测绘出版社 SURVEYING AND MAPPING PRESS

Positional Error Model of Line Segments with Modeling and Measuring Errors Using Brownian Bridge

Xiaohua TONG[1,2], Lejingyi ZHOU[1,2], Yanmin JIN[1,2]

1. *College of Surveying and Geo-informatics, Tongji University, Shanghai* 200092, *China*; 2. *Shanghai Key Laboratory for Planetary Mapping and Remote Sensing for Deep Space Exploration, Shanghai* 200092, *China*

Abstract: Spatial linear features are often represented as a series of line segments joined by measured endpoints in surveying and geographic information science. There are not only the measuring errors of the endpoints but also the modeling errors between the line segments and the actual geographical features. This paper presents a Brownian bridge error model for line segments combining both the modeling and measuring errors. First, the Brownian bridge is used to establish the position distribution of the actual geographic feature represented by the line segment. Second, an error propagation model with the constraints of the measuring error distribution of the endpoints is proposed. Third, a comprehensive error band of the line segment is constructed, wherein both the modeling and measuring errors are contained. The proposed error model can be used to evaluate line segments' overall accuracy and trustability influenced by modeling and measuring errors, and provides a comprehensive quality indicator for the geospatial data.

Key words: spatial data; line segment; modeling error; measuring error; Brownian bridge

Citation: Xiaohua TONG, Lejingyi ZHOU, Yanmin JIN. Positional Error Model of Line Segments with Modeling and Measuring Errors Using Brownian Bridge[J]. Journal of Geodesy and Geoinformation Science, 2023, 6(2): 1-10. DOI: 10.11947/j.JGGS.2023.0201.

1 Introduction

Error modeling of spatial data has always been a core scientific issue in the field of surveying and mapping and Geographic Information Science (GIS)[1-5]. The positional error modeling of points has been well studied and solved, while the error modeling of the linear features is still a challenge and needs to be further developed.

Since the 1950s, researchers have made great contributions to modeling the errors of spatial linear features. The ε-band model[6-7] is a representative deterministic model, in which the width of the error band is a fixed value throughout the line segment. Then the error distribution of the line segments' endpoints is commonly used to derive statistical error models[8-13]. Literature [8] assumed that the endpoints' coordinate errors of line segments are independent and subject to binary normal distribution, and the errors of the interpolated points along the line segment were derived by the error propagation from the endpoints. In this error band model, the error of the middle point on the line segment is smaller than the errors of the two endpoints, which indicates that the shape of the error band is a "dumbbell" like concave set. Furthermore, Literature [14] presented the generic G-band model based on the stochastic process. In this method, the line segment is regarded as a composite of infinite linearly interpolated points. The distributions of the line segments and the analytical expression of the error band boundary were derived statistically. Literature [15] proposed a statistical simulation error band model which regards the whole line segment as a random variable. In this model, the error distribution of the line segment was obtained based on the statistical characteristics of the

Received date: 2023-03-01; accepted date: 2023-05-30
Foundation support: National Natural Science Foundation of China (Nos. 42071372; 42221002)
First author: Xiaohua TONG, male, PhD, professor, majors in planetary mapping and remote sensing for deep space exploration.
E-mail: xhtong@tongji.edu.cn

endpoints of the line segment and the pre-determined confidence level. Literatures [16—17] proposed a standard error band model based on the global error probability distribution of line segments. In this method, an accurate analytical expression of this error band based on the probability density distribution of the line segment was derived.

The shapes of the above error bands change gradually from equal width band (ε-band) to the "dumbbell" like shape, which is narrower in the middle part of the line segment than at the endpoints. Namely, the position errors in the middle part of the line segment are smaller than that at the two endpoints. In the above error models, only the measuring errors of the endpoints were considered, and the errors of the points between the endpoints can be derived based on the endpoints' measuring error. The continuous actual geographic features are usually captured as discrete points connected by straight line segments in the feature representation[18]. However, besides the measuring errors of the endpoints, there also would be positional differences between this approximated representation formed by line segments and the actual continuous features[19-21]. Here, we denote these positional differences caused by the line segments representation as the modeling errors. Ignoring the modeling errors may cause inaccurate error modeling for the line segments[22-23].

Some research attention has been focused on the modeling error for linear features. Literature [19] proposed a method to depict the modeling errors using empirically determined functions. One suggested modeling error funtion is the exponential function, which is dependent on the distance from the center of gravity of the line. Literature [22] defined the magnitude of the modeling error as the maximum distance between the quadratic curve and the line segment. Literature [21] proposed a method to evaluate the modeling error by randomly inserting sub-vertices with a uniform distribution along the line segment between two adjacent observed points. Some researchers used the positional differences between spline curves and line segments to describe modeling errors[24-25]. The positional differences between line segments and their represented actual features are unknown and random. Therefore, the modeling errors are random and may have different forms with respect to different actual geographic features. It is necessary to establish a general error modeling method that can quantify the random and various modeling errors. Brownian motion, which is a continuous stochastic process, is popularly employed to describe the motions of moving objects[26-28]. Brownian motion is constituted by random positional displacements. Different shapes of curves can be obtained by the combining of random positional displacements. Therefore, various linear features that have different shapes can be modeled by using the Brownian motion. This paper focuses on the error modeling method based on the Brownian motion.

This paper aims to present an error model of line segments by using the Brownian motion, in which both the modeling and measuring errors are taken into account. With the measuring errors of the endpoints as the prior constraints, the Brownian bridge error model is obtained to depict the overall error distribution of line segments.

The rest of the paper is structured as follows. Following the introduction, the Brownian bridge error model of line segments with modeling and measuring errors is presented in Section 2. In Section 3, numerical experiments and the results are presented, and the proposed error model is used to compute the error distribution of line segments with endpoints' prior positional errors. Finally, the conclusions are presented in Section 4.

2 Brownian Bridge Error Model of Line Segments with Modeling and Measuring Errors

2.1 Position distribution of arbitrary points on actual geographical features

Line segments are the essential components of the linear feature representation. A straight line segment represents the actual geographical feature only through the two observed endpoints. However, the actual geographical feature represented by the line segment, such as a segment of a road or coastline,

is generally irregular and non-linear. This leads to the modeling error, which is the positional difference between the actual geographical feature and the line segment. In addition, the modeling errors are different with respect to different features. Therefore, the Brownian motion is employed to describe the positional distribution of an arbitrary point on the geographical feature, which is represented by a line segment.

It is assumed that a segment of a geographic linear feature in the real world is represented as a line segment $\boldsymbol{Z}_0\boldsymbol{Z}_1$. The endpoint coordinate vector of line segment $\boldsymbol{Z}_0\boldsymbol{Z}_1$ is $\boldsymbol{S}_2=(\boldsymbol{Z}_0^T,\boldsymbol{Z}_1^T)^{\mathrm{T}}=(X_0,Y_0,X_1,Y_1)^{\mathrm{T}}$, which follows the four-dimensional normal distribution, namely

$$\boldsymbol{S}_2\sim\boldsymbol{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}) \tag{1}$$

where, $\boldsymbol{\mu}=(\boldsymbol{\mu}_0^T,\boldsymbol{\mu}_1^T)^{\mathrm{T}}=((\mu_{0,x},\mu_{0,y}),(\mu_{1,x},\mu_{1,y}))^{\mathrm{T}}$ is the expectation vector, $\boldsymbol{\mu}_i$ is the expectation of the endpoint $\boldsymbol{Z}_i$; $\boldsymbol{\Sigma}=\delta\boldsymbol{I}_{4\times4}$ is the covariance matrix of the endpoint coordinate vector $\boldsymbol{S}_2$ of $\boldsymbol{Z}_0\boldsymbol{Z}_1$, and $\boldsymbol{I}_{4\times4}$ is the 4×4 identity matrix; and the covariance matrix at the endpoint $\boldsymbol{Z}_i$ is denoted as $\boldsymbol{\Sigma}_{i,i}(i=0,1)$.

Fig.1 presents a possible case of the position of an actual linear geographic feature and its representation of line segment $\boldsymbol{Z}_0\boldsymbol{Z}_1$. In this figure, the black line denotes the positional expectation $\boldsymbol{\mu}_0\boldsymbol{\mu}_1$ of line segment $\boldsymbol{Z}_0\boldsymbol{Z}_1$, the green ellipses are the standard error ellipses of the endpoints' positions, and the blue irregular line represents a possible position of the actual feature represented by line segment $\boldsymbol{Z}_0\boldsymbol{Z}_1$.

The positions of the points on the actual geographic features represented by the line segment can be expressed as a stochastic process as follows

$$\{\boldsymbol{Z}(t),t\in[0,d]\} \tag{2}$$

where, t is the distance between the left endpoint's positional expectation $\boldsymbol{\mu}_0$ and the positional expectation $\boldsymbol{\mu}(t)$ of $\boldsymbol{Z}(t)$, $t=|\boldsymbol{\mu}_0\boldsymbol{\mu}(t)|$, and d represents the distance between the positional expectation of the two endpoints ($\boldsymbol{Z}_0$ and $\boldsymbol{Z}_1$), $d=|\boldsymbol{\mu}_0\boldsymbol{\mu}_1|$. The position of the actual geographical feature (as shown by the irregular blue line in Fig.1) is one of the sample curves $z(t)$ ($z\in\boldsymbol{Z},t\in[0,d]$) of the stochastic process $\{\boldsymbol{Z}(t),t\in[0,d]\}$.

The difference between the position $\boldsymbol{Z}(t)$ of a point on the actual feature and its positional expectation $\boldsymbol{\mu}(t)$ is denoted as the error vector $\boldsymbol{\xi}(t)$ (as shown by the black arrow in Fig.1), and can be expressed as follows

$$\boldsymbol{\xi}(t)=\boldsymbol{Z}(t)-\boldsymbol{\mu}(t) \tag{3}$$

The position of the actual geographic feature represented by line segment $\boldsymbol{Z}_0\boldsymbol{Z}_1$ is unknown and random. By assuming that the position of the actual feature is subject to a normal distribution and its positional change within a small range is relatively smooth and uniform, the Brownian motion stochastic process can be then used to describe the position $\boldsymbol{Z}(t)$ and the corresponding error vector $\boldsymbol{\xi}(t)$ of the actual linear feature.

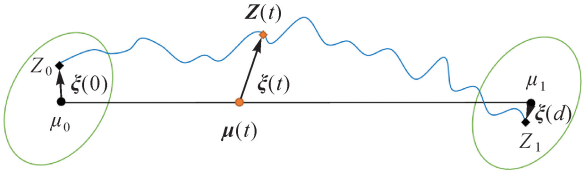

Fig.1 Line segment Z_0Z_1 and the corresponding actual geographic feature

The error vector $\boldsymbol{\xi}(t)$ is affected by the positional errors of both the line segment's left endpoint $\boldsymbol{Z}_0$ and right endpoint $\boldsymbol{Z}_1$. At any point $\boldsymbol{\mu}(t)$, the error vector obtained by the error propagation of the measuring error of the endpoint $\boldsymbol{Z}_i$ combined with the modeling error is denoted as "the error vector associated with the endpoint $\boldsymbol{Z}_i$", namely $\boldsymbol{\xi}_i(t)$. The overall error vector $\boldsymbol{\xi}(t)$ consists of two error vectors of $\boldsymbol{\xi}_0(t)$ and $\boldsymbol{\xi}_1(t)$, which are associated with endpoints $\boldsymbol{Z}_0$ and $\boldsymbol{Z}_1$, respectively. Therefore, the overall error vector $\boldsymbol{\xi}(t)$ can be expressed as follows

$$\boldsymbol{\xi}(t)=\boldsymbol{\xi}_0(t)+\boldsymbol{\xi}_1(t) \tag{4}$$

where, $\{\boldsymbol{\xi}_0(t),t\in[0,d]\}$ and $\{\boldsymbol{\xi}_1(t),t\in[0,d]\}$ are also Brownian motion stochastic processes under stationary and normal assumptions, and their variance in the unit distance is $\boldsymbol{\Sigma}_{B_i}$. Namely, $\boldsymbol{\xi}_i(t)=\boldsymbol{B}_i(t)$, and $\boldsymbol{B}_i(t)\sim N(0,t\boldsymbol{\Sigma}_{B_i})$. The position of any point of the actual linear feature can be further expressed as

$$\boldsymbol{Z}(t)=\boldsymbol{\xi}_0(t)+\boldsymbol{\xi}_1(t)+\left(1-\frac{t}{d}\right)\boldsymbol{\mu}_0+\frac{t}{d}\boldsymbol{\mu}_1 \tag{5}$$

The error vector $\boldsymbol{\xi}(t)$ is affected by the measuring errors of both the left and right endpoints $\boldsymbol{Z}_0$ and

$\boldsymbol{Z}_1$, and the positional expectation of $\boldsymbol{Z}(t)$ is $\left(1-\frac{t}{d}\right)\boldsymbol{\mu}_0+\frac{t}{d}\boldsymbol{\mu}_1$. The error propagation of the measuring errors of left and right endpoints $\boldsymbol{Z}_0$ and $\boldsymbol{Z}_1$, combined with the modeling error, will be further studied in the next section.

2.2 Integrated modeling and measuring error propagation method with the constraints of endpoints' measuring errors

The overall error vector $\boldsymbol{\xi}(t)$ is influenced by both the measuring errors of the endpoints and the modeling error of the line segment. However, there are only the measuring errors at the two endpoints. Therefore, the constraints that the error vector $\boldsymbol{\xi}(t)$ at the endpoints should be equal to the endpoints' measuring errors need to be introduced.

The error vector $\boldsymbol{\xi}_1(t)$ associated with the right endpoint is discussed firstly. Based on the constraints of endpoints' measuring errors and the conditional distributions of multivariate normal distributions[29], the error vector $\boldsymbol{\xi}_1(t)$ with constraints can be calculated as follows

$$\boldsymbol{\xi}_1(t)=\boldsymbol{B}_1(t)\mid_{\text{constraints}}=\boldsymbol{B}_1(t)-\frac{t}{d}\boldsymbol{B}_1(d)+\frac{t}{d}(\boldsymbol{Z}_1-\boldsymbol{\mu}_1),t\in[0,d] \tag{6}$$

where, $\boldsymbol{B}_1(t)-\frac{t}{d}\boldsymbol{B}_1(d)$ is the modeling error at the point $\boldsymbol{\mu}(t)$ ($t\in[0,d]$) on the expected line segment that is propagated from the error at the endpoint $\boldsymbol{Z}_1$, and is denoted as "the modeling error associated with $\boldsymbol{Z}_1$"; and $\frac{t}{d}(\boldsymbol{Z}_1-\boldsymbol{\mu}_1)$ is the measuring error at $\boldsymbol{\mu}(t)$ ($t\in[0,d]$) that is propagated from the error at the endpoint $\boldsymbol{Z}_1$, denoted as "the measuring error associated with $\boldsymbol{Z}_1$".

It is assumed previously that the distribution of $\boldsymbol{\xi}_1(t)$ follows Brownian motion without considering the constraints of endpoints' measuring errors. After the introduction of the constraints of the endpoints' measuring errors, the distribution of $\boldsymbol{\xi}_1(t)$ still partially follows the Brownian motion but is constrained at the right endpoint, which is transformed into a Brownian bridge. Therefore, the error vector $\boldsymbol{\xi}_1(t)$ at the right endpoint is equal to the measuring error of the right endpoint, namely

$$\boldsymbol{Z}_1-\boldsymbol{\mu}_1=\boldsymbol{B}^*(d)=\boldsymbol{B}_1(d)-\frac{d}{d+a}\boldsymbol{B}_1(d+a),a>0 \tag{7}$$

where, $\boldsymbol{B}^*(t)$ is a Brownian bridge stochastic process which is followed by the error vector $\boldsymbol{\xi}_1(t)$ at the right endpoint, $\boldsymbol{B}^*(t)=\boldsymbol{B}_1(t)-\frac{t}{d+a}\boldsymbol{B}_1(d+a)$; and a is the parameter that affects the magnitude of the modeling error.

According to Eq. (7), the covariance matrix of the Brownian motion $\{\boldsymbol{B}_1(t),t\in[0,d]\}$ and the measuring error related to $\boldsymbol{Z}_1$ $\left(\text{i.e., } \frac{t}{d}(\boldsymbol{Z}_1-\boldsymbol{\mu}_1)\right)$ is obtained as follows

$$\operatorname{cov}\left(\boldsymbol{B}_1(t),\frac{t}{d}(\boldsymbol{Z}_1-\boldsymbol{\mu}_1)\right)=\frac{at^2}{d(d+a)}\boldsymbol{\Sigma}_{B_1} \tag{8}$$

where, $\boldsymbol{\Sigma}_{B_1}$ represents the variance of the error associated with $\boldsymbol{Z}_1$ per unit distance. By calculating the variance of the measuring error of the right endpoint in Eq. (7), the relationship between $\boldsymbol{\Sigma}_{B_1}$ and the variance $\boldsymbol{\Sigma}_{1,1}$ of measuring error at $\boldsymbol{Z}_1$ is obtained as follows

$$d\left(1-\frac{d}{d+a}\right)\boldsymbol{\Sigma}_{B_1}=\boldsymbol{\Sigma}_{1,1} \tag{9}$$

By substituting Eq. (7) into Eq. (6), the error vector $\boldsymbol{\xi}_1(t)$ with the constraint of measuring errors considering the correlation between the modeling error and measuring error, is expressed as follows

$$\boldsymbol{\xi}_1(t)=\boldsymbol{B}_1(t)-\frac{t}{d+a}\boldsymbol{B}_1(d+a) \tag{10}$$

where, $\boldsymbol{B}_1(t)$ is a Brownian motion stochastic process that the error vector $\boldsymbol{\xi}_1(t)$ follows before considering endpoints' constraints.

Similarly, the error vector $\boldsymbol{\xi}_0(t)$ related to the left endpoint can be further obtained when the endpoint's measuring error is taken as a constraint. It is formulated as follows

$$\boldsymbol{\xi}_0(t)=\boldsymbol{B}_0(d-t)-\frac{d-t}{d+a}\boldsymbol{B}_0(d+a) \tag{11}$$

The relationship between the variance $\boldsymbol{\Sigma}_{B_0}$ of the error associated with $\boldsymbol{Z}_0$ in the unit distance and the variance $\boldsymbol{\Sigma}_{0,0}$ of measuring error at $\boldsymbol{Z}_0$ is obtained as follows

$$d\left(1-\frac{d}{d+a}\right)\boldsymbol{\Sigma}_{B_0}=\boldsymbol{\Sigma}_{0,0} \tag{12}$$

By substituting Eqs. (10) and (11) into Eq. (5), the position of any point of the actual linear feature represented by the line segment $\boldsymbol{Z}_0\boldsymbol{Z}_1$ is expressed as

$$\boldsymbol{Z}(t)=\boldsymbol{B}_0(d-t)-\frac{d-t}{d+a}\boldsymbol{B}_0(d+a)+\left(1-\frac{t}{d}\right)\boldsymbol{\mu}_0+\boldsymbol{B}_1(t)-\frac{t}{d+a}\boldsymbol{B}_1(d+a)+\frac{t}{d}\boldsymbol{\mu}_1,t\in[0,d] \quad (13)$$

where, $\boldsymbol{B}_i(t)$ is a Brownian motion stochastic process that the error vector $\boldsymbol{\xi}_i(t),(i=0,1)$, follows before considering endpoints' constraints, and $d\left(1-\frac{d}{d+a}\right)\boldsymbol{\Sigma}_{B_i}=\boldsymbol{\Sigma}_{i,i}$; and $B_i(t)\sim N(0,t\boldsymbol{\Sigma}_{B_i}),(i=0,1)$.

By combining Eqs. (13), (9) and (12), the positional expectation and variance of $\boldsymbol{Z}(t)$ are derived as

$$\boldsymbol{\mu}(t)=\left(1-\frac{t}{d}\right)\boldsymbol{\mu}_0+\frac{t}{d}\boldsymbol{\mu}_1 \quad (14)$$

$$\boldsymbol{\Sigma}(t)=\frac{(d-t)(a+t)}{ad}\boldsymbol{\Sigma}_{0,0}+\frac{t(d+a-t)}{ad}\boldsymbol{\Sigma}_{1,1} \quad (15)$$

The parameter a in Eq. (15) is unknown and needs to be determined. According to Eqs. (9) and (12), the value of parameter a affects the variance $\boldsymbol{\Sigma}_{B_i}(i=0,1)$. Therefore, a is the parameter that affects the magnitude of the modeling error. The method of determining the parameter a of the line segment model is further explained below.

In order to determine the parameter a, the relationship between the modeling error and characteristics of line segment $\boldsymbol{Z}_0\boldsymbol{Z}_1$ needs to be researched. The variation trend of error variance of the line segment $\boldsymbol{Z}_0\boldsymbol{Z}_1$ is discussed at the position $\boldsymbol{\mu}\left(\frac{d}{2}\right)$, which is the expected midpoint of $\boldsymbol{Z}_0\boldsymbol{Z}_1$. The difference between the variance $\boldsymbol{\Sigma}\left(\frac{d}{2}\right)$ of the line segment's error at $\boldsymbol{\mu}\left(\frac{d}{2}\right)$ and the variance $\boldsymbol{\Sigma}(0)$ of the left endpoint's error is defined as an error increment $\Delta\boldsymbol{\Sigma}$ as follows

$$\Delta\boldsymbol{\Sigma}=\boldsymbol{\Sigma}\left(\frac{d}{2}\right)-\boldsymbol{\Sigma}(0)=\frac{d}{2a}\boldsymbol{\Sigma}(0) \quad (16)$$

Firstly, only the position of the left endpoint is taken as the prior information in the Brownian motion. Under stationary and normal assumptions, the position of the actual feature follows the Brownian motion i.e., $\boldsymbol{B}(t)$; and the error's variance in the unit distance of the line segment is $\Sigma(\boldsymbol{B}(1))$. According to the positional constraint of the left endpoint and the conditional distributions of multivariate normal distributions[29], the position $\boldsymbol{Z}(t)$ of the linear feature follows the Brownian motion starting at the left endpoint $\boldsymbol{Z}_0$, namely

$$\boldsymbol{Z}(t)=\boldsymbol{B}(t)\mid_{\text{constraints}}=\boldsymbol{Z}_0+\boldsymbol{B}(t) \quad (17)$$

From the definition of the Brownian motion, any point $\boldsymbol{B}(t)$ at the Brownian motion follows a normal distribution, i.e., $\boldsymbol{B}(t)\sim N(0,t\Sigma(\boldsymbol{B}(1)))$, and the variance of $\boldsymbol{B}(t)$ is $\Sigma(\boldsymbol{B}(t))=t\Sigma(\boldsymbol{B}(1))$. Therefore, the variance of $\boldsymbol{Z}(t)$ is expressed as follows

$$\boldsymbol{\Sigma}_0(t)=\boldsymbol{\Sigma}_0(0)+\Sigma(\mathrm{B}(t))=\boldsymbol{\Sigma}_0(0)+t\Sigma(\boldsymbol{B}(1)) \quad (18)$$

The $\Sigma(\boldsymbol{B}(1))$ in Eq. (18) is a fixed value and is independent with of the distance d. According to Eq. (18) and the definition of error increment in Eq. (16), $\Delta\boldsymbol{\Sigma}_0$ is reformulated as follows

$$\Delta\boldsymbol{\Sigma}_0=\boldsymbol{\Sigma}_0\left(\frac{d}{2}\right)-\boldsymbol{\Sigma}_0(0)=\Sigma\left(\boldsymbol{B}\left(\frac{d}{2}\right)\right)=\frac{d}{2}\Sigma(\mathrm{B}(1))\propto d \quad (19)$$

In Eq. (19), error increment $\Delta\boldsymbol{\Sigma}_0$ is expressed as the product of the distance d with a constant independent of d. Therefore, the error increment $\Delta\boldsymbol{\Sigma}_0$ is directly proportional to the distance d between $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$, namely $\Delta\boldsymbol{\Sigma}_0\propto d$.

However, after the introduction of the positional distribution of right endpoint $\boldsymbol{Z}_1$ in the Brownian motion, $\Delta\boldsymbol{\Sigma}$ is constrained to be smaller than $\Delta\boldsymbol{\Sigma}_0$ because of the addition of more prior information. Therefore, according to Eq. (16), it is suggested to use the power function to depict the relationship between parameter a and the distance d, namely

$$a=d^c \quad (20)$$

Based on the above analysis, $\Delta\boldsymbol{\Sigma}$ increases with the increment of d, and the amplitude of the increasement is less than that in the directly proportional relationship, so the power c is greater than 0 and less than 1, i.e., $0<c<1$. Therefore, it is suggested in this paper that parameter a can be approximately expressed as

$$a=d^{0.5} \quad (21)$$

By substituting Eq. (21) into Eqs. (13) and (15), and combing Eqs. (13)—(15), the positional distribution of any point of the actual linear feature can be obtained. In the next section, a comprehensive error boundary will be constructed to describe the range of probable positions where actual features may occur.

2.3 Analytical expression of the error boundary

In this section, based on the error ellipses of positions between the endpoints of the line segment, the error boundary model of line segments is further constructed to describe the possible range of positions of the actual linear features represented by line segments.

When the standard deviations of the measuring errors at the endpoints of the line segment in the x and y directions are equal, the error ellipse of the point $\boldsymbol{\mu}(t)$ ($t \in [0,d]$) is a circle with center of $\boldsymbol{\mu}(t)$ and radius of $R\sqrt{\Sigma_x(t)}$, where R is the ratio of the axis' length of the error ellipse to that of the standard error ellipse. $\Sigma_x(t)$ is the variance of $\boldsymbol{Z}(t)$ in the x direction and is an element of the covariance matrix of $\boldsymbol{Z}(t)$, $\boldsymbol{\Sigma}(t)=\begin{pmatrix}\Sigma_x(t) & \Sigma_{x,y}(t)\\ \Sigma_{x,y}(t) & \Sigma_y(t)\end{pmatrix}$, which is calculated by Eq. (15) as follows

$$\Sigma_x(t)=\frac{(d-t)(a+t)}{ad}\delta+\frac{t(d+a-t)}{ad}\delta \tag{22}$$

The vertical distances from the envelope of the error ellipses of all points on the line segment to the positional expectation $\boldsymbol{\mu}_0\boldsymbol{\mu}_1$ are calculated. This vertical distance is denoted as a function $g(k)$ and is used to represent the width of the error band. It is assumed that the foot of perpendicular is $\boldsymbol{u}(k)$, and $\boldsymbol{u}(k)=(u_x(k),u_y(k))^{\mathrm{T}}$; $\boldsymbol{u}(k)$ is located on the line where the expected line segment $\boldsymbol{\mu}_0\boldsymbol{\mu}_1$ falls; k is the distance from $\boldsymbol{u}(k)$ to the expectation of the left endpoint ($\boldsymbol{\mu}_0$), and $k=\mathrm{sign}\left(\frac{\mu_{0x}-u_x(k)}{\mu_{0x}-\mu_{1x}}\right)|\boldsymbol{\mu}_0\boldsymbol{u}(k)|$.

Therefore, the vertical distance $g(k)$ from the boundary of the error band to the line segment $\boldsymbol{\mu}_0\boldsymbol{\mu}_1$ can be written as

$$g(k)=\max_{t\in[0,d]}\sqrt{R^2\Sigma_x(t)-(k-t)^2} \tag{23}$$

By substituting Eq. (22) into Eq. (23), the vertical distance $g(k)$ should satisfy the following function

$$g^2(k)=\max_{t\in[0,d]}h_k(t) \tag{24}$$

where $h_k(t)$ is a quadratic equation of t, which is expressed as follows

$$h_k(t)=-\left(\frac{2R^2\delta+ad}{ad}\right)\left(t-\frac{\frac{R^2\delta}{a}+k}{\frac{2R^2\delta}{ad}+1}\right)^2+R^2\delta-k^2+\left(\frac{2R^2\delta+ad}{ad}\right)\left(\frac{\frac{R^2\delta}{a}+k}{\frac{2R^2\delta}{ad}+1}\right)^2 \tag{25}$$

$h_k(t)$ increases and then decreases on $t\in R$, and takes the maximum value at $t=t_m(k)$. The maximum point $t_m(k)$ is obtained by calculating the root of the first derivative of the function $h_k(t)$, and is expressed as follows

$$t_m(k)=\frac{R^2d\delta+kad}{2R^2\delta+ad}=\frac{d}{2}+\frac{2adk-ad^2}{4R^2\delta+2ad} \tag{26}$$

By substituting Eq. (26) into Eq. (25), the maximum value of $h_k(t)$ is derived as $H(k)$

$$H(k)=\max_{t\in R}h_k(t)=h_k(t_m(k))=R^2\delta-k^2+\left(\frac{2R^2\delta+ad}{ad}\right)\left(\frac{R^2d\delta+kad}{2R^2\delta+ad}\right)^2 \tag{27}$$

By combining Eq. (24) and Eq. (27), it is known that when $0\leqslant t_m(k)\leqslant d$, $g^2(k)=H(k)$; when $t_m(k)<0$ or $t_m(k)>d$, $g^2(k)=h_k(0)$ or $h_k(d)$, and $g^2(k)\neq H(k)$. The position of the error band boundary of the line segment defined by the vertical distance $g(k)$ is summarized as follows

$$g^2(k)=\begin{cases}R^2\delta-k^2, & k\in[-R\sqrt{\delta},-\frac{R^2}{a}\delta)\\ -\frac{2R^2\delta}{2R^2\delta+ad}k^2+\frac{2R^2d\delta}{2R^2\delta+ad}k+R^2\delta+\frac{R^4d\,\delta^2}{2R^2a\delta+a^2d}, & k\in[-\frac{R^2}{a}\delta,d+\frac{R^2}{a}\delta]\\ R^2\delta-(k-d)^2, & k\in(d+\frac{R^2}{a}\delta,d+R\sqrt{\delta}]\end{cases} \tag{28}$$

According to Eq. (28), the error boundary of the line segment consists of three parts. When $k \in [-R\sqrt{\delta}, -\frac{R^2}{a}\delta)$ and $k \in (d+\frac{R^2}{a}\delta, d+R\sqrt{\delta}]$, the two ends of the error band boundary are circular arcs and are symmetric about the line segment's positional expectation $\boldsymbol{\mu}_0\boldsymbol{\mu}_1$. When $k \in [-\frac{R^2}{a}\delta, d+\frac{R^2}{a}\delta]$, the middle part of the error band boundary is two elliptic arcs intercepted by the interval of $k \in [-\frac{R^2}{a}\delta, d+\frac{R^2}{a}\delta]$. And the two elliptic arcs are symmetric with respect to $\boldsymbol{\mu}_0\boldsymbol{\mu}_1$.

By setting the first derivative of the boundary equation to zero, we can find that the extreme point of the error band boundary is near the middle point of the line segment. Furthermore, the second derivative of the boundary equation is negative. Therefore, it is found that from the left endpoint to the right endpoint, the error band's width increases firstly and then decreases. It is obvious that the error band's width in the line segment's middle part is greater than that at the two endpoints. Therefore, compared with the error band presented by Caspary and Scheuring[8], which is a concave set without considering the modeling error, the obtained Brownian bridge error band is a convex set, which describes the range of probable positions of the actual ground objects represented by line segments. In this model, the modeling and measuring errors are integrated to form the total error. It can be used to evaluate the overall accuracy and trustability of the line segment. It is designed for the case that there are both modeling and measuring errors contained in the line segment. In the next section, the Brownian bridge error model proposed in this paper will be verified and analyzed.

3 Numerical Experiments and Discussion

In this section, some numerically simulated line segments were used to verify the proposed Brownian bridge error model. Meanwhile, the effects of the parameters δ and d on the shape of the error band were analyzed. δ is the variance of the endpoint, and d is the distance between the positional expectation of the two endpoints.

The parameters of δ and d are two critical parameters that affect the size of the error band. Therefore, in order to analyze the different effects of δ and d on the error band, we divided the experiments into two cases. In Case 1, there were six line segments with the same variance δ but different distances d. In Case 2, there were six line segments with the same distance d but different variance δ.

It is assumed that the coordinate vector s_2 consists of two endpoints, $\boldsymbol{s}_2 = (\boldsymbol{z}_0^T, \boldsymbol{z}_1^T)^{\mathrm{T}} = (x_0, y_0, x_1, y_1)^{\mathrm{T}}$, and the covariance matrix of the line segment's two endpoints is $\boldsymbol{\Sigma} = \delta \boldsymbol{I}_{4\times4}$.

Tab.1 shows the endpoints coordinates of six line segments and their corresponding parameters of δ and d in Case 1. In this case, all the line segments' endpoints have the same error variance δ of 1 m^2, while their distances d between the positional expectations of the two endpoints are different.

Tab.2 shows the endpoints coordinates of the other six line segments and their corresponding parameters of δ and d in Case 2. In this case, all the line segments have the same distance d, while the error variances δ are different. In both cases, the parameter R of the error band boundary is set to 3, which means that the ellipses with three standard deviations were used to construct the error bands.

In order to evaluate the errors of the line segments in both cases, four parameters reflecting the characteristics of the error band were calculated and shown in Tab.1 and Tab.2, respectively. The calculated parameters include a, $Tr\left(\frac{d}{2}\right)$, $g\left(\frac{d}{2}\right)$, and p. Among these four parameters, a is the parameter affecting the modeling error; $Tr\left(\frac{d}{2}\right)$ represents the covariance matrix's trace at the point $(\mu\left(\frac{d}{2}\right))$, which has the maximum error on the line segment Z_0Z_1; $g\left(\frac{d}{2}\right)$ denotes the vertical distance from the error boundary to the point $\mu\left(\frac{d}{2}\right)$, which is the

maximum half-width of the proposed error band; p is the error's increment of $\mu\left(\frac{d}{2}\right)$ relative to the left endpoint Z_0, and $p=\frac{g\left(\frac{d}{2}\right)}{R\sqrt{\delta}}-1$.

It can be seen from the results in Tab.1 that, when the error variance δ of the line segment's endpoints is fixed, with the increasing of the distance d, the parameter a, the covariance matrix's trace $Tr\left(\frac{d}{2}\right)$, the vertical distance $g\left(\frac{d}{2}\right)$, and the error's increment p all increased.

Tab.1 Original data of line segments and derived error parameters in Case 1

Line No.	x_0 /m	y_0 /m	x_1 /m	y_1 /m	δ /m^2	d /m	a /m	$Tr\left(\frac{d}{2}\right)$ /m^2	$g\left(\frac{d}{2}\right)$ /m	p
1	0	0	25	0	1	25	5.000	7.000	5.613	0.871
2	0	0	30	0	1	30	5.477	7.477	5.801	0.934
3	0	0	35	0	1	35	5.916	7.916	5.986	0.990
4	0	0	40	0	1	40	6.325	8.325	6.121	1.040
5	0	0	45	0	1	45	6.708	8.708	6.260	1.087
6	0	0	50	0	1	50	7.071	9.071	6.389	1.828

As shown in Tab.2, when the distance d of line segments is fixed, with the increment of the error variance δ, the error trace $Tr\left(\frac{d}{2}\right)$ and the vertical distance $g\left(\frac{d}{2}\right)$ also increase, but the parameter a and the error increment p remain unchanged.

Tab.2 Original data of line segments and derived error parameters in Case 2

Line No.	x_0 /m	y_0 /m	x_1 /m	y_1 /m	δ /m^2	d /m	a /m	$Tr\left(\frac{d}{2}\right)$ /m^2	$g\left(\frac{d}{2}\right)$ /m	p
1	0	0	40	0	0.5	40	6.325	4.162	4.328	1.040
2	0	0	40	0	1	40	6.325	8.325	6.121	1.040
3	0	0	40	0	1.5	40	6.325	12.487	7.496	1.040
4	0	0	40	0	2	40	6.325	16.649	8.656	1.040
5	0	0	40	0	2.5	40	6.325	20.811	9.677	1.040
6	0	0	40	0	3	40	6.325	24.974	10.601	1.040

Based on the above two cases of experiments, it can be seen that: ① The parameter a, the error indicators of the line segment ($Tr\left(\frac{d}{2}\right)$ and $g\left(\frac{d}{2}\right)$) and the error's increment p increased with the increasing of the distance d; ② The value of the error variance δ only affects the value of the intermediate point' error of the line segment but does not affect the parameter a and error's increment p.

Fig. 2 shows the boundary of the Brownian bridge error band of line segment 4 in Tab.1 (also the same line segment 2 in Tab.2). This Brownian bridge error band is composed of two circular arcs on the left and right sides and two elliptic arcs on the upper and lower side. In this figure, the four red points are the intersection of the circular arcs and the ellipse arcs. The left and right circles are the endpoints' error ellipses, $g\left(\frac{d}{2}\right)$ is the vertical distance (half-width of the error band) from the error boundary at the point with the maximum error to the positional expectation $\mu_0\mu_1$. $r(0)$ is the radius of the error ellipse at the endpoint, and $r(0)=R\sqrt{\delta}$. The width of the error band increases and then decreases from the left endpoint to the right endpoint. The maximum value is reached at $\mu\left(\frac{d}{2}\right)$. The shape of

the proposed Brownian bridge error band is spindle-shaped which is wider in the middle part of the line segment than at the endpoints. The shape of the error band means that the position errors in the middle part of the line segment are larger than that at the two endpoints.

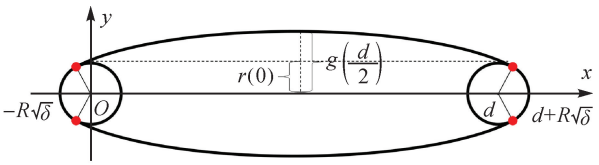

Fig.2 Boundary of Brownian bridge error band

According to the derived error parameters of line segment 4 shown in Tab.1, the width $G\left(\frac{d}{2}\right)$ of the error band at the point with the maximum error on the line segment Z_0Z_1 can be computed as

$$G\left(\frac{d}{2}\right)=2g\left(\frac{d}{2}\right)=2\sqrt{\left(1+\frac{\sqrt{d}}{2}\right)}r(0)=12.241\ \text{m}$$

The width $G\left(\frac{d}{2}\right)$ of the error band at the point with the maximum error is $\sqrt{\left(1+\frac{\sqrt{d}}{2}\right)}$ times the diameter $2r(0)$ of the error ellipse at the endpoint Z_0. In this example, the error band's width at the middle point of the line segment increased by about 104% compared with that at the endpoints. Regarding the other line segments in the examples, the error band's width at the middle point increased by at least 87%. On the contrary, with respect to the line error band model proposed by Caspary and Scheuring[8], the width of the error band at the middle point is about 29% smaller than that at the endpoints. The Brownian bridge error model of the line segment become a convex set because of the introduced modeling error. The range and width of the Brownian bridge error band can be used as indices to evaluate the positional accuracy of the spatial line segment.

4 Conclusions

A Brownian bridge error model of line segments is proposed for modeling the positional errors of spatial line segments containing both modeling and measuring errors. The stochastic process of Brownian motion is used to depict the positional distributions of the actual geographic features represented by the line segments. In the Brownian motion, the measuring error distributions of the endpoints are used as the prior constraints, such that the Brownian motion is constrained and become the Brownian bridge model. A propagation method of modeling-measuring integrated error is proposed by using the Brownian bridge. Finally, the analytical expression of the Brownian bridge error boundary is established.

Two cases of numerical experiments were conducted to verify the proposed error modeling method. Several conclusions can be drawn. ① The proposed error band is obtained to be a convex set. Compared with the error band proposed by Caspary and Scheuring[8] which is dumbbell-shaped, the proposed error band is spindle-shaped, which is wider in the middle part of the line segment than at the endpoints. The shape of the proposed error band indicates that the position errors in the middle part of the line segment are larger than that at the two endpoints. This error distribution conclusion is consistent with the regulations in surveying and mapping, which is the more distant from the known and observed point, the larger the error is. ② The shape of the proposed method is related the endpoints' variance and the length of the line segment. And a proximate method for estimating the modeling error parameter is proposed.

In summary, the proposed Brownian bridge error model is characterized by taking into account both the modeling and measuring errors of the line segments. It provides a solution for the challenge of error modeling for geographic features that has received a lot of research attention over the past more than 20 years. It can be used to evaluate line segments' overall accuracy and quality trustability. Nevertheless, the error boundary of this model is established with the assumption that the error distributions of the endpoints are independent. In the future study, the correlation between the endpoints of line segments will be further considered.

References

[1] National Center for Geographic Information and Analysis. The

research plan of the national center for geographic information and analysis[J]. International Journal of Geographical Information Systems, 1989, 3(2): 117-136. DOI: 10.1080/02693798908941502.

[2] GOODCHILD M F. Imprecision and spatial uncertainty[M]// SHEKHAR S, XIONG Hui. Encyclopedia of GIS. Boston: Springer, 2008: 480-483. DOI: 10.1007/978-0-387-35973-1_592.

[3] GONG Jianya, LI Dajun. Entropy-based models for positional uncertainty of line segments in GIS[J]. Survey Review, 2011, 43(322): 390-401. DOI: 10.1179/003962611X13055561708786.

[4] TONG Xiaohua, XIE Huan, LIU Shijie, et al. Uncertainty of spatial information and spatial analysis[M]//LENG Shuying, GAO Xizhang, PEI Tao, et al. The Geographical Sciences During, 1986-2015: From the Classics to the Frontiers. Singapore: Springer, 2017: 511-522. DOI: 10.1007/978-981-10-1884-8_25.

[5] SHI Wenzhong, CHEN Pengfei, ZHANG Xiaokang. Reliability analysis in geographical conditions monitoring[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(10): 1620-1626. DOI: 10.11947/j.AGCS.2017.20170377.

[6] PERKAL J. On epsilon length[J]. Bulletin de i'Academie Polonaise des Sciences, 1956, 4: 399-403.

[7] CHRISMAN N R. A theory of cartographic error and its measurement in digital data base[C]//Proceedings of Auto-Carto. Bethesda: American Congress on Surveying and Mapping, 1982, 5: 159-168.

[8] CASPARY W, SCHEURING R. Positional accuracy in spatial databases[J]. Computers, Environment and Urban Systems, 1993, 17(2): 103-110. DOI: 10.1016/0198-9715(93)90040-C.

[9] GOODCHILD M F, HUNTER G J. A simple positional accuracy measure for linear features[J]. International Journal of Geographical Information Science, 1997, 11(3): 299-306. DOI: 10.1080/136588197242419.

[10] LIU Wenbao. A theory of uncertainty in spatial data within GIS [D]. Wuhan: Wuhan Technical University of Surveying and Mapping, 1995.

[11] LEUNG Y, YAN Jianping. A locational error model for spatial features[J]. International Journal of Geographical Information Science, 1998, 12(6): 607-620. DOI: 10.1080/136588198241699.

[12] SHI Wenzhong. A generic statistical approach for modelling error of geometric features in GIS[J]. International Journal of Geographical Information Science, 1998, 12(2): 131-143. DOI: 10.1080/136588198241923.

[13] TONG Xiaohua, SHI Wenzhong, LIU Dajie. Uncertainty model of circular curve features in GIS[J]. Acta Geodaetica et Cartographica Sinica, 1999, 28(4): 325- 329.

[14] SHI Wenzhong, LIU Wenbao. A stochastic process-based model for the positional error of line segments in GIS[J]. International Journal of Geographical Information Science, 2000, 14(1): 51-66. DOI: 10.1080/136588100240958.

[15] TONG Xiaohua, SUN Tong, FAN Junyi, et al. A statistical simulation model for positional error of line features in Geographic Information Systems (GIS)[J]. International Journal of Applied Earth Observation and Geoinformation, 2013, 21: 136-148. DOI: 10.1016/j.jag.2012.08.004.

[16] SHI Wenzhong, YOU Yangsheng, SHAO Pan, et al. Standard deviation of line objects in geographic information science[J]. Annals of GIS, 2014, 20(1): 39-51. DOI: 10.1080/19475683.2013.862297.

[17] WAN Yiliang. Research on the theory for spatial data quality inspection and assessment [D]. Wuhan: Wuhan University, 2015.

[18] LONGLEY P A, GOODCHILD M F, MAGUIRE D J, et al. Geographic information science and systems[M]. 4th ed. [S. l.]: Wiley, 2015.

[19] ALESHEIKH A A, BLAIS J A R, CHAPMAN M A, et al. Rigorous geospatial data uncertainty models for GISs[M]// JATON A, LOWELL K. Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources. Chelsea, Michigan: Ann Arbor Press, 195-202.

[20] TINNACHOTE C, CHEN X. An approach for object-based positional error model[J]. Journal of Spatial Science, 2005, 50(1): 1-12. DOI: 10.1080/14498596.2005.9635034.

[21] DE BRUIN S. Modelling positional uncertainty of line features by accounting for stochastic deviations from straight line segments[J]. Transactions in GIS, 2008, 12(2): 165-177. DOI: 10.1111/j.1467-9671.2008.01093.x.

[22] LAN Yueming, TAO Benzao. Combined quantification of line feature uncertainties in GIS[J]. Geomatics and Information Science of Wuhan University, 2003, 28(5): 559-561.

[23] HU Shengwu, YU Xu. Research progress about spatial data uncertainty [J]. Journal of Henan Polytechnic University (Natural Science), 2016, 35(6): 815-822.

[24] SUN Tong, TONG Xiaohua. A comprehensive curve error model based on broken-line approximation in curve position[J]. Engineering of Surveying and Mapping, 2010, 19(3): 26-30. DOI: 10.3969/j.issn.1006-7949.2010.03.009.

[25] CAI Jianhong, LI Deren, ZHU Daolin. Positional uncertainty and visualization of curves with modeling errors added[J]. Science of Surveying and Mapping, 2012, 37(1): 44-46.

[26] COX D R, MILLER H D. The theory of stochastic processes [J]. Journal of the Royal Statistical Society Series C: Applied Statistics, 1965, 16(2). DOI:10.1007/978-3-642-61943-4.

[27] HORNE J S, GARTON E O, KRONE S M, et al. Analyzing animal movements using Brownian bridges [J]. Ecology, 2007, 88(9): 2354-2363. DOI: 10.1890/06-0957.1.

[28] BEARUP D, BENEFER C M, PETROVSKII S V, et al. Revisiting Brownian motion as a description of animal movement: a comparison to experimental movement data[J]. Methods in Ecology and Evolution, 2016, 7(12): 1525-1537. DOI: 10.1111/2041-210X.12615.

[29] ARNOLD B C. Characterizations involving conditional specification[J]. Journal of Statistical Planning and Inference, 1997, 63(2): 117-131. DOI: 10.1016/S0378-3758(97)00005-0.