Web-based Surface Fitting System for Geological Field Data using Free and Open Source Software

Nonogaki, S.,[1,] Nemoto, T.,[2] and Masumoto, S.[2]
[1]Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology
Central 7, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8567, Japan, E-mail: s-nonogaki@aist.go.jp
[2]Department of Geosciences, Graduated School of Science, Osaka City University, Osaka, Japan

Abstract
The role of Web services has become more and more important for managing geological information. With the help of web services, one can find geological information relatively easily. However, one may want to process geological information on the web. The purpose of this study is to utilize geological information on the Web. The authors developed a Web-based surface fitting system for geological field data using Free and Open Source Software. In this system, the surface fitting is performed based on smoothing algorithm with bi-cubic B-spline function. Available geological field data are elevation data and dip data. In this paper, we describe the detail of the system and give examples of surface fitting. As a result of this study, it is possible to generate a distribution map of the data, a contour map of the calculated surface, and a Digital Elevation Model from geological filed data on the Web. In conclusion, the system promotes the utilization of geological information on the web.

1. Introduction

The work here is of utilization of geological filed data on the Web. Geological filed data are quite useful for analyses of subsurface structure. Especially borehole cores, geochemical concentrations, and trend information play an important role to understand subsurface condition. In general, geological filed data are observed irregularly in geological survey. It is hard to understand subsurface condition directly from the data. In order to utilize geological field data in practical analyses, we usually perform a surface fitting and generate a grid data like DEMs (Digital Elevation Models). A considerable number of studies have been conducted on algorithms for surface fitting in the field of geology as well as in many other fields (e.g. Matheron, 1963, Dubrule and Kostov, 1986, Inoue, 1986 and El Abbass et al., 1990). In addition, various systems have been developed based on the algorithms. However, most of systems work in stand-alone PC. It means that we need to make a suitable environment for installing the systems on our own PC. Further, these systems are generally very expensive. These become a big obstacle to utilization of geological field data as well as to dissemination of the systems. Recently the concern with a management of geological field data on the Web has been growing. However, most of management systems just have a capability of storing the data. Very few systems have a capability of processing the data. In this study, in order to promote the utilization of geological information, we have developed a new system for surface fitting with geological field data on the Web. The system has been implemented by FOSS (Free and Open Source Software). Bi-cubic B-spline function is adopted as a surface for fitting to the data. As a result, it is possible to generate a DEM from geological field data without preparing a new environment of PC. The DEM is saved as an ASCII text file that can be utilized in various GISs or Web-GISs. In conclusion, the new system leads to a higher utilization of geological information.

2. Algorithm for Surface Fitting

There are many surface fitting algorithms with B-spline for geological field data (e.g. Bhattacharyya, 1969 and Inoue, 1986). However, most of them use only equality constraints from elevation to determine a shape of surface. On the other hand, Nonogaki et al., (2008) proposed a unique algorithm that uses constraints from slope information and inequality constraints from elevation as well as normal equality constraints from elevation. Use of several types of constraints must improve reliability of the result. Thus, we adopt this algorithm in surface fitting for Web-GIS. As the purpose of this paper is concerned, it is not necessary to discuss the algorithm in details. In this chapter, we describe only important parts of the algorithm. Refer to Nonogaki et al., (2008) for more details on the algorithm.

2.1 Geological Field Data Available in Surface Fitting

Two types of geological filed data are used as constraints on surface fitting. One is elevation data that constrain a height of the surface. Suppose that a surface can be expressed in $z = f(x, y)$ and that an elevation z_p is obtained at a point (x_p, y_p). A possible constraint from the point is as follow:

$$f(x_p, y_p) - z_p = 0$$

Equation 1a

$$f(x_p, y_p) - z_p < 0$$

Equation 1b

$$f(x_p, y_p) - z_p > 0$$

Equation 1c

Constraint (1a) is used in cases that the surface passes through the point. Constraint (1b) and (1c) is used in cases that the surface passes under the point and over the point respectively. The three types of constraints are distinguished by an optional parameter l_p, which is defined by: $l_p = 0$ for the data (1a), $l_p < 0$ for (1b), and $l_p > 0$ for (1c). We call the data given by (1a) as "equality elevation data" and call the data given by (1b) and (1c) as "inequality elevation data". The other is dip data that constrain a trend of the surface. Let ϕ be an azimuth direction of maximum slope of the surface. ϕ is measured clockwise from north. Let θ be a slope angle. Suppose that the dip data (ϕ_q, θ_q) is obtained at a point (x_q, y_q). A possible constraint from the point is as follow:

$$f_x(x_q, y_q) + \sin\phi_q \tan\theta_q = 0$$

Equation 2a

$$f_y(x_q, y_q) + \cos\phi_q \tan\theta_q = 0.$$

Equation 2b

2.2 Bi-Cubic B-Spline Function

Bi-cubic B-spline function is used to approximate a surface. Let $\Omega = [x_{min}, x_{max}] \times [y_{min}, y_{max}]$ be a rectangular domain in x-y plane. Dividing Ω into $M_x \times M_y$ sections, the surface in Ω can be expressed in a quadratic form:

$$f(x, y) = \sum_{i=1}^{M_x+3} \sum_{j=1}^{M_y+3} c_{ij} \, N_i(x) N_j(y)$$

Equation 3

Where $N_i(x)$ and $N_j(y)$ are normalized cubic B-spline bases with respect to x and y respectively. c_{ij} are the constants for the products $N_i(x)N_j(y)$.

2.3 Determination of Optimal Surface

There may be many feasible surfaces that satisfy the constraints from geological field data. We assume that an optimal surface must be the smoothest one among the feasible surfaces. Based on the exterior penalty function method, we find the surface that minimizes an augmented objective function:

$$Q(f) = \{m_1 J_1(f) + m_2 J_2(f)\} + \alpha \{R_H(f) + \gamma R_D(f)\}$$

Equation 4

Where $J_1(f)$ and $J_2(f)$ are functional that evaluate the flatness and smoothness of the surface respectively. $R_H(f)$ and $R_D(f)$ are functional that evaluate the goodness of fit to elevation data and dip data respectively. m_1, m_2, α, and γ are parameters that adjust not only weight balances but also dimensions between four functionals.

3. Surface Fitting System

Several studies have been made on development of "stand-alone" system for management of geological field data (e.g. McInerney et al., 2005 and Mathers and Kessler, 2008). However, there has been no study that tried to develop a system for surface fitting with geological field data on the Web. In this study, we have developed a "Web-based" surface fitting system for geological field data. All the system for surface fitting has been implemented by FOSS. The main aim of the system is to generate a DEM using geological field data on the Web. In this chapter, we describe the system configuration and the procedures for surface fitting.

3.1 System Configuration

CPU used in this system is Intel(R) Xeon(R) Processer X3470 (Clock speed: 2.93GHz). Amount of RAM is 16GB. The system works on LINUX OS environment. It consists of Web server, mapping tool, and some other software. Table 1 shows the software components making up the system. GMT is used for generating 2D maps such as distribution maps of geological field data and contour maps of calculated surface. Python and FORTRAN are used for numerical processing such as calculating

Table 1: Software components of the system

Items	Name	Version
Operating System	CentOS	5.4 (kernel 2.6.18)
Web Server	Apache	2.2.3
Mapping Tool	GMT	4.5.3
Other Software	Python	2.4.3
	GCC-GFORTRAN	4.1.2
	PHP	5.1.6
	JAVA	1.6.0

statistics of the data, surface fitting and generating 3D map. PHP and JAVA are used for passing numerical parameters and for manipulating behaviors of check boxes and buttons on browser screens.

3.2 Procedures for Surface Fitting

There is no need to prepare any special environments on our own PC. All operations can be done for free through the Web browser such as Internet Explorer or Mozilla Fire Fox. However, a basic authentication is needed to access the system. There are five steps to perform a surface fitting and to generate a DEM for geological field data. Figure 1 shows an example of operation screens during a surface fitting with elevation data modified from Table 5.11 in Davis (1986).

Step1: Upload the Field Data

First of all, geological field data must be uploaded into the system. Elevation data and dip data are available as stated above. In the case of elevation data, information of each data is described in one line with a format (*id*, *x*, *y*, *z*, *l*). Additionally, the final line must be a specific data (0, 9e9, 9e9, 9e9). In the case of dip data, information of each data is described in one line with a format (*id*, *x*, *y*, *z*, ϕ, θ). A final line must be a data (0, 9e9, 9e9, 9e9, 9e9, 9e9). Maximum number of field data is 20,000,000. The example of data files is shown in the browser screen (Figure 1a).

Step2: Generate a Distribution Map

In this step, in order to obtain criteria for surface fitting, with the help of a distribution range of geological field data, we generate a distribution map of the data (Figure 1b). The required information is as follow: (i) geographical coordinates of the map edge, (ii) intervals of grid lines that will be drawn on the map, and (iii) horizontal and vertical size of the map in cm. File name for the distribution map is automatically given by the system. The distribution map can be saved as PNG (Portable Network Graphics) format or PS (Post Script) format.

Step3: Determine an Optimal Surface

In this step, with the help of the distribution map, we give a set of parameters for surface fitting (Figure 1c). The required information is as follow: (i) calculation range (domain Ω), (ii) number of small cells within the calculation range (number of sections inside domain Ω: M_x and M_y), and (iii) some other parameters. Maximum number of small cells with respect to each axial direction (M_x and M_y) is 300. Dip data are not necessarily in calculation. However, text box for dip data must be filled. A string "dummy" is given by default. Parameter gamma is available when dip data are uploaded. There are two kinds of output files. One is a file for bi-cubic B-spline function of the optimal surface. It includes information for defining the equation (3). The other is a file for values of functionals in equation (4) such as $J_1(f)$, $J_2(f)$, $R_H(f)$ and $R_D(f)$. The names of each file are automatically given as well as that of the distribution map.

Step4: Generate a DEM

In this step, we generate a DEM for the calculated surface. When the system finishes the calculation, it shows a contour map of the calculated surface on the browser screen (Figure 1d). In addition, the system shows a link to another screen under the contour map. In the link, we can confirm not only a list of numerical values of functionals in equation (4) but also 3D visualization of the calculated surface (Figure 1e). Both the contour map and the list of numerical values can be saved as PNG format and PS format. Further, 3D visualization of the surface can be saved as VRML (Virtual Reality Modeling Language) format. Among the numerical values displayed on the browser screen, functionals $R_H(f)$ (Rh) and $R_D(f)$ (Rd) are very important for verifying the goodness of fit. These values are residual mean squares (RMS) between calculated surface and geological filed data that do not satisfy constraints. Square root of RMS is given by "Error" displayed on the right side of RMS. For example, in figure 1e, Rh is 2.76×10^{-4} and square root of Rh is 1.66×10^{-2}. With the help of numerical and visual information, we determine whether the result is feasible or not. After obtaining an acceptable result, we generate a DEM. The required information is as follow: (i) file name and file format for DEM, (ii) output range, and (iii) number of grids. The output range must be within calculation range. Maximum number of grids with respect to each axial direction is 5,000. There are two types of DEM formats: one is "Horizon Grid" format and the other is "GRASS ASCII" format.

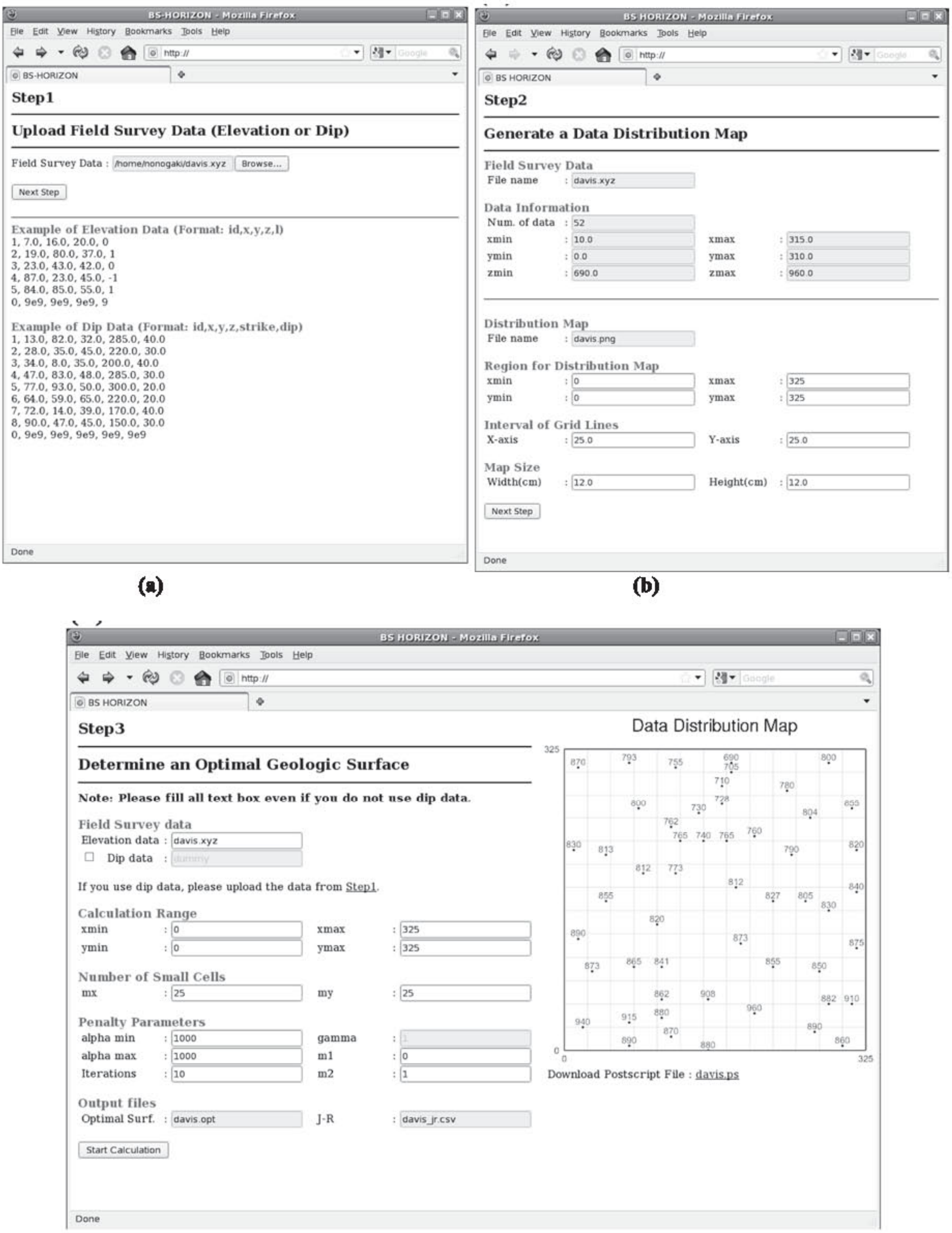
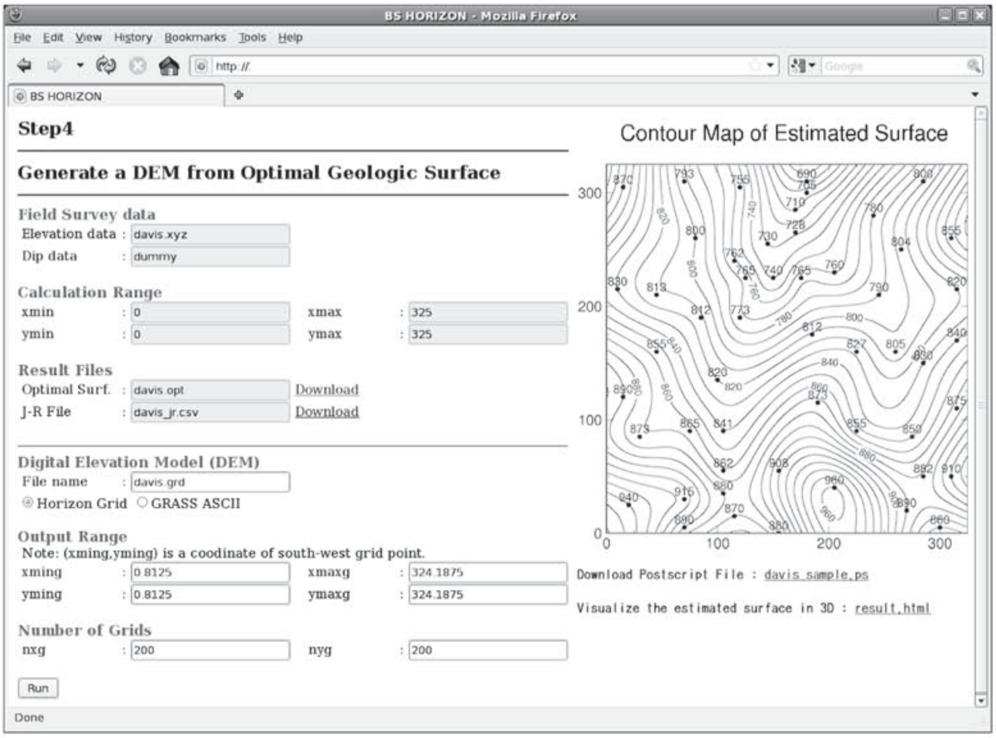
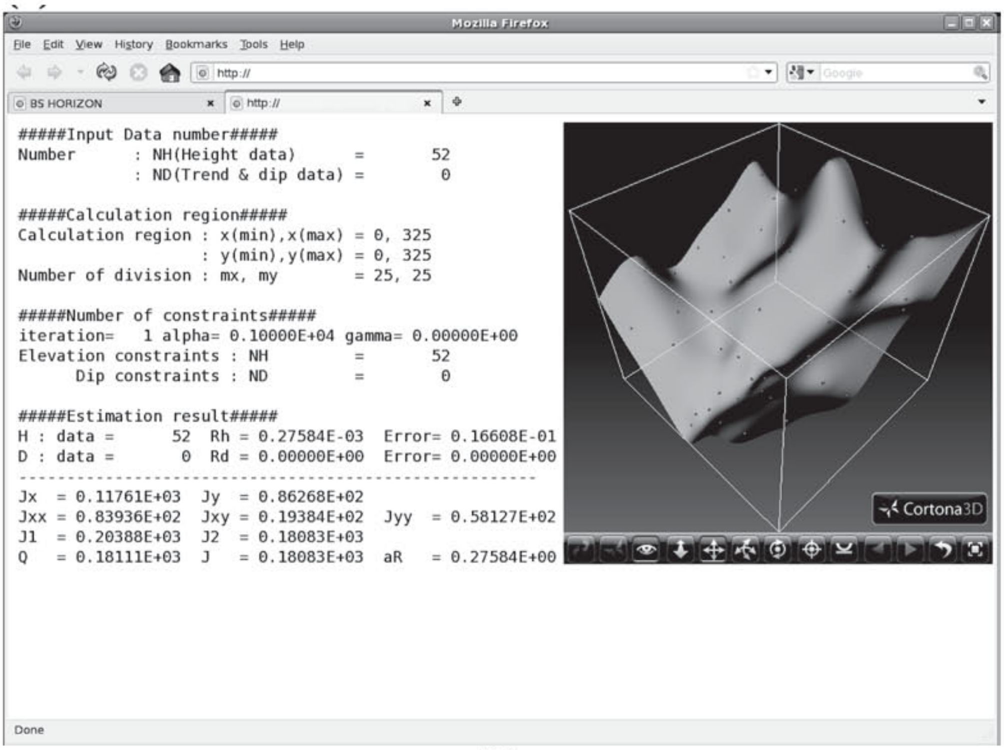


Figure 1: An example of operation screens. (a): screen for data upload, (b): screen for generation of distribution map, (c): screen for determination of an optimal surface

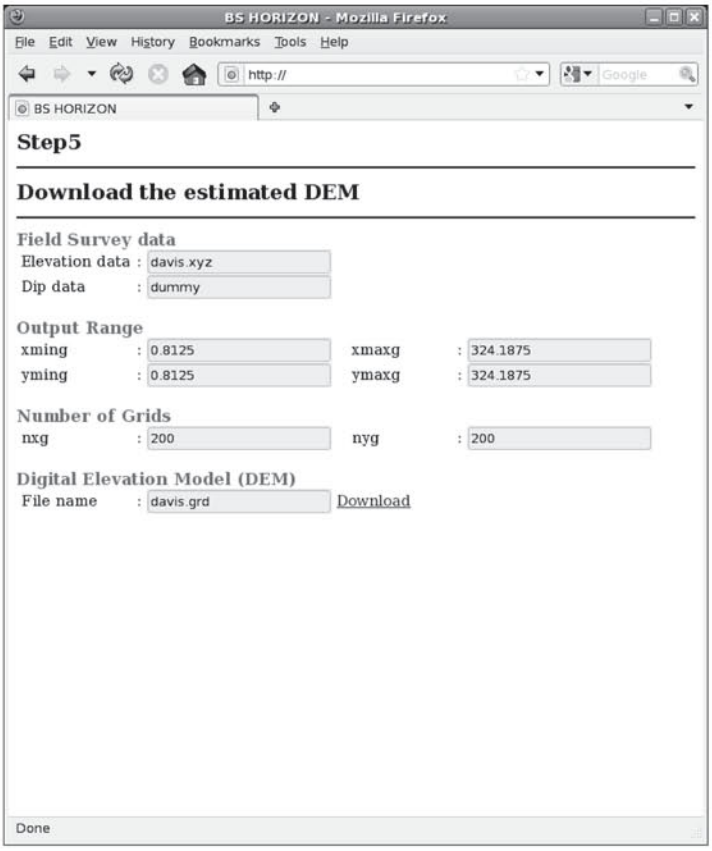


(d)



(e)

Figure 1: (d): screen for generation of DEM and
(e): screen for functional values and 3D visualization of the optimal surface



(f)

Figure 1: (f): screen for download of the DEM

Step5: Download the Result

In this step, we just save the DEM in a specified file after confirming the parameters of DEM generation (Figure 1f).

4. Examples

4.1 Surface Fitting with Elevation Data

Figure 2 shows an example of surface fitting with elevation data that provide constraints (1a), (1b), and (1c). Figure 2a shows a distribution map of the data. In figure 2, types of constraints are distinguished by symbols. A black dot means a location of the data that provides constraint (1a). A down-pointing triangle means a location of the data that provides constraint (1b). A up-pointing triangle means a location of the data that provides constraint (1c). Numerical value near the symbol is elevation at the data point. Domain Ω for surface fitting is [0, 100] × [0, 100]. Parameters used in calculation are as follows: $M_x = M_y = 10$, $\alpha_{min} = 1$, $\alpha_{max} = 10^2$, N_{ITR} (number of iteration) = 10, $m_1 = 0$, and $m_2 = 1$. Computation time is less than 1 second. In numerical evaluation, Rh is 2.79×10^{-4} and Error (square root of Rh) is 1.67×10^{-2}. Considering digits of elevation data, we can safely say that these values are sufficiently small. Figure 2b shows a contour map of the calculated surface. It is clear that the surface satisfies all types of data. As a result, we can conclude that the calculated surface is feasible.

4.2 Surface Fitting with Dip Data

Figure 3 shows an example of surface fitting with dip data. Domain Ω for surface fitting is [420000, 432500] × [5883000, 5897000]. Parameters used in calculation are as follows: $M_x = 125$, $M_y = 140$, $\alpha = 10^4$, $\gamma = 10^3$, $m_1 = 0$, and $m_2 = 1$. Computation time is 12.4 seconds. Figure 3a shows a contour map of the calculated surface. In figure 3a, a symbol of dip data is given by a long bar and short spike perpendicular to the long bar. Short spike means the azimuth direction of maximum slope. Numerical value in parenthesis near the symbol of dip data is slope angle. In numerical evaluation, Rh is 5.48×10^{-10} and square root of Rh is 2.34×10^{-5}. Rd is 3.92×10^{-10} and square root of Rd is 1.98×10^{-5}. Considering digits of elevation data and dip data, we can conclude that the result is feasible. Figure 3b shows a 3D visualization of the surface. An outline of folding structure is easily confirmed.

4.3 Application to 3D Geological Modeling

As mentioned in chapter 3, the surfaces derived from this system can be saved as GRASS ASCII format, which is one of the formats supported by GDAL (Geospatial Data Abstraction Library). Thus, it is possible to utilize the calculated surfaces in various GISs or Web-GISs.

Figure 4 shows an example of use of the surfaces in 3D geological modeling based on borehole data. In recent year, there are several Web-based systems for 3D geological modeling (e.g. Nemoto et al., 2003 and Masumoto et al., 2008). The developed system has a potential of interoperating with those systems.

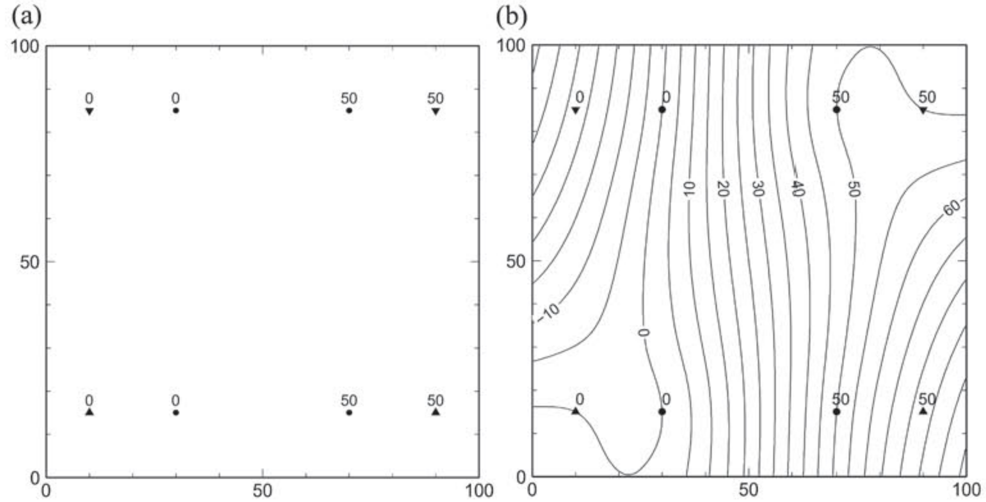


Figure 2: An example of surface fitting with elevation data. (a) distribution map of the data, (b) contour map of the calculated surface

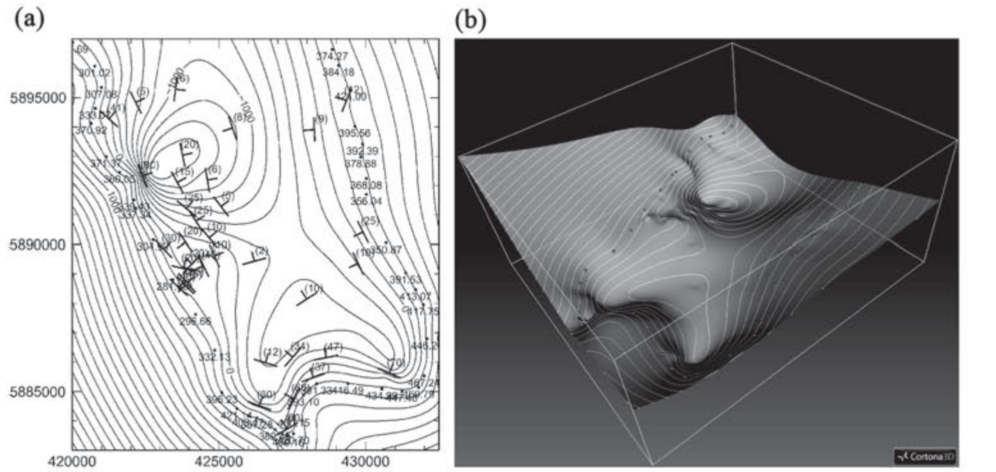


Figure 3: An example of surface fitting with dip data. (a) contour map of the calculated surface, (b) 3D visualization of the surface by VRML

Figure 4: An example of use of the surface in 3D geological modeling based on borehole data with GRASS GIS

5. Conclusions

The Web-based surface fitting system for geological field data has been developed using FOSS. This system has an advantage that inequality elevation data and dip data are available as well as normal equality elevation data. It contributes to higher-reliable surface fitting with geological filed data. Further, this system enables us to generate a DEM from geological field data for free without preparing any new environments of PC. The DEM can be saved not only as Horizon format but also as GRASS ASCII format. Therefore, it is easy to utilize the surfaces derived from this system in various GISs or Web-GISs. This promotes the utilization of geological information on the Web. At this moment, processing speed of the system is not so fast. For example, surface fitting with several thousand geological field data takes about 820 seconds in the case of $M_x = M_y = 300$, about 130 seconds in the case of $M_x = M_y = 200$, and about 7 seconds in the case of $M_x = M_y = 100$.

Processing speed may be a limiting factor in dissemination of the system. The topic of our further study is to speed up process using multithreaded program. Moreover, the implemented systems adopt an original rule in parameter passing for surface fitting. It may be a limiting factor not only in dissemination of the system but also in interoperation with other systems. In order to solve this problem, we will adopt OGC standards like WPS (Web Processing Service) in future enhancements to the system.

Acknowledgement

The authors would like to thank the anonymous reviewers for their useful comments. Thanks also to Prof. Venkatesh Raghavan for helpful suggestions. This study was supported by KAKENHI (22500094; Grant-in-Aid for Scientific Research by Japan Society for Promotion of Science).

References

Bhattacharyya, B. K., 1969, Bicubic Spline Interpolation as a Method for Treatment of Potential Field Data. *Geophysics*, 34, 402-423.

Davis, J. C., 1986, Statistics and Data Analysis in Geology. 2nd edition, John Wiley & Sons, Inc., New York.

Dubrule, O. and Kostov, C., 1986, An Interpolation Method Taking into Account Inequality Constraints: I. Methodology. *Mathematical Geology*, 18, 33-51; II. Practical Approach. *Mathematical Geology*, 18, 53-73.

El Abbass, T., Jallouli, C., Albouy, Y., and Diament, M., 1990, A Comparison of Surface Fitting Algorithms for Geophysical Data. *Terra Nova*, 2, 467-475.

Inoue, H., 1986, A Least-Squares Smooth Fitting for Irregularly Spaced Data: Finite-element approach using the cubic B-spline basis. *Geophysics*, 51, 2051-2066.

Masumoto, S., Nonogaki, S., Ninsawat, S., Iwamura, S., Sakurai, K., Raghavan, V., Nemoto, T. and Shiono, K., 2008, Development of Prototype System for Three Dimensional Geologic Modeling Based on Web-GIS. *Proc. GIS IDEAS2008*, 83-88.

Matheron, G., 1963, Priciple of geostatistics. *Economic Geology*, 58, 1246-1266.

Mathers, S. J., and Kessler, H., 2008, GSI3D the software and Methodology to Build Systematic Near-Surface 3-D Geological Models - Version 2.6, British Geological Survey Open Report, OR/08/064. British Geological Survey, Nottingham.

McInerney, P., Guillen, A., Courrioux, G., Calcagno, P., and Lees, T., 2005, Building 3D Geological Models Directly from the Data? A New Approach to Broken Hill, Australia. *Digital Mapping Techniques '05 - Workshop Proceedings: U.S. Geological Survey Open-File Report 2005-1428*, 119-130.

Nemoto, T., Raghavan, V., Masumoto, S. and Shiono, K., 2003, Development of an Online Spatial Information System for Geologic Modeling - SISGeM. *Geoinformatics*. 14, 63-68.

Nonogaki, S., Masumoto, S. and Shiono, K., 2008, Optimal Determination of Geologic Boundary Surface using Cubic B-Spline. *Geoinformatics*. 19, 61-77.