Tectonophysics 633 (2014) 232–245

Contents lists available at ScienceDirect

Tectonophysics

journal homepage: www.elsevier.com/locate/tecto

Post-processing for uncertainty reduction in computed 3D geological models

Jorge Kazuo Yamamoto [a], Katsuaki Koike [b], Antonio Tadashi Kikuda [a], Ginaldo Ademar da Cruz Campanha [a,*], Arthur Endlen [a]

[a] Institute of Geosciences, University of Sao Paulo, Sao Paulo, Brazil
[b] Graduate School of Engineering, Kyoto University, Kyoto, Japan

ARTICLE INFO

Article history:
Received 31 March 2014
Received in revised form 30 June 2014
Accepted 13 July 2014
Available online 22 July 2014

Keywords:
Multiquadric equations
Computer geological models
Uncertainty
Resampling
Post-processing

ABSTRACT

Computer aided geological modeling is always subject to uncertainty because it results from interpolation of sample data. Sample data are subject to uncertainties coming from location error, stochastic nature of the geological variable and imprecise knowledge. The lack of knowledge due to limited sampling data can be viewed as the main source of uncertainty. Besides, in terms of computer modeling another source of uncertainty is the shape complexity shown by geological units. Considering that just few drill holes are available the constructed model is going to present large uncertainties. However, an uncertain model cannot be used for a precise interpretation of geological setting. To overcome this problem, this paper proposes a novel approach for uncertainty reduction and enhancement of the reliability of the geological model. This procedure is based on resampling the computed model and post-processing using this new sample. Results of post-processing are much better than the first interpolated geological model. Besides, this procedure can be repeated until the final result is considered acceptable within the limited extent of the sample data.

1. Introduction

3D geological models built from sampling data essentially contain uncertainties. Moreover, the uncertainty increases with decreasing sample size. It is important to note that additional data not always means further uncertainty reduction depending on location of the new data. This paper concerns building 3D geological models from geologic logging of drill holes that are irregularly spaced in the horizontal and vertical directions. According to Welmann et al. (2010, p. 142), sources of uncertainty associated with geological data can be classified in three categories: (1) imprecision and measurement error; (2) stochastic nature of the geological variable and (3) imprecise knowledge. In the first category we have uncertainty coming from location error. For example, drill hole deviations will cause errors on determination of three dimensional coordinates and consequently on thickness estimation and geologic contact definition. The last two categories are related to the lack of knowledge because insufficient sampling. In the early stages of mineral exploration the lack of knowledge is the main source of uncertainty because few data are available. Considering modern drilling and surveying techniques, the location error is insignificant. However, the geologist must provide the best geological model in the light of available data. It means giving a result as certain as possible. In this paper, 3D geological models are built from the interpolation of types of a categorical variable such as facies, formations, rock types, and degree of alteration. Interpolation of different types of a categorical variable is possible after transforming them into indicator functions. Thus, for each type its indicator function is interpolated. Because we have a number of interpolated indicator functions corresponding to the different types, the most likely type is given by the greatest indicator value. The variance associated with the greatest indicator value is the measure of uncertainty. The resulting interpolated geological model will present uncertainties closer to geological boundaries (contacts and faults). Evidently, uncertainties will be large when this model is computed from few samples, and this uncertain model is unacceptable even considering the lack of knowledge as given by the sample. A computed 3D geological model is unacceptable when it presents artifacts coming from interpolation of few data. In this paper, we are aimed to improve the accuracy of geological model and propose to resample the computed model by using this new sample to post-process only uncertain zones. Moreover, the post-processed geological model can be resampled and post-processed again. Theoretically, this process can be implemented repeatedly. Usually, the resulting model after the second resampling is more accurate than the first interpolated model. Besides, post-processing improves the geological continuity of formations.

2. Measuring uncertainties in geological modeling

Wellmann et al. (2010, p. 143–145) proposed a method to incorporate uncertainty in geological models. This method is based on five-step

* Corresponding author. Tel.: +55 11 30913946; fax: +55 11 30914258.
E-mail address: ginaldo@usp.br (G.A.C. Campanha).

http://dx.doi.org/10.1016/j.tecto.2014.07.013
0040-1951/

procedure, in which the first concerns construction of the initial model using all available data (contacts, faults and orientation measurements); since input data are subject to uncertainties the proposed method assigns probability distributions (normal distribution for gradual contact; uniform distribution when direct contact is missing and discrete distribution when the position of correct contact is unknown); in the next step different input data sets are drawn from assigned probability distributions; simulation of multiple geological models based on new input data sets; and finally visualization and post-processing of results. Because there are multiple realizations of the geological model, uncertainties are viewed by displaying all realizations simultaneously (Wellmann et al., 2010, p. 145). Besides that, for complex geological settings, Wellmann et al. (2010, p. 146) proposed to use indicator function that is one if location x belongs to a given formation F and zero otherwise.

$$I_F(x) = \begin{cases} 1\ for x \in F \\ 0\ for x \notin F \end{cases}$$

According to these authors, because there are multiple realizations, the probability of the formation F occurring at location x can be computed as the mean value of all indicator functions.

$$P_F(G) = \sum_{k \in n} \frac{I_{F_k}}{n}$$

Wellmann et al. (2011) suggested a further improvement on the former method (Wellmann et al., 2010, p. 748–750) that is the use of the concept of information entropy after Shannon (1948). They use the same five-step procedure (Wellmann et al., 2010, p. 143–144) in which added the computation of the information entropy H associated with the mean value of all indicator functions.

$$H = -\sum_{i=1}^{N} p_i \log p_i. \quad (1)$$

where N is total of possible outcomes and the logarithm is to base 2 (Wellmann et al., 2011, p. 748).

Silva and Deutsch (2012, p. 307-1) presented an approach for modeling multiple rock types with a measure of uncertainty that gives the confidence for a rock type prevailing at an unsampled location. This method is based on distance function, which is negative when it is within a domain and positive when outside (Silva and Deutsch, 2012, p. 307-1). According to this procedure, available samples are coded as indicator functions:

$$I_k(u_\alpha) = \begin{cases} k\ if\ u_\alpha \in domain\ k \\ 0\ if\ u_\alpha \notin domain\ k \end{cases}.$$

Now, the indicator function is the input of a function F that returns the distance (Silva and Deutsch, 2012, p. 307-2):

$$F(I_k(u_\alpha)) = dF_k(u_\alpha) = \begin{cases} -d\ if\ I_k(u_\alpha) = k \\ +d\ if\ I_k(u_\alpha) = 0 \end{cases}.$$

Considering that we have K domains we will have K distance functions on sample locations. Distance functions are interpolated K times for an unsampled location (Silva and Deutsch, 2012, p. 307-2). The most likely domain at an unsampled location u_o is given by the minimum distance function (Silva and Deutsch, 2012, p. 307-3):

$$I_k^*(u_o) = F^{-1}\left(\min\left\{dF_1^*(u_o, l_1), dF_2^*(u_o, l_2), \ldots, dF_K^*(u_o, l_K)\right\}\right).$$

As a measure of uncertainty, Silva and Deutsch (2012, p. 307-3) proposed to define a U coefficient as the ratio between variance of all estimated distances but the minimum ($dF_m(u_o)$) and the total variance:

$$U(u_o) = \frac{Var\left[dF_1(u_o), \ldots, dF_{m-1}(u_o), dF_{m+1}(u_o), \ldots, dF_K(u_o)\right]}{Var\left[dF_1(u_o), \ldots, dF_{m-1}(u_o), dF_m(u_o), dF_{m+1}(u_o), \ldots, dF_K(u_o)\right]}$$

where $dF_m(u_o)$ is the minimum interpolated distance. The range of U coefficient is between zero and one, where zero means no uncertainty and one large uncertainty (Silva and Deutsch, 2012, p. 307-3). According to these authors, the U coefficient can be calculated at all interpolated locations representing a measure of uncertainty.

Yamamoto et al. (2012) proposed a different approach to derive uncertainty associated with the geological model. Actually, the geological model results from the interpolation of types of a categorical variable composed of K types. Thus, available data are transformed into K indicator functions (Yamamoto et al., 2012, p. 147) that are used to interpolate unsampled locations. The indicator function for the k[th] type is:

$$I(x_i; k) = \begin{cases} 1\ if\ type\ k\ is\ present\ at\ location\ x_i \\ 0\ if\ type\ k\ is\ not\ present\ at\ location\ x_i \end{cases}. \quad (2)$$

As the most suitable interpolation method, Yamamoto et al. (2012, p. 148) proposed to use multiquadric equations (Hardy, 1971, p. 1907) instead of indicator kriging. Although indicator kriging is the most common approach for interpolation of indicator functions, this method requires K indicator semivariograms. However, this is very difficult to obtain them when we have types presenting few data points and consequently few pairs for semivariogram computation (Yamamoto et al., 2012, p. 147).

The indicator function for the k[th] type can be interpolated at an unsampled location x_o based on n neighboring indicator values as (Yamamoto et al., 2012, p. 148):

$$I^*(x_o; k) = \sum_{i=1}^{n} w_i I(x_i; k). \quad (3)$$

This equation was proved to be equivalent to the original multiquadric equations of Hardy (1971, p. 1907) by Yamamoto and Landim (2013, p. 110–111).

The weights $\{w_i, i = 1, \ldots, n\}$ are the solution of a system of linear equations (Yamamoto et al., 2012, p. 148):

$$\begin{cases} \sum_{j=1}^{n} w_j \varphi\left(x_j - x_i\right) + \mu = \varphi(x_o - x_i)\ \text{for}\ i = 1, n \\ \sum_{j=1}^{n} w_j = 1 \end{cases}. \quad (4)$$

For interpolation of indicator functions the multiquadric kernel gives good results. The multiquadric kernel is:

$$\varphi(x) = \sqrt{|x|^2 + C} \quad (5)$$

where $|x|$ is the norm of a vector in $\mathbf{R}^n$ and C is a positive constant. This constant is known as shape parameter and the accuracy of multiquadric interpolation depends on this parameter (Bayona et al., 2011, p. 7384–7385).

The variance associated with the k[th] interpolated type can be computed as (Yamamoto et al., 2012, p. 149):

$$S_o^2(x_o; k) = \sum_{i=1}^{n} w_i \left[I(x_i; k) - I^*(x_o; k)\right]^2. \quad (6)$$