CrossMark

Entropy-Based Weighting in One-Dimensional Multiple Errors Analysis of Geological Contacts to Model Geological Structure

Weisheng Hou[1,2] · **Chanjie Cui**[1,2] ·
Liang Yang[1,3] · **Qiaochu Yang**[1,2] · **Keith Clarke**[4]

Abstract In each step of geological modeling, errors have an impact on measurements and workflow processes and, so, have consequences that challenge accurate three-dimensional geological modeling. In the context of classical error theory, for now, only spatial positional error is considered, acknowledging that temporal, attribute, and ontological errors—and many others—are part of the complete error budget. Existing methods usually assumed that a single error distribution (Gaussian) exists across all kinds of spatial data. Yet, across, and even within, different kinds of raw data (such as borehole logs, user-defined geological sections, and geological maps), different types of positional error distributions may exist. Most statistical methods make a priori assumptions about error distributions that impact their explanatory power. Consequently, analyzing errors in multi-source and conflated data for geological modeling remains a grand challenge in geological modeling. In this study, a novel approach is presented regarding the analysis of one-dimensional multiple errors in the raw data used for model geological structures. The analysis is based on the relationship between spatial error distributions and different geological attributes. By assuming that the con-

✉ Weisheng Hou
houwsh@mail.sysu.edu.cn

Keith Clarke
kclarke@geog.ucsb.edu

1 School of Earth Sciences and Engineering, Sun Yat-sen University, B410 Dihuan Building, No. 135 Xingangxi Road, Guangzhou 510275, Guangdong Province, China

2 Guangdong Provincial Key Laboratory of Mineral Resource Exploration and Geological Processes, Guangzhou 510275, China

3 Present Address: Geological Sciences, Stanford University, Stanford, CA 94035, USA

4 Department of Geography, 1720 Ellison Hall, University of California, Santa Barbara, Santa Barbara, CA 93106-4060, USA

tact points of a geological subsurface are decided by the geological attributes related to both sides of the subsurface, this assumption means that the spatial error of geological contacts can be transferred into specific probabilities of all the related geological attributes at each three-dimensional point, which is termed the "geological attribute probability". Both a normal distribution and a continuous uniform distribution were transferred into geological attribute probabilities, allowing different kinds of spatial error distributions to be summed directly after the transformation. On cross-points with multiple raw data with errors that follow different kinds of distributions, an entropy-based weight was given to each type of data to calculate the final probabilities. The weighting value at each point in space is decided by the related geological attribute probabilities. In a test application that accounted for the best estimates of geological contacts, the experimental results showed the following: (1) for line segments, the band shape of geological attribute probabilities matched that of existing error models; and (2) the geological attribute probabilities directly show the error distribution and are an effective way of describing multiple error distributions among the input data.

1 Introduction

With the goal of knowing the true spatial distribution of geological conditions, three-dimensional geological modeling employs a series of techniques and three-dimensional structural or geostatistical models for applications such as geological mapping, civil engineering, reservoir appraisal, mineral targeting, and risk assessment (Bistacchi et al. 2008; Collon et al. 2015; Royse 2010; Wu et al. 2015). In such applications, error in the three-dimensional model can translate into serious consequences of risk, failure, or cost. To minimize the error, measurements of geological attributes are made at points, along transects, as depth fields, or as a two-dimensional mapped surface. Three-dimensional models of a subsurface or geological attribute field are then assembled based on the different datasets, such as borehole data, cross-sections, remotely sensed variables, and geological maps. These data help geologists to visualize, understand, and quantify the geological processes under way and the nature of the current status of the geological form that these processes have created. Ideally, all available multi-source data are reconciled together or conflated to construct a reliable three-dimensional geological model (Fasani et al. 2013; Kaufmann and Martin 2008; Li et al. 2013; Lindsay et al. 2012). However, each type of data carries its own uncertainties. Errors and uncertainties due to combining these pieces of data can be analyzed in addition to the uncertainties of the individual datasets. Combinatorial errors and uncertainty may, consequently, affect the precision of three-dimensional models and the results of any subsequent assessment or calculation. This problem is widely recognized to be a major challenge in each step of three-dimensional geological model creation (Caumon et al. 2009; Jessell et al. 2010; Jones et al. 2004; Li et al. 2013; Tacher et al. 2006; Wellmann and Regenauer-Lieb 2012).

In recent years, strategies for error and uncertainty assessments of raw data and model structures including topology have been proposed for different types of applications (Jones et al. 2004; Li et al. 2013; Lindsay et al. 2012, 2013; Tacher et al. 2006; Thiele et al. 2016; Wellmann and Regenauer-Lieb 2012). A framework and quality detection technology for error analysis of three-dimensional geological modeling data has been presented (Zhu and Zhuang 2010). However, the assessment and accounting of the impact of errors during the process of constructing a three-dimensional geological model remain unclear because the errors in the input data are random, diversely distributed, and vague. Yet, existing geostatistical methods usually assume that all errors follow a single distribution. Therefore, it is necessary to build an aggregated theoretical model for multiple errors analysis that can reveal the complete impact of error and uncertainty on constructed three-dimensional geological models.

While constructing three-dimensional geological structural models, input data are usually transferred into geological contact information according to the rock type or other geological properties. Picking contacts from cuttings, cores, or logs is often performed by geologists or geophysicists. Many studies by petrophysicists describe methods to analyze errors or uncertainties of velocity and geological structures while processing seismic images (Fomel 1994; Fomel and Landa 2014; Pon and Lines 2005; Thore et al. 2002; Grubb et al. 2001). When constructing a three-dimensional geological structure, the impact of errors and uncertainties from geophysical interpretation, field observation, etc. on the three-dimensional structures should be a special concern. Therefore, this study discusses methods for error analysis of the input data, rather than error analysis approaches for picking contacts.

From the viewpoint of mapping, a novel approach is presented to simulate the impact of varying distributions of multiple errors in geological contacts obtained from different types of data sources for constructing three-dimensional geological structural models. Two assumptions are made: (1) each type of data has spatial errors regardless of data consistency, and its error follows one type of probability density function (pdf); and (2) the errors that exist in different data sources are independent, which means that input data may be provided by different geologists or agencies; therefore, Bayes' theorem cannot be directly utilized in aggregating errors. The proposed method is based on the relationship between a geological attribute (such as lithological character) distribution and the spatial error distribution. Additionally, the method is focusing on analyzing error distributions without multiple z-values for the same x and y location. This study aims to (1) construct the error distribution relationship between the spatial errors and their related geological attributes within geological databases and (2) construct an aggregate model for a one-dimensional multiple errors analysis model at cross-points. A cross-point is the intersection in three-dimensional space of planar or linear geological features. The proposed approach is novel in that (1) it addresses multiple errors in one workflow simultaneously and (2) the error distribution is generated by a stochastic variable around all the boundaries instead of only at the contact points. This provides a more realistic model of the actual pattern of error and uncertainty surrounding the features in a geological model, for example, along fault lines and contacts.

2 Related Research

Error in three-dimensional geological modeling is mainly attributed to three components: error in measurements in the input data, error generated during the modeling process, and errors caused by imprecise knowledge (Bárdossy and Fodor 2004; Mann 1993; Wellmann et al. 2010). Error from the input data is a type of observation error that exists in most raw data. Examples of measurement error of raw data include incorrect orientation of the structure and imprecise interpretation, such as the stratigraphic information of borehole cores or the geological interpretation along cross-sections using seismic data or borehole information. This study focuses on analyzing the distribution of this type of error. Error generated during the modeling process is a type of stochastic, and inherently random, error that shows up in the process of interpolation of the subsurface or is caused by combining data. The last category of uncertainty refers to the result of incomplete or imprecise knowledge of structural existence or persistence, general conceptual ambiguities, or the need for generalizations. Thus, for example, a sedimentary outcrop may be defined by visually interpreting vegetation patterns or by interpolating a boundary from a spatially distributed set of measures. The precise rock type may be poorly defined or rely on arbitrary definitions, such as percent sand. Geological contacts are often classified according to values of different geological properties, such as velocity, rock type, density, and electricity resistivity. In the process of seismic imaging, the issue of structural uncertainty in the domain of velocity estimation has been discussed previously (Pon and Lines 2005; Thore et al. 2002). Researchers also studied the impact of velocity uncertainties on migrated images and AVO attributes (Fomel and Landa 2014; Grubb et al. 2001).

This study only discusses methods for analyzing the spatial error or uncertainty that exists in input data. Except for the dynamic modeling method (Chauvin and Caumon 2015), input data are transferred into geological contacts and are extracted as spatial data, such as points and segments in most approaches, to construct three-dimensional geological structures. Hence, the first type of error manifested is the positional error of geological contacts or boundaries, which is usually regarded as a kind of spatial error. Therefore, positional error methods for simulating spatial error are used to analyze errors in geological data. Based on the stage of modeling, approaches to the analysis of error or uncertainty can be classified into two types. One type of error analysis methods for three-dimensional geological modeling usually focuses on the impact of variabilities introduced by humans or computationally during data collection, processing, and representation in the final model (Bond et al. 2012; Jones et al. 2004; Lindsay et al. 2012; Tacher et al. 2006; Thore et al. 2002). This type of solution strives to reduce the effects of the errors before uncertainties are integrated into a final model. The other type of methods relates to quantifying the impacts of errors and uncertainties on the final three-dimensional model (Caumon et al. 2009; Jessell et al. 2010; Jones et al. 2004; Tacher et al. 2006; Viard et al. 2011; Wellmann et al. 2010; Wellmann and Regenauer-Lieb 2012). These methods simultaneously evaluate the impacts of the input data and all possible three-dimensional models.

On the other hand, recent research has focused on the problem from the viewpoint of alternative spatial data models (Chilès et al. 2004; Shi 1994; Shi and Liu 2000; Wellmann et al. 2010; Zhang et al. 2009): the vector data model or the raster model.

When the input data and the three-dimensional final model are organized as a vector data model using points, lines, areas, and volume, the main objective is to simulate the random error distribution in the input data or in the modeling process and its impact on the final model based on the random error theorem. Based on the assumption of error independence between sampling points, the error characteristics of digitized geological input data are measured with aggregated statistics, such as the mean squared positional error, the mean directional error, the error interval, or the error ellipsoid. Errors related to interpolation are usually described by the semivariogram or covariance matrix (Bárdossy and Fodor 2004; Guillen et al. 2008; Tacher et al. 2006; Zhang et al. 2009). Approaches to the analysis of errors along curves or segments have been developed based on randomness in point error analysis. Assuming that errors of segment endpoints are independent and follow a normal distribution, error distributions along a line segment can be inferred by the error ellipsoid and its confidence body of arbitrary points based on different parameters, such as the ε-band (Shi 1994; Zhang et al. 2009), error bands (Caspary and Scheuring 1993; Dutton 1992; Zhang and Tulip 1990), G-band (Shi and Liu 2000), and H-band (Li et al. 2002). When the input data and three-dimensional final model are represented with raster models, such as voxel or grid models, the idea of error analysis is totally different. In this circumstance, geostatistical methods or information entropy-based methods are the best choice for error or uncertainty analysis of each vertex within the whole model space (Calcagno et al. 2008; Chilès et al. 2004; Guillen et al. 2008; Wellmann et al. 2010; Wellmann and Regenauer-Lieb 2012).

Another key problem that should be solved in this study is to aggregate multiple error distributions in input data. From the perspective of a statistician, this kind of problem is treated as combining prior and preposterior probabilities into a posterior probability based on the assumption of data independence or conditional independence (Caers and Hoffman 2006). Geological data, however, usually cannot satisfy this assumption. The approaches taken in previous studies to handle the aggregating probabilities for geological data are diverse because each method is proposed for a specific application. For example, Tarantola and Valette (1982) proposed the concept of conjunction and disjunction of probabilities for inverse problems. Journel (2002) presented the Tau model from a broad perspective, and Polyakova and Journel (2007) improved it by using the Nu model. Allard et al. (2012) reviewed most of the available techniques for aggregating probability distributions and found that methods based on the product of probabilities are preferred. However, a suitable method should be selected from these methods based on the product of probabilities for the specific problem in this study.

A significant improvement in the accuracy of a three-dimensional geological model is made by fusing multiple input data (Collon et al. 2015; Kaufmann and Martin 2008; Travelletti and Malet 2012). However, across, and even within, different kinds of raw data (such as borehole logs, user-defined geological sections, and geological maps), different types of positional error distributions may exist. For example, error along the contact between two formations in a drill hole or outcrops may be expressed by either a normal distribution with a standard deviation around an expected value, a lognormal distribution, or a discrete probability distribution (Davis 2002; Wellmann et al. 2010). Therefore, the assumption of one single distribution cannot reflect the impacts of all of the types of errors in the process of modeling and the final three-dimensional model.

3 Methodology

3.1 Overview

As mentioned above, the assumption of a single error distribution can neither reflect the aggregate quality measurement of multiple source data nor reveal the impact of the differences of multiple errors on three-dimensional model construction. Current research has focused on the analysis of positional error for points. In reality, subsurface feature positions and shapes for three-dimensional geological modeling are determined by sets of related geological attributes and their distributions. Variation in point positions and their error distributions is part of the variation in the geological attributes in the entire space or volume. Therefore, some type of inherent relationship may exist between the spatial error and the geological attributes of contacts. Within error bounds defined by a threshold value, the probability of a point location is the probability that the contact showed up in measurements or samples. Additionally, the probability is the likelihood of encountering a certain type of rock or soil at a given location (Tacher et al. 2006). When spatial errors of contacts are defined with three coordinates, such as x, y, and z, and are transferred into the probability of a particular type of geological attribute, such as rock type, the different errors can be aggregated by linear weighted summation because multiple errors are represented as a single variable.

This research introduces a method for mapping the relationships among multi-source error distributions in three-dimensional geological data and attributes. The procedure for error analysis in the proposed method can be summarized by the following stages (Fig. 1):

1. The method begins with a best-guess geological interpretation of observation data, including stratigraphic information in the borehole, cross-sections, or other data. The contact points on each type of source data are assigned an error distribution and confidence interval according to the observations. In this study, we were principally concerned with the contact information for boreholes and cross-sections.
2. For each contact point, the boundary error value is given according to the confidence interval, which captures areas where the interface is uncertain.
3. An error band of a line segment in a cross-section is estimated from the error distribution at the endpoints of the segment, where error types might be different. The notion of distributed probability is introduced. We infer formulations for mapping geological attribute probabilities and the spatial error distribution of contacts for different geological formations. Then, the error distribution in three-dimensional space is transferred into a probability field for each of the contact points.
4. The probability values in the error band of line segments in geological cross-sections are calculated by the same method as in stage 3. Based on the information entropy method, the entropy weighting was calculated for each type of error in the source data. All types of errors on the same contact point are aggregated by a linear summation with entropy weighting. Then, the error distribution of each
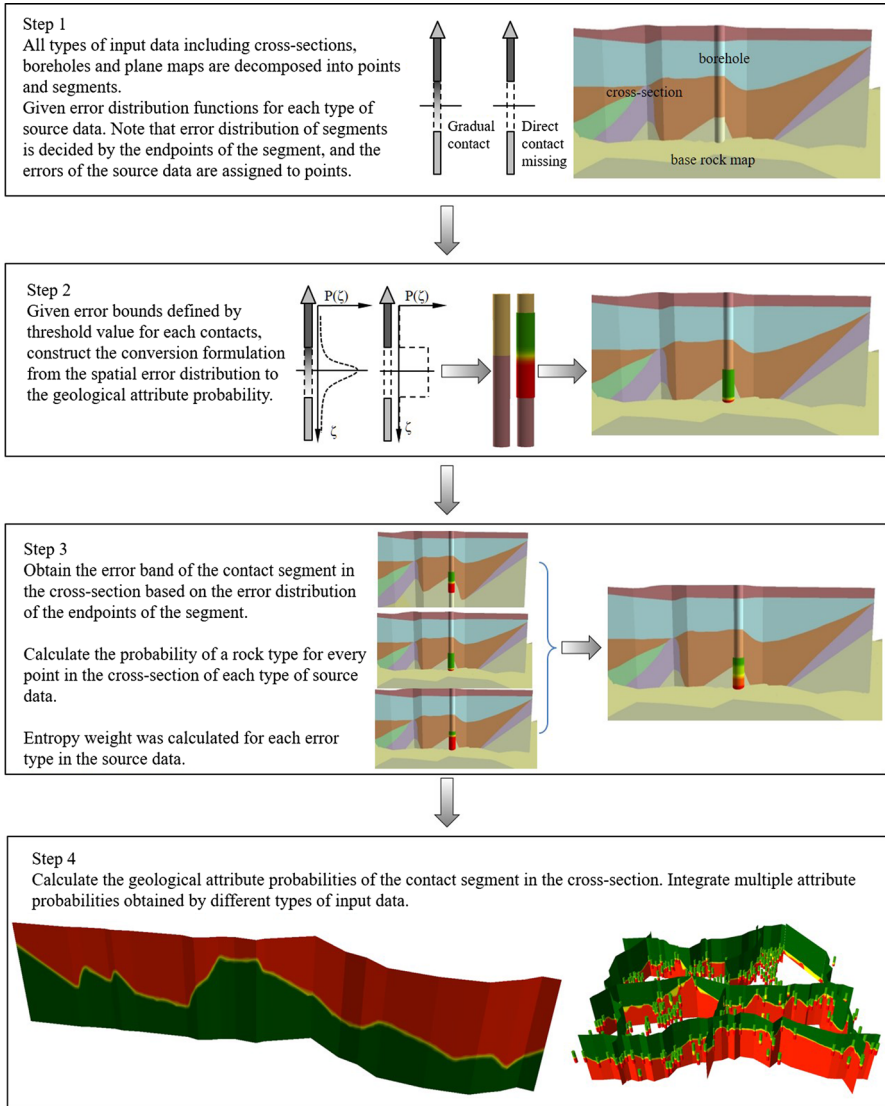
Fig. 1 Simplified workflow of the proposed method

rock type in the entire three-dimensional space can be estimated and visualized by the probability field.

It is important to note that multiple errors are directly represented in detail by probability fields on the basis of the prior model. This method does not assume the single error distribution on one contact point, as noted by Tacher et al. (2006).

3.2 Spatial Error and Geological Attribute Probability

Every point in space actually exists and falls within the boundaries of strata as chosen by geologists according to the proximal geological attributes, such as geological age, depth, and lithological characteristics. Therefore, errors along geological boundaries appear because uncertainties exist in the measurements and definitions of geological attributes. Accordingly, the problem of positional errors at spatial points can be treated as a set of uncertainties in the related geological attributes. Probability is used to express the spatial error distribution as presented by Tacher et al. (2006). In this study, errors in the subsurface or interpretation of the strata are assumed to be the chance of appearance of a particular rock type at a spatial point. The chance is a probability of the event that "rock type A is found at the location $z(u)$", which is termed the "geological attribute probability". The geological attribute probability is a set of probabilities of geological attributes that decide the classification of the strata and not just the rock type. The cumulative distribution function $F(u)$ of the probability density function $f(Z)$ is often used to calculate probabilities $P(A)$ within a confidence interval

$$P(A) = F(u) = \int_a^b f(z)\mathrm{d}z, \tag{1}$$

where a and b are bounded values of the confidence interval at a chosen level. Here, the $P(A)$ is a function of position or z-value.

3.3 Geological Attribute Probability of an Error on a Contact Line

Except for contact points in borehole data, contact lines extracted from cross-sections are often used as source data for three-dimensional geological modeling because they represent the geological understanding and consensus of geologists. Here, the cross-sections around boreholes are discussed without taking into consideration those sections interpreted from geophysical data, such as reflection or refraction seismic sections. In this situation, contact lines are decomposed into several line segments or more. The error distribution of the line segments is decided by the errors of the endpoints of the line segment.

In geographic information system (GIS) literature, the error distribution of a line segment has been investigated for half a century. Based on the epsilon band (Perkal 1966) and statistical theory, other error models have been developed, such as the error bands (Li et al. 2002; Shi and Liu 2000; Zhang and Tulip 1990), the confidence region (Dutton 1992), the H-band (Chilès et al. 2004), and the G-band (Calcagno et al. 2008) models. These methods make the same assumption that the endpoints of a line segment follow the same error distribution, such as the normal distribution (Tong et al. 2013). However, in the three-dimensional geological multi-source context, the endpoints of a line segment may follow different error distributions, such the bimodal distribution or the P-norm distribution with $p = 1.60$ (Tong and Liu 2004). In this study, the error distribution of an arbitrary point on a segment was derived based on the error bands.
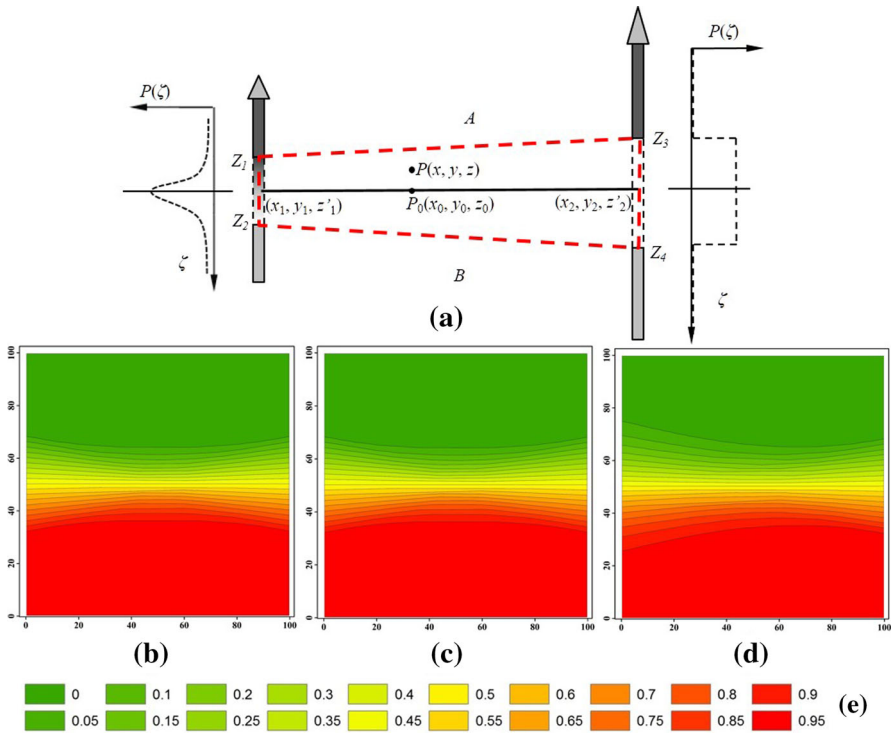
Fig. 2 Experiment of error distributions of a line segment. **a** Line segment between two contact points of boreholes. **b**–**d** Distributions of the probability that a point belongs to strata B, in which the error endpoints follow different kinds of error distribution; **b** errors of both endpoints follow a normal distribution; **c** errors of both endpoints follow a continuous distribution; **d** errors of the left endpoint follow a normal distribution and those of the right endpoint follow a continuous distribution. **e** Legend of the probability that a point belongs to strata B

Given that the depth positions of endpoints (x_1, y_1, z'_1) and (x_2, y_2, z'_2) of a line segment on a cross-section are independent random variables z'_1 and z'_2 (Fig. 2a), for which errors follow distributions $f_1(z)$ and $f_2(z)$, the depth value of an arbitrary point on the segment (except for the endpoints) can be expressed as $z = C_1 z'_1 + C_2 z'_2$, where $C_1 + C_2 = 1$ and $C_1 = \frac{\sqrt{(x-x_2)^2+(y-y_2)^2+(z-z'_2)^2}}{(x_1-x_2)^2+(y_1-y_2)^2+(z'_1-z'_2)^2}$. The error density function at an arbitrary point z_0 on the line segment is

$$f(z_0) = \frac{1}{C_2} \int_{-\infty}^{\infty} f_1(z'_1) f_2 \left(\frac{z_0 - C_1 z'_1}{C_2} \right) dz_1. \qquad (2)$$

For an arbitrary point $P(x, y, z)$, right above point $P_0(x, y, z)$ on the line segment (Fig. 2a), the necessary and sufficient condition for belonging to the upper stratum A is that the boundary of stratum A is beneath point P, which means that $z_0 \leq z$. Therefore, the probability value of $P(A, z)$ equals the cumulative distribution function of the probability density function $f(z)$ from negative infinity to z. Therefore, different results are obtained for different distribution functions, such as the normal distribution

function, the uniform distribution function, or a discrete distribution function, on the endpoints. Here, the normal distribution and continuous uniform distribution are discussed in detail. A continuous uniform distribution is taken into consideration when a contact between two formations is not outcropped or when a segment is missing a borehole sample. Usually, the strata boundaries in the cross-section or virtual borehole used in three-dimensional geological modeling are not outcropped. Three types of combinations are presented, as follows (Fig. 2b–d):

1. If the errors of the two endpoints both follow a normal distribution, the probability value of $P(A, z)$ at an arbitrary point (x, y, z) can be expressed as

$$P(A, z) = F(z) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{z} e^{-\frac{(z_0-\mu)^2}{2\sigma^2}} \, dz_0, \tag{3}$$

   where μ is the expected value and σ is the variance.

2. Given that the errors of the two endpoints both follow a continuous uniform distribution $f_1(z) = \begin{cases} \frac{1}{2a} & |z| \le a \\ 0 & |z| > a \end{cases}$ and $f_2(z) = \begin{cases} \frac{1}{2b} & |z| \le b \\ 0 & |z| > b \end{cases}$, the probability value of $P(A, z)$ on an arbitrary point $P(x, y, z)$ can be represented as three conditional instances. If $aC_1 = bC_2$, $P(A, z)$ should be

$$P(A, z) = \begin{cases} 0 & z < -aC_1 - bC_2 \\ \frac{[z+(aC_1+bC_2)]^2}{8abC_1C_2} & -aC_1 - bC_2 \le z < 0 \\ \frac{-z^2+2(aC_1+bC_2)z+(aC_1+bC_2)^2}{8abC_1C_2} & 0 \le z < aC_1 + bC_2 \\ 1 & z \ge aC_1 + bC_2 \end{cases}. \tag{4}$$

   If $aC_1 > bC_2$, $P(A, z)$ should be

$$P(A, z) = \begin{cases} 0 & z < -aC_1 - bC_2 \\ \frac{[z+(aC_1+bC_2)]^2}{8abC_1C_2} & -aC_1 - bC_2 \le z < bC_2 - aC_1 \\ \frac{1}{2} + \frac{z}{2aC_1} & bC_2 - aC_1 \le z < aC_1 - bC_2 \\ \frac{-z^2+2(aC_1+bC_2)z}{8abC_1C_2} + 1 - \frac{bC_2}{2aC_1} & aC_1 - bC_2 \le z < aC_1 + bC_2 \\ 1 & z \ge aC_1 + bC_2 \end{cases}. \tag{5}$$

If $aC_1 < bC_2$, $P(A, z)$ should be

$$P(A, z) = \begin{cases} 0 & z < -aC_1 - bC_2 \\ \frac{[z+(aC_1+bC_2)]^2}{8abC_1C_2} & -aC_1 - bC_2 \le z < aC_1 - bC_2 \\ \frac{1}{2} + \frac{z}{2bC_2} & aC_1 - bC_2 \le z < bC_2 - aC_1 \\ \frac{-z^2+2(aC_1+bC_2)z}{8abC_1C_2} + 1 - \frac{aC_1}{2bC_2} & bC_2 - aC_1 \le z < aC_1 + bC_2 \\ 1 & z \ge aC_1 + bC_2 \end{cases} \quad .$$

(6)

3. Given that the error at one endpoint follows a continuous uniform distribution $f_1(z) = \begin{cases} \frac{1}{2a} & |z| \le a \\ 0 & |z| > a \end{cases}$ and the other follows the normal distribution $f_2(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-z^2}{2\sigma^2}}$, the error distribution function of an arbitrary point $P(x, y, z)$ can be expressed as (Liu 1999)

$$f(z) = \frac{1}{C_2} \int_{-\infty}^{\infty} f_1(z_1) f_2(z_2) dz_1 = \frac{1}{C_2} \int_{-a}^{a} \frac{1}{2a\sigma\sqrt{2\pi}} e^{\frac{-\varphi(\frac{z_0-C_1z_1}{C_2})^2}{2\sigma^2}} dz_1. \quad (7)$$

Then, $P(A, z)$ should be

$$P(A, z) = \frac{1}{2} \tag{8}$$
$$+ \frac{C_2}{2aC_1} \left\{ \left(\frac{z + C_1 a}{C_2} \right) \varphi_0 \left(\frac{\frac{z+C_1a}{C_2}}{\sigma} \right) - \left(\frac{z - C_1 a}{C_2} \right) \varphi_0 \left(\frac{\frac{z-C_1a}{C_2}}{\sigma} \right) + \frac{\sigma}{\sqrt{2\pi}} \left(e^{-\frac{(\frac{z_0+C_1a}{C_2})^2}{2\sigma^2}} - e^{-\frac{(\frac{z_0-C_1a}{C_2})^2}{2\sigma^2}} \right) \right\},$$

where $\varphi_0 \left(\frac{\frac{z+C_1a}{C_2}}{\sigma} \right) = \varphi \left(\frac{\frac{z+C_1a}{C_2}}{\sigma} \right) - \frac{1}{2}$, $\varphi_0 \left(\frac{\frac{z-C_1a}{C_2}}{\sigma} \right) = \varphi \left(\frac{\frac{z-C_1a}{C_2}}{\sigma} \right) - \frac{1}{2}$, and φ is a standard normal distribution function. The geological probabilities at an arbitrary point on a line segment can be obtained using this method, and geological attribute probability distributions of an entire line feature are composed by related probabilities of a finite number of connected segments.

When the error distributions of the two endpoints of a line segment both follow normal distributions, the probabilities that a point belongs to strata B become distributed as an error band (Fig. 2b), as proven by Shi and Liu (2000). In Fig. 2c, the width of the probability distribution belonging to strata B around the segment looks like an error band, in which the maximum values are located at the endpoints and the minimum values are located in the middle of the segment. The reason this occurs is that the error distribution is a linear combination of two identical distribution functions. From Eqs. (7) and (8) in case 3, the geological attribute probability is a cumulative function of the normal distribution function in the confidence interval $[-a, a]$, although errors

of each endpoint follow a different distribution. The shape of the probability distribution that belongs to strata B looks like the result of two normal distributions with different confidence intervals.

3.4 Aggregating Multiple Errors by Geological Attribute Probabilities Using Entropy-Based Weighting

After the conversion using the method proposed above, every type of error was represented by the related geological attribute probabilities. The next step is to combine the information from different sources in a probabilistic framework. Studies have presented diverse approaches. A general formulation of all pooling methods was presented by Allard et al. (2012)

$$T_G(A) = Z + U(A) + \left(1 - \sum_{i=1}^{n} w_i\right) T_0(A) + \sum_{i=1}^{n} w_i T_i(A), \tag{9}$$

in which T is related to probabilities in the following way: $T \equiv P$ for all linear pooling methods; $T \equiv \ln P$ for methods based on the product of probabilities; and $T \equiv \ln O = \ln P - \ln(1 - P)$ for methods based on the product of odds. $U(A)$ is an updating likelihood when considering the general log-linear pooling; it is the logarithm of the Nu parameter for the Nu model. $T_0(A)$ is the prior probability and Z is a normalizing constant. The weight w_0 has been set equal to $\left(1 - \sum_{i=1}^{n} w_i\right)$ to respect external Bayesianity. The form of Eq. (9) is a kind of a generalized linear pooling method. The aggregated probabilities can be calculated by any item of the formula (Allard et al. 2012). All of the methods that can be included in Eq. (9) have some parameters, such as weights, that should be estimated or preset by users, except for entropy-based methods, such as the maximum entropy method. The estimated or preset parameters may introduce new error or uncertainty into the final results. Since entropy was introduced in geostatistics as a measure of the uncertainty of a prior distribution model (Christakos 1990), the entropy-based methods including the maximum entropy approach were widely used in earth science, especially in estimating an unknown target probability distribution.

In this study, an additive method with entropy-based weighting is presented to aggregate multiple errors. Given that N types of source data were involved in the cross-area, n possible stratigraphic attributes exist at the cross-point (x, y, z), and the probability value that the cross-point belongs to stratum i in source data j is $P_{ij}(x, y, z)$. The $P_{ij}(x, y, z)$ can be calculated using Eqs. (1) and (2). The information regarding the entropy of point (x, y, z) can be interpreted as the amount of missing information with respect to the lithological probabilities for the point. Then, an entropy-based weight value for each kind of source data can be described as

$$\omega_j = \frac{1 - H_j}{\sum\limits_{j=1}^{N} (1 - H_j)}$$

$$H_j = - \sum_{i}^{n} P_{ij}(x, y, z) \log P_{ij}(x, y, z), \tag{10}$$

where H_j is an entropy value for the source data j. The w_j value is equal to 0 when n is 1. Note that the w_j here is not dictatorial, as described by Allard et al. (2012), because $w_j \neq 0$ when $i \neq j$. For the same type of source data, the entropy-based weight value of each point in space may be different because it is determined by the corresponding geological attribute probabilities rather than by the error type. The error in each type of data source has an entropy value that is conditioned to the probabilities, according to Eq. (10). Therefore, the final value of the geological property probability of a position is calculated as

$$P_i(x, y, z) = \sum_{j}^{N} \omega_j P_{ij}(x, y, z). \tag{11}$$

When a new data source is introduced, the entropy value of the point will change because the new data source provides another probability of the geological attribute. Equation (11) is a linear pooling method with a nonlinear weight value w_j. The aggregated probability distribution is often clearly multi-modal.

3.5 Experimental Data and Method Testing

The method was tested by analyzing the error distribution at the intersections of different types of data. Several hundred borehole samples, a map of the basement rock, and six intersected cross-sections were collected, as shown in Fig. 3 (Hou et al. 2016). In Fig. 3, there are six cross-sections and nearly 200 boreholes. Boreholes DC043 and DC091 are on the line of section EW2. Sections EW2 and SN1 cross at the position of DC091. Cross-sections were digitized from a scanned image. The basement rock map was digitized according to the isobaths of the basement rocks. The stratigraphic divisions were given for each borehole according to lithological information and test data. To simplify the presentation of the data, only the Quaternary layer and basement rock are specifically identified in Fig. 3. In the intersections, ideally, the contacts of the basement rock and the Quaternary layer extracted from the borehole, the cross-section, and the basement rock map should be at the same point. In fact, the contact points between the Quaternary layer and basement rock from the borehole are not always consistent with those extracted from the basement rock map, such as the point at borehole DC037 shown in Fig. 3.

3.5.1 Attribution Aggregation for a Single Point

In this study, the contacts between the Quaternary layer and basement rock for three cross-points are extracted, as shown in Fig. 3, to test the proposed method. Among
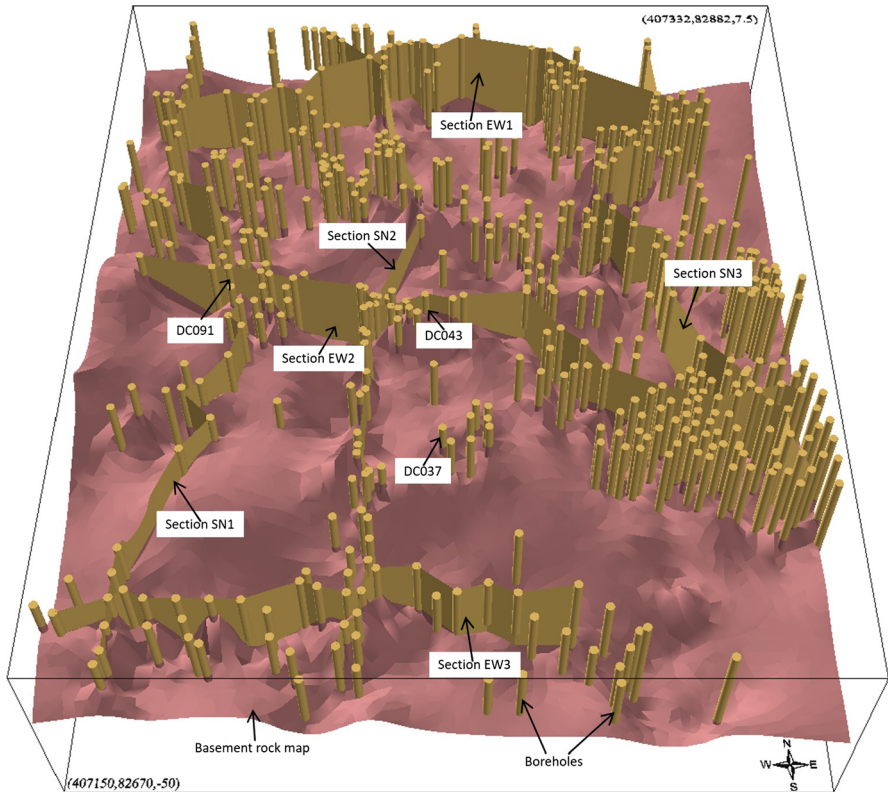
Fig. 3 Testing data. To simplify the problem, strata from borehole cores are classified as two types of strata: Quaternary stratum (marked in yellow) and basement rock. Sections EW2 and SN1 crossed at DC091. Probabilities along boreholes DC043, DC091, DC037, and section EW2 and on the geological map are specifically discussed in the testing section

these data, the contacts of borehole DC037 were not used to interpolate the basement rock map. The position of the contact point from the basement rock in borehole DC037 is inconsistent with that from the basement rock map, as shown in Fig. 4a. The contacts of the other two boreholes DC091 and DC043 are reconciled to the contacts with the basement rock map. Errors in the geological contacts from the borehole and basement rock map follow the normal distribution function with a standard deviation of 0.667 m.

In forming the attribution probability, Fig. 4b, c showed the error distributions around the contact points in the three boreholes and cross-points on the basement map. The borehole DC037 was added after the basement rock map was interpolated with the contacts from the other boreholes and cross-sections. The expected aggregated value of the contact along the borehole DC037 locates between the contact position extracted from the borehole DC037 and that extracted from the basement rock map (Fig. 4d). As shown in Fig. 5a, the curve of the aggregated attribution probability of the basement rock contact point in DC037 is approximately the average of the other two probabilities because the contact position of the basement rock that is extracted from
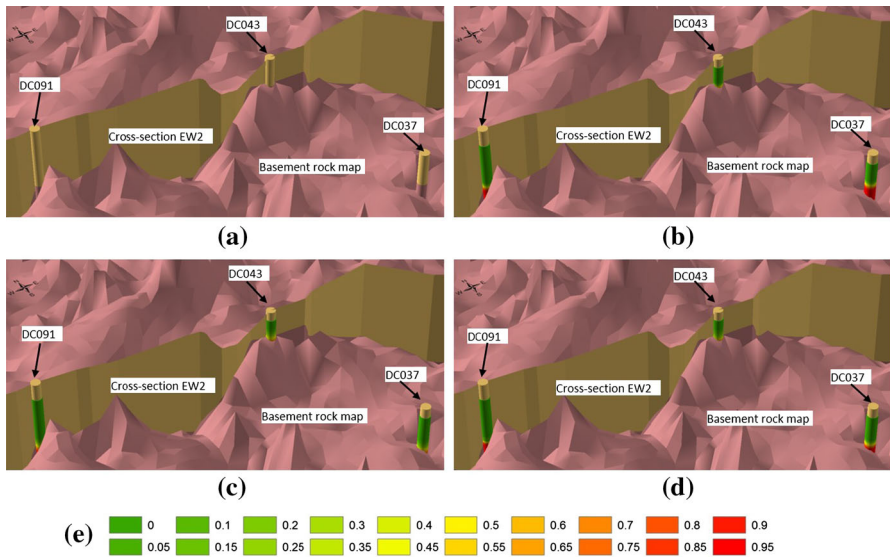
Fig. 4 Error aggregation for different source data at a cross-point: **a** positions of the cross-points in the boreholes, the cross-section, and the geological map; **b** attribution distribution of the basement rock in the borehole; **c** attribution distribution of the basement rock of the contact points extracted from the basement rock map; **d** aggregated results of **b** and **c**; and **e** legend of the probability of the basement rock distribution

borehole DC037 (approximately -7 m) is far away from the corresponding position from the basement rock map (approximately -10 m). Note that the aggregated curve has two points that are tangential to the other two curves where the probability value is 0.5, marked by dashed lines in Fig. 5. For example, the aggregated probability equals the probability value of the basement rock in DC091, while the probability of the basement rock on the point of the basement rock map is 0.5 (Fig. 5c). If the base value is 2 in Eq. (10), the entropy value is 1.0 when the probability value is equal to 0.5. Then, the entropy weight value of the corresponding source is 0. This result means that one data source cannot provide any information to judge which stratum the contact point belongs to when the probability value of the current data source is 0.5.

3.5.2 Attribution Aggregation Around the Cross-Section Using the Segment

An experiment regarding the error distributions for basement rock boundaries is presented to demonstrate the implementation of the proposed method. Basement rock boundaries were extracted from cross-section EW2 and the basement rock map along the EW2 cross-section. The error of the boundaries of the basement rock extracted from the cross-section EW2 was assumed to follow a continuous distribution because these data are not obtained by outcropping, as noted by Tong et al. (2013), and the errors of the cross-point between the cross-section and the borehole follow a normal distribution. Boundaries of the basement rock with errors that follow a normal distribution were extracted according to the cross-over lines between the cross-section and the basement rock map. Figure 6b, c show the probability distributions of the
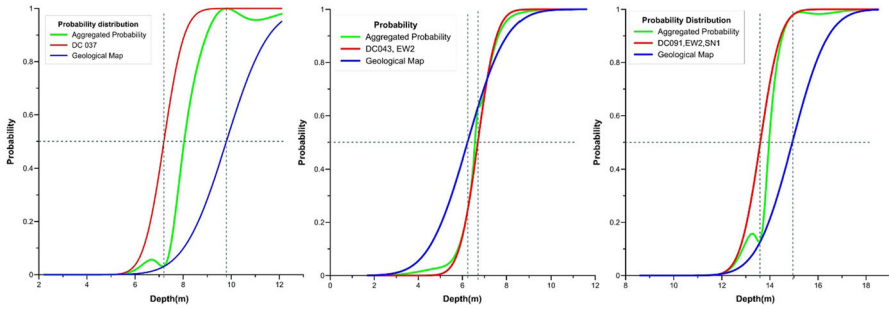
Fig. 5 Comparison of the attribution probability curves of the basement rock before and after aggregation. The left, middle, and right figures are the attribution probability curves of the basement rock for the contact points on the position of boreholes DC037, DC043, and DC091, respectively. The point at which the dashed lines cross marks the position where the aggregated probability is 0.5

basement rock boundaries from the EW2 cross-section (Fig. 6a) and the basement rock. In the cross-section, the segments that composed the strata boundaries were treated separately. Errors on endpoints of one segment can follow different distributions. Although the expected value of the basement rock depth in different source data has the same trend, huge differences exist in several local positions, such as the position marked by the black arrow in Fig. 6b, c. The attribute probability distribution of the basement rock appeared to have many glitches in Fig. 6c because the contact points were densely distributed on the basement rock map. The yellow area in Fig. 6d is much wider than that in Fig. 6b, c, as shown by the position marked with the black arrow, because the boundaries of the basement rock in the two types of data do not coincide with each other. This result illustrates that the aggregated probability value is, essentially, a weighted average of the other two data sources. The attribution probability curves were extracted up to a depth of − 21 m from Fig. 6b–d (in Fig. 7). The probabilities of the basement rock boundaries extracted from the cross-section EW2 (red curves in Fig. 7) and from the basement rock map (green curves in Fig. 7) are not coincident with each other. The entropy-weighted aggregated probability curve is located between other curves of the two sources. Additionally, as shown in the left image of Fig. 8, the probabilities of the basement rock contacts on the cross-sections and borehole cores were calculated. The probability values are aggregated on these contact points where several kinds of source data cross, such as the contact point along DC091.

The method presented here is an example of a one-dimensional approach. To extend the method to a three-dimensional situation, a very simple flow is presented, and the analytical solution for a triangle is derived. Assuming a triangle, as shown in Fig. 9, the coordinates of each endpoint are $A(x_1, y_1, z_1)$, $B(x_2, y_2, z_2)$, and $C(x_3, y_3, z_3)$. Point P is an arbitrary point in the triangle and point D is the intersection point of line AP and line CB. The coordinates of point D are $D(x_D, y_D, z_D)$. The lengths of line segments in the image can be represented as: $|CD| = a$, $|BD| = b$, $|AP| = c$, and $|PD| = d$. Let $C_1 = b/(a + b)$, $C_2 = a/(a + b)$, $C_3 = d/(c + d)$, and $C_4 = c/(c + d)$. The depth of point P can be represented as

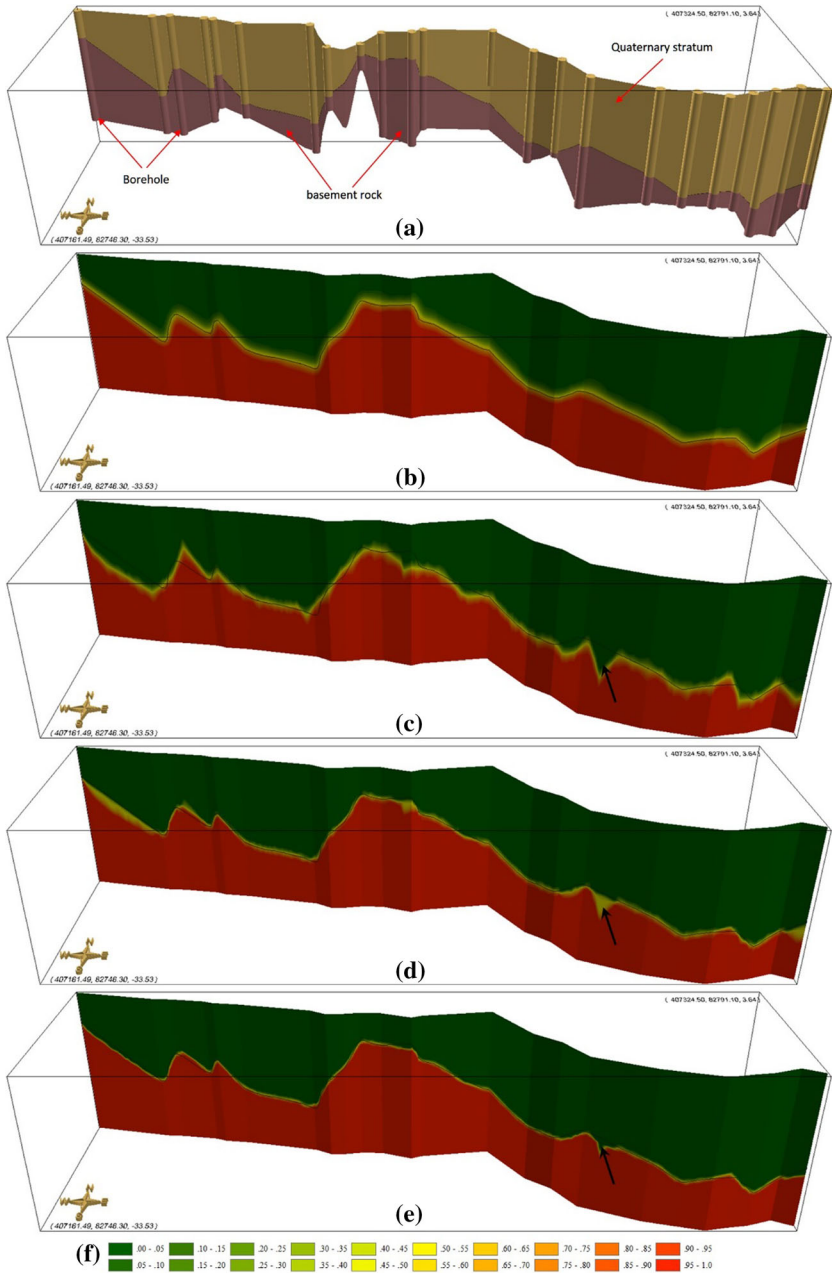$$Z_0 = C_3 z_3 + C_4 z_D \quad \text{and} \quad Z_D = C_1 z_1 + C_2 z_2. \quad (12)$$

Fig. 6 Probability distribution of the basement rock depth along the cross-section EW2, where the black line marks the depth of − 21 m: **a** the cross-section EW2 represented by the Quaternary layer and the basement rock; **b** the probability distribution of the base rock boundaries that are extracted from cross-section EW2; **c** the probability distribution of the base rock boundaries that are extracted from the basement rock map along cross-section EW2; **d** the aggregated result of **b** and **c** with entropy-based weighting; **e** the aggregated probability results of the basement rock depth along the cross-section EW2 using the maximum entropy method; and **f** the legend of the probability that the point belongs to the basement rock
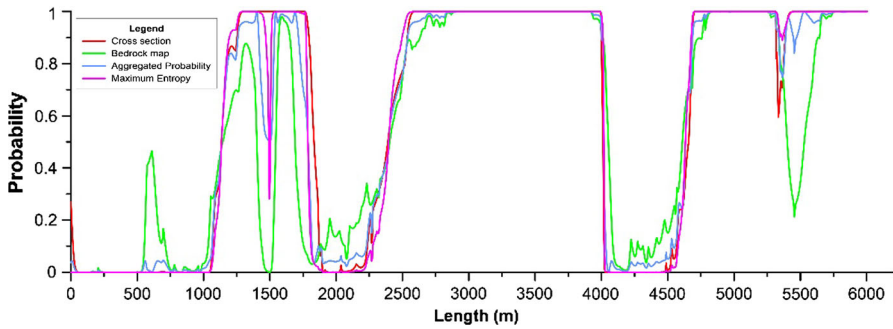
Fig. 7 Attribution probability curves of the basement rock using the weighted average and maximum entropy method at the depth of − 21 m, which are extracted from Fig. 6b–e. The purple curve represents the result of the maximum entropy method, and the navy blue curve is the aggregated result of the weighted average method
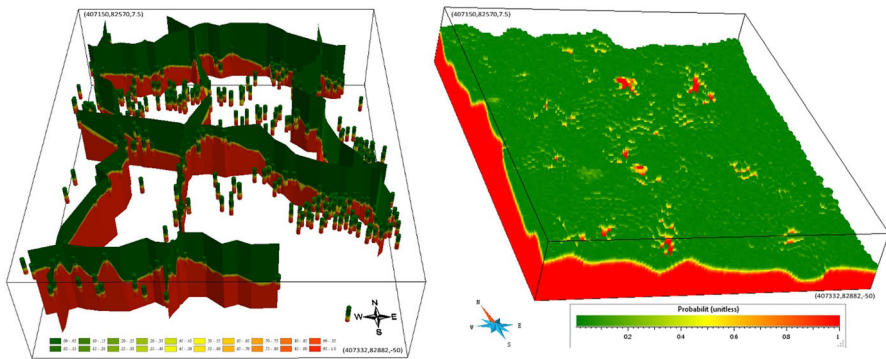

Fig. 8 Results of the probability distribution of the basement rock. In the left figure, the aggregated probability of the basement rock, which is extracted from boreholes, cross-sections, and the basement rock map, is shown, along with the boreholes and cross-sections. The right figure illustrates the probability distribution of the basement rock surface without a zero value

The error distribution of the points in the triangle can be calculated according to the method presented in this paper. The basement rock surface is represented by triangle networks. Then, the attribute probabilities of every point on the surface can be calculated. The right image of Fig. 8 illustrates the integrated probability distribution of the basement rock shown along the cross-sections boreholes, and shows the probability distribution of the basement rock surface without a zero value.

4 Discussion

In this study, the entropy value was used as a weighting value for different error distributions of the input data. This is a key factor that influenced the final probability. The weight value was calculated by the probabilities rather than introducing new parameters, as in some other methods. When the probability of a point is much larger than 0.5, the geological attribute on the point is relatively certain. In this circumstance,
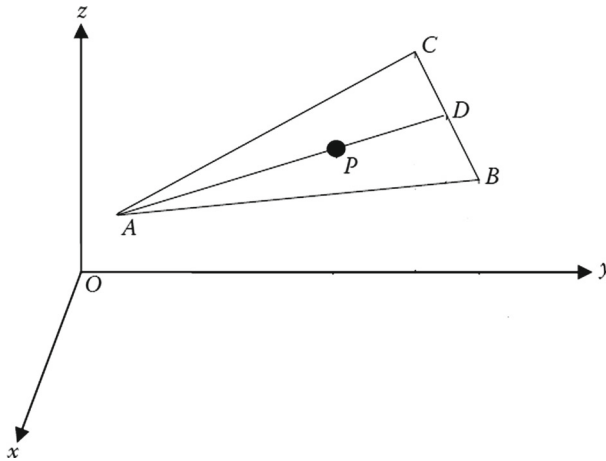
Fig. 9 Diagram of a triangle

the contribution of this type of data source is greater than that of other types of data sources in determining the geological attribute. When the probability equals 0.5 and the entropy is 1.0, the geological property on the point is the most unclear, and this data source contributes the least to classifying the geological attributes. Therefore, the weight value is determined by the error characteristics, including the error distribution function and its corresponding parameters, such as the confidence interval and standard deviation.

Essentially, the method presented in this study is a generalized linear pooling method aggregating multiple errors from different data sources. Many disciplines, including statistics, management sciences, and geostatistics, have studied aggregating information from distinct sources. The methods found in the literature can be classified into multiplicative-based and additive-based aggregation operators (Allard et al. 2012). In the methods that can be included in Eq. (9), the maximum entropy method is the only approach that does not introduce any extra parameters. Therefore, the results of the aggregating probabilities were compared (as shown in Figs. 6e and 7). The maximum entropy is a logarithmic linear pooling method and is more sensitive than an entropy-based weighting method. The range of error distribution (yellow area shown in Fig. 6e) in the maximum entropy method is narrower than that in the entropy-based weighting method. For example, at the position marked by the black arrow in Fig. 6, the aggregated probability is approximately 0.5 with the entropy-based weighting method, while the probability calculated by the maximum entropy method is 1.0. From Fig. 7, the probabilistic curve extracted from the result by the entropy-based weighting method appears turbulent from 1500 to 2000 m and from 3500 to 4000 m. However, the curve calculated by the maximum method is much smoother. When the probability is less than 0.5, the aggregated probability is greater than that of the entropy-based weighting method, and vice versa.

The method presented in this study is actually a type of linear method in which the weight values are not preset by users or estimated by introducing some other

parameters. The weight value is decided by the characteristics of the error distribution itself. Although the linear method is intrinsically sub-optimal, the presented method can reveal the error range and details of different input data.

The error type is assigned for each contact position from different sources in the presented method. In this study, the error types of contact points are not classified by the stratigraphic information around the contact points. However, this information will not affect the aggregated process and applications because the probabilities around the segments of all the boundaries were calculated separately. Therefore, the error type at each contact point will impact the probability distributions instead of the workflow. The additive method used in this study has an implicit assumption that the error distribution of an arbitrary point of one segment should be limited by the error distribution of the endpoints, which is why the uncertainties on the segment seem smaller than the endpoints of the segment in Figs. 2 and 6.

From Fig. 7, the integrated probability curve (pink curve in Fig. 7) based on the maximum entropy method has the same waveform as the curve obtained using the entropy-based method, but the result with the maximum entropy method is much smoother. Notice that most values of the integrated result with the maximum entropy method are smaller than those values obtained using the entropy-based method (navy blue curve in Fig. 7), except for some positions, such as the position at 1500 m.

Limitations of our method are mainly related to the error types and the conversion from error distribution to geological attribute probabilities. Classification of the error types for contact points is a time-consuming and tedious computation that reduces the efficiency. For classification, it could be reasonable to build an error-type appointment rule based on the stratigraphic information. In this paper, the error distribution has been discussed in the depth direction instead of in two or more dimensions. This avoids the overlay problem on the conjunct point between segments. At this point, conversion formulas for high-dimension error distribution functions should be inferred based on the positional error of the line features or curve lines. This would help us analyze the topological error in the source data. Additionally, the implementation is limited to possible outcomes. This study presents a one-dimensional method and does not analyze a case with multiple z-values, such as an overturned fold, dome structures, and thrust faults.

5 Conclusions

In this study, a one-dimensional multiple errors analysis method is presented for geological contacts that are typically used to build geological models. The foundation of the approach is that the spatial errors of contacts or boundaries between strata can be represented by geological attribute probabilities, of which it is assumed that errors are caused by imprecise measurement or vague classification of geological attributes, such as rock type. The error distribution functions and prior guesses are accounted for, since we performed the analysis based on statistical manipulations. At each location of input data, the probabilities of each kind of rock type can be evaluated using this method. The regulation of different sources and types of errors using the entropy-based weighting is presented. Therefore, in this method, the final chance of finding a certain

rock type at one particular point relies on the probabilities of different types of error distributions.

The approach is simple and intuitive for analyzing and visualizing errors in multi-source data. At a contact point, the error distribution of a rock type is calculated using the error of depth of the contact point. The final result is multiple models after integration if more than one input datum exists for the contact point. Similar to the error band used in the geographic information system (GIS), the integrated one-dimensional error density distribution functions along a segment were derived, where errors at the endpoints followed different distribution functions, such as a normal distribution or a continuous distribution. By assuming that the error at a random point of a segment is a combination of the error distributions of the endpoints, the shape of the probability distributions appears to be similar to the error bands presented by Tong and Liu (2004) and Tong et al. (2013). Because the presented method is based on error assessment of a probabilistic model (Tacher et al. 2006), the assumption that only one kind of error distribution can be defined for all source data is not limited to this application. In the two experiments, we have shown how we can aggregate multiple types of errors at cross-points. An advantage of our method is that it is not limited to a specific error distribution and that the workflow itself is always performed in the same way. The test results illustrated that the integrated probability is a weighted average result, although the weights are nonlinear. Like the maximum entropy method, the presented method does not estimate parameters for integration because the weight value was decided by the probabilities of the rock type based on the input data.

Assuming that the uncertainty away from an observation should be higher than the uncertainty at an observation, uncertainties are characterized and measured by information entropy values in the method proposed by Wellmann and Regenauer-Lieb (2012). In this study, the error distribution at arbitrary points of a segment is assumed to follow or be limited by the error distributions of the endpoints of the segments, yet it is still measured by the error distribution function. Therefore, the uncertainty at a random point of a segment is smaller than that at the endpoints. Additionally, totally different results were obtained because of different assumptions.

Several aspects of this research are to be developed further, including an analysis method for higher dimensional error distributions. Additional deterministic vector information, such as dip measurements, should be taken into consideration in the process. Regardless, the proposed method expands and refines the treatment of error and uncertainty in geology and opens new possibilities for more sophisticated uses of three-dimensional geological models in instances in which errors in the model can translate into serious consequences of risk, failure, or cost.

References

Allard D, Comunian A, Renard P (2012) Probability aggregation methods in geoscience. Math Geosci 44:545–581

Bárdossy G, Fodor J (2004) Evaluation of uncertainties and risks in geology—new mathematical approaches for their handling. Springer, New York

Bistacchi A, Massironi M, Dal Piaz GV, Dal Piaz G, Monopoli B, Schiavo A, Toffolon G (2008) 3D fold and fault reconstruction with an uncertainty model: an example from an Alpine tunnel case study. Comput Geosci 34:351–372

Bond CE, Lunn RJ, Shipton ZK, Lunn AD (2012) What makes an expert effective at interpreting seismic images? Geology 40:75–78

Caers J, Hoffman T (2006) The probability perturbation method: a new look at Bayesian inverse modeling. Math Geol 38:81–100

Calcagno P, Chilès J-P, Courrioux G, Guillen A (2008) Geological modelling from field data and geological knowledge: Part I. Modelling method coupling 3D potential-field interpolation and geological rules. Phys Earth Planet Inter 171:147–157

Caspary W, Scheuring R (1993) Positional accuracy in spatial databases. Comput Environ Urban Syst 17:103–110

Caumon G, Collon-Drouaillet P, Le Carlier de Veslud C, Viseur S, Sausse J (2009) Surface-based 3D modeling of geological structures. Math Geosci 41:927–945

Chauvin B, Caumon G (2015) Building folded horizon surfaces from 3D points: a new method based on geomechanical restoration. In: Schaeben H, Delgado RT, Gerald van den BK, Regina van den B (eds) Proceedings of the 17th annual conference of the International Association for Mathematical Geosciences (IAMG) 2015, Freiberg, Germany, September 2015, pp 39–48

Chilès J-P, Aug C, Guillen A, Lees T (2004) Modelling the geometry of geological units and its uncertainty in 3D from structural data: the potential-field method. In: Proceedings of the international symposium "Orebody modelling and strategic mine planning—uncertainty and risk management models", Perth, Western Australia, November 2004, pp 313–320

Christakos G (1990) A Bayesian/maximum-entropy view to the spatial estimation problem. Math Geol 22:763–777

Collon P, Steckiewicz-Laurent W, Pellerin J, Laurent G, Caumon G, Reichart G, Vaute L (2015) 3D geo-modelling combining implicit surfaces and Voronoi-based remeshing: a case study in the Lorraine Coal Basin (France). Comput Geosci 77:29–43

Davis CJ (2002) Statistics and data analysis in geology, 3rd edn. Wiley, New York

Dutton G (1992) Handling positional uncertainty in spatial databases. In: Proceedings of the 5th international symposium on spatial data handling, Charleston, SC, USA, August 1992, pp 460–469

Fasani GB, Bozzano F, Cardarelli E, Cercato M (2013) Underground cavity investigation within the city of Rome (Italy): a multi-disciplinary approach combining geological and geophysical data. Eng Geol 152:109–121

Fomel S (1994) Method of velocity continuation in the problem of seismic time migration. Russ Geol Geophys 35:100–111

Fomel S, Landa E (2014) Structural uncertainty of time-migrated seismic images. J Appl Geophys 101:27–30

Grubb H, Tura A, Hanitzsch C (2001) Estimating and interpreting velocity uncertainty in migrated images and AVO attributes. Geophysics 66:1208–1216

Guillen A, Calcagno P, Courrioux G, Joly A, Ledru P (2008) Geological modelling from field data and geological knowledge: Part II. Modelling validation using gravity and magnetic data inversion. Phys Earth Planet Inter 171:158–169

Hou WS, Yang L, Deng DC, Ye J, Keith C, Yang ZJ, Zhuang WM, Liu JX, Huang JC (2016) Assessing quality of urban underground spaces by coupling 3D geological models: the case study of Foshan City, South China. Comput Geosci 89:1–11

Jessell WM, Ailleres L, de Kemp AE (2010) Towards an integrated inversion of geoscientific data: what price of geology? Tectonophysics 490:294–306

Jones RR, McCaffrey KJ, Wilson RW, Holdsworth RE (2004) Digital field data acquisition: towards increased quantification of uncertainty during geological mapping. Geol Soc Lond Spec Publ 239:43–56

Journel AG (2002) Combining knowledge from diverse sources: an alternative to traditional data independence hypotheses. Math Geol 34:573–596

Kaufmann O, Martin T (2008) 3D geological modelling from boreholes, cross-sections and geological maps, application over former natural gas storages in coal mines. Comput Geosci 34:278–290

Li DJ, Gong JY, Xie GS, Du DS (2002) Error entropy band for linear segments in GIS. Geomat Inf Sci Wuhan Univ 31:82–85 (in Chinese)

Li X, Li P, Zhu H (2013) Coal seam surface modeling and updating with multi-source data integration using Bayesian Geostatistics. Eng Geol 164:208–221

Lindsay MD, Aillères L, Jessell MW, de Kemp EA, Betts PG (2012) Locating and quantifying geological uncertainty in three-dimensional models: analysis of the Gippsland Basin, southeastern Australia. Tectonophysics 546–547:10–27

Lindsay MD, Jessell MW, Ailleres L, Perrouty S, de Kemp E, Betts PG (2013) Geodiversity: exploration of 3D geological model space. Tectonophysics 594:27–37

Liu Z (1999) Integration of uncertainties and distribution. Phys Exp 19(5):17–19 (in Chinese)

Mann CJ (1993) Uncertainty in geology. In: Davis JC, Herzfeld UC (eds) Computers in geology—25 years of progress. Oxford University Press, New York, pp 241–254

Perkal J (1966) On the length of empirical curves. Discussion Paper 10, Michigan Inter-University Community of Mathematical Geographers, Ann Arbor, MI

Polyakova EI, Journel AG (2007) The Nu expression for probabilistic data integration. Math Geol 39:715–733

Pon S, Lines LR (2005) Sensitivity analysis of seismic depth migrations. Geophysics 70:S39–S42

Royse KR (2010) Combining numerical and cognitive 3D modelling approaches in order to determine the structure of the Chalk in the London Basin. Comput Geosci 36:500–511

Shi WZ (1994) Modelling positional and thematic uncertainties in integration of remote sensing and geographic information systems. ITC, Enschede

Shi WZ, Liu WB (2000) A stochastic process-based model for the positional error of line segments in GIS. Int J Geogr Inf Sci 14:51–66

Tacher L, Pomian-Srzednicki I, Parriaux A (2006) Geological uncertainties associated with 3-D subsurface models. Comput Geosci 32:212–221

Tarantola A, Valette B (1982) Inverse problems = quest for information. J Geophys 50:159–170

Thiele ST, Jessell MW, Lindsay M, Wellmann JF, Pakyuz-Charrier E (2016) The topology of geology 2: topological uncertainty. J Struct Geol 91:74–87

Thore P, Shtuka A, Lecour M, Ait-Ettajer T, Cognot R (2002) Structural uncertainties: determination, management, and applications. Geophysics 67:840–852

Tong XH, Liu DJ (2004) Probability density function and estimation for error of digitized map coordinates in GIS. J Central South Univ Technol 11:69–74

Tong XH, Sun T, Fan JY, Goodchild MF, Shi WZ (2013) A statistical simulation model for positional error of line features in Geographic Information Systems (GIS). Int J Appl Earth Obs Geoinf 21:136–148

Travelletti J, Malet J-P (2012) Characterization of the 3D geometry of flow-like landslides: a methodology based on the integration of heterogeneous multi-source data. Eng Geol 128:30–48

Viard T, Caumon G, Lévy B (2011) Adjacent versus coincident representations of geospatial uncertainty: which promote better decisions? Comput Geosci 37:511–520

Wellmann JF, Regenauer-Lieb K (2012) Uncertainties have a meaning: information entropy as a quality measure for 3-D geological models. Tectonophysics 526–529:207–216

Wellmann JF, Horowitz FG, Schill E, Regenauer-Lieb K (2010) Towards incorporating uncertainty of structural data in 3D geological inversion. Tectonophysics 490:141–151

Wu Q, Xu H, Zou X, Lei H (2015) A 3D modeling approach to complex faults with multi-source data. Comput Geosci 77:126–137

Zhang GY, Tulip J (1990) An algorithm for the avoidance of sliver polygons and clusters of points in spatial overlay. In: Proceedings of the 4th international symposium on spatial data handling, Zurich, Switzerland, July 1990, pp 141–150

Zhang JX, Zhang JP, Yao N (2009) Geostatistics for spatial uncertainty characterization. Geo Spat Inf Sci 12:7–12

Zhu LF, Zhuang Z (2010) Framework system and research flow of uncertainty in 3D geological structure models. Min Sci Technol (China) 20:306–311