certain determinative function $V(x, y)$ determined on the investigated region S where (x, y) are the coordinates of the point. The task is to construct function $V(x, y)$ for the whole investigate region S on the basis of known data $V_i=V(x_i, y_i)$ – the values measured in points (x_i, y_i) and another information on the investigate object (phenomenon). After that geofield value in any point of the investigated region S may be simply calculated by the formula.

Determinative methods of geofield recovery often called interpolators may be divided into local and global interpolators. Implementation of any global interpolator may be reduced to carrying out of two main stages:

- determinative function $V(x, y)$ (it may be, for example, solution of linear equation system or statistical calculation) should be calculated taking into account all initial points;
- geofield values are calculated in the required points by determinative function $V(x, y)$ directly.

Spline interpolation, trend analysis (approximation by polynomials) and others are often implemented as global interpolators. Global interpolators should be used there where the phenomenon may be described by determinative function over the whole investigated region. In this case surface recovered by global interpolator is smooth, as a rule.

At large quantity of initial points (tens of thousands) use of some methods is inefficient: for example, if time of calculation of determinative function depends quadratically on a number of initial points. Besides, global interpolators smooth substantially the surface at approximation. Therefore, to surmount such defects local interpolators are used [4].

Implementation of any local interpolator may be reduced to carrying out the following stages:

- finding the nearest initial points to geofield design point;
- calculating determinative function $V(x, y)$ on the basis of the found points;
- computing geofield values directly by determinative function $V(x, y)$;
- carrying out the first three stages for all computed points of geofield.

The method of inversely weighted distances, local approximation by polynomials, radial elemental function and others are implemented in the form of local interpolators [1, 4]. The advantage of local interpolators on global ones is in the fact that changing a number of the nearest points it is possible to control the degree of filtration of initial data and therefore, the magnitude of surface smoothing.

Using determinative methods a number of problems occurs. *Firstly*, it is necessary to justify a selection of interpolator. As it is known, such justification requires additional information on the nature of recovered phenomenon. Such information is often weakly formalized or absent at all. *Secondly*, many methods have several adjustable parameters requiring sensible use. And, *thirdly*, it is not always possible to compare parameters of the selected method with additional information on the recovered geofield.

Accuracy of geofield recovery

It is known that the problem of geofield recovery is ill-posed [1–5]. Therefore, it may be solved by different methods and with different accuracy. In each method for solving this problem the specified representations of geofield nature are introduced. As a result, each method supports different interpretation of initial data. The main complexity consists in construction of the most accurate surface corresponding most of all to reality (described phenomenon). Achievement of this aim is connected with a number of «success factors»:

- presence of sufficient amount of initial valid data;
- deep understanding of the researched phenomenon;
- application of adequate mathematical methods of describing laws of phenomenon propagation;
- presence of convenient tools for studying data, constructing surfaces and estimating their validity.

To solve the problem of geofield recovery accuracy the approach based on tool post-processing of recovered geofield was suggested by the authors. The matter of this approach consists in using tools of surface deformation. The suggested approach may be used at solving the problem of geofield recovery when the initial data are weakly formalized and therefore they can not be used in the existing methods of recovery.

Let us assume that it is necessary to recover the field of formation top and bottom of oil deposit productive stratum. The initial data for this problem is information on the depth of formation top and bottom of this seam in some wells. It is also known that such seam thickness in region *B* equals to zero (in geology this case is called bed thinning), Fig. 1.

Using traditional methods and algorithms of geofield recovery does not allow solving correctly this problem (Fig. 1, *a*). However, using additional information given in the form of network-pattern it is possible to implement such bed thinning (Fig. 1, *b*).

The technique may be interpreted in the following way: surface is extended till the geofield value is obtained in the network-pattern specified by user.

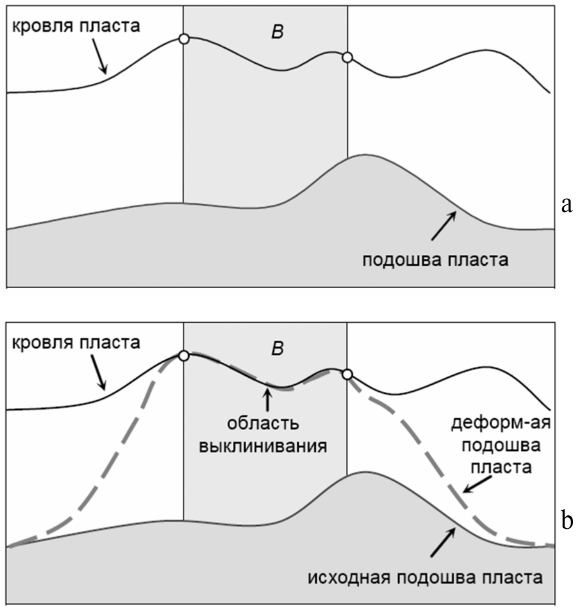

Fig. 1. *The example of refining oil stratum configuration: a) before editing and b) after*
Кровля пласта – formation top; подошва пласта – surface of stratum; область выклинивания – thinning region; деформированная подошва пласта – deformed surface of stratum; исходная подошва пласта – original surface of stratum

It should be noted that use of weakly formalized data at geofield recovery represents an iterative process including construction of residual maps, error probability maps, calculation of different statistical indicators as well as geofield visualization that is processing of heterogeneous data having spatial binding. Therefore, it is sensible to use for such estimation the geoinformation approach within which the main methods of tridimensional analysis are implemented.

Our experiment in the field of solving the problems of geofield recovery allows making the following conclusions:

- none of the methods of geofield recovery is the best from the point of view of accuracy;
- at a great number of initial points all strict interpolators give just the same results;

- to detect the main (local) laws it is necessary to use methods of global (local) trend-analysis;
- it is unallowable to use strict interpolators in the conditions when initial data contain errors.

Problems of method numerical efficiency

Solving the problems of geofields recovery always requires high memory capacities and time expenditures. In spite of computer science rapid development the problem of numerical efficiency of methods and algorithms of geofield recovery is still urgent. Modern computer resources allow processing huge arrays of spatial data however calculation time may be still unallowably large. Use of high-performance algorithms allows reducing calculation time in tens or even hundred times. Search algorithms using spatial indexing may be referred to such algorithms [6].

Known methods of spatial indexing have their disadvantages – they do not take into account specific character of geofield digital model presented in the form of regular network.

Traditionally, geofield is recovered in the following way:

- regular network cell is chosen;
- *points* of initial set being in the circle of specified radius form recoverable regular network node are searched;
- value of recoverable node of regular network is calculated;
- if regular network is not recovered the first stage is fulfilled.

The authors developed the original algorithm of geofield recovery taking into account regular network specific character which is actually a raster.

The matter of suggested algorithm consists in the fact that *nodes* of regular network being in the circle of specified radius *are calculated* for each point of initial set of data.

Having specified the radius and using coordinates of the points of initial data set as coordinates of search centre of circle all cells of regular network being in specified radius from the point of initial data set may be calculated without significant time expenditures. To calculate geofield value in a cell the same calculations should be carried out for all points of initial data set.

Information whether a point of initial data set takes part in calculation of network cell value (be in specified radius) is held in two-dimensional matrix with the dimension equal to dimension of recoverable regular network. Each cell of the matrix corresponds to the regular network cell and contains a bit vector a number of elements of which equals to a number of initial data. Each element of such vector contains information on the fact whether the point of initial data set with appropriate index takes part in calculations or not.

Thus, algorithm of geofield recovery may be briefly described in the following way:

- indices of boundary cells of the circle with specified radius are determined by the algorithm of Brezenheim [7];
- a point of initial data set is selected;
- cells of recoverable regular network moved away from the point of initial data set to the specified radius are determined;
- serial number of the point of initial data set is stored for each cell;
- if not all the points of initial data set are passed – stage 2;
- the value of field degree is calculated for each cell of regular network.

The suggested algorithm of calculation of geofield recovery allowed reducing calculation time more than twice at large number of initial points (Fig. 2).

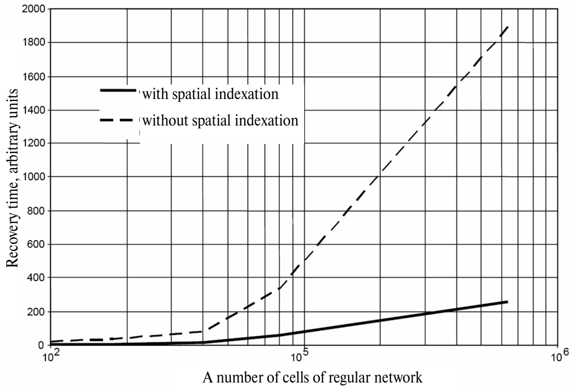

Fig. 2. *Comparison of numerical efficiency of geofield recovery algorithms*

Conclusion

Problems occurring at geofield recovery may be divided into methodological and technological. The problem of selection and justification of recovery method, its parameters, their values, models, function etc. are referred to the first type. These problems overcoming depends to a large extent on theoretical and practical knowledge of the researcher himself. Insufficient training does not allow him to use fully mathematical apparatus and tools of modern facilities for solving the given problems.

Technological problems, in their turn, may be also divided into problems of numerical efficiency of the applied methods and problems of integration of heterogeneous spatial data. The carried out analysis of the existing systems showed that the most perspective ones when solving the problems of geofield recovery are geoinformation systems having a proper set of functions for implementing spatial analysis of geofields. In the range of geoinformation approach the authors developed the approach allowing refining the results of geofield recovery at the stage of post-processing relying on the data which can not be used in traditional methods of geofield recovery.

Investigations of geofield recovery algorithm suggested by the authors showed its high efficiency at a great number of initial data. Taking into account the above said the conclusion may be made on the fact that application of spatial indexing algorithms is obligatory while developing geofield recovery methods.

The given approaches were successfully used in program package SurfMapper developed by the authors.

The work was implemented with the support of RFBR grant № 06-07-89190.

REFERENCES

1. Aronov V.I. Methods of constructing maps of geological-geophysical features and geometrization of oil and gas pool on computer. – Moscow: Nedra, 1990. – 301 p.
2. Kanevskiy M.F., Demyanov V.V. Introduction into methods of analyzing environmental data // Problems of environmental and natural resources. – 1999. – № 11. – P. 4–11.
3. Kovin R.V., Markov N.G. Geoinformation technologies for analyzing two-dimensional geofields. – Tomsk: Tomsk University Press, 2006. – 166 p.
4. Burrough P.A., McDonnell R.A. Principles of Geographical Information Systems. – Oxford: Oxford University Press, 1998. – 333 p.
5. Gousie M.B., Williams G., Agnitti T., Doolittle N. CompSurf: An Environment for Exploring Surface Reconstruction Methods on a Grid // Computers & Geosciences. – 2003. – № 9. – P. 1165–1173.
6. Laslo M. Computational geometry and computer graphics on C++ / Translated from English. – Moscow: «BINOM» Press, 1997. – 304 p.
7. Preparata F., Sheimos M. Computational geometry: Introduction. – Moscow: Mir, 1989. – 478 p.

Received on 16.10.2007

UDC 504.064(4)

APPLICATION OF JOHNSON DISRIBUTION TO THE PROBLEM OF AEROSPACE IMAGES CLASSIFICATION

Yu.B. Burkatovskaya, N.G. Markov, A.S. Morozov, A.P. Serykh*

Tomsk Polytechnic University
*Tomsk State University
E-mail: julya@mail.tsu.ru

Solving the problem of aerospace images classification it was suggested to approximate distribution density of image characteristics by Johnson distribution. The possibilities of such approach were investigated and its availability was shown.

Introduction

Aerospace images (AI) are an important source of data on Earth surface state and used, for example, for fire recognition, estimation of territory ecological state etc. Solving the problem of AI automated interpretation is mainly connected with the development of new methods, algorithms and software tools for such image classification [1–3]. The most widespread approach in this case is application of the theory of image statistic recognition [4] using various classifiers. Classification of earth surface types on AI may be carried out using both parametric and nonparametric classifiers. However, the most part of these classifiers do not satisfy increasing requirements to accuracy and speed of classification therefore, the task of development of new high-performance methods and approaches to AI classification is urgent.

In this paper the approach allowing solving the problem of AI classification using Johnson distribution is developed. The results of studying such approach efficiency are given.

Problem of aerospace image classification

In general case AI consists of N points and includes K data channels. According to electromagnetic spectrum region used for AI formation all AI are divided into ultraviolet, optical, infrared and radar [1]. At present panchromatic (they are also called single-channel) AI are used for analyzing earth surface dynamics. They include one more or less wide spectral region as well as multisensor images (called multi-channel or multiregion) including several layers (channels) obtained simultaneously in various narrow spectrum ranges. Original AI are represented both in widespread formats (for example, Windows Bitmap) and in special formats of geoinformation systems (for example, ER Storage, GeoTiff, Imagine Image) for further processing and interpretation.

Let this or that method is fit for determining that studied AI has M types (classes) of earth surface such as water, vegetation, coniferous forest etc. A sample (AI pixel group) called learning corresponds to each class. Subject to this it is necessary to solve the problem of AI classification that is to refer each point (pixel) of an image to one of classes M.

Parametric approach (it is based on hypothesis of normal distribution of feature values) is often used at AI classification with training or methods of nonparametric statistics are used within nonparametric approach or neuronet classification. In the first approach kernel rate