Chin. Geogra. Sci. 2012 Vol. 22 No. 6 pp. 695–704
doi: 10.1007/s11769-012-0544-y
www.springerlink.com/content/1002-0063

Errors Prediction for Vector-to-Raster Conversion Based on Map Load and Cell Size

LIAO Shunbao[1, 2], BAI Zhongqiang[2, 3], BAI Yan[2, 3]

(*1. College of Environment and Planning, Henan University, Kaifeng 475004, China; 2. Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China*; *3. Graduate University of Chinese Academy of Sciences, Beijing 100049, China*)

Abstract: Vector-to-raster conversion is a process accompanied with errors. The errors are classified into predicted errors before rasterization and actual errors after that. Accurate prediction of the errors is beneficial to developing reasonable rasterization technical schemes and to making products of high quality. Analyzing and establishing a quantitative relationship between the error and its affecting factors is the key to error prediction. In this study, land cover data of China at a scale of 1∶250 000 were taken as an example for analyzing the relationship between rasterization errors and the density of arc length (DA), the density of polygon (DP) and the size of grid cells (SG). Significant correlations were found between the errors and DA, DP and SG. The correlation coefficient (R^2) of a model established based on samples collected in a small region (Beijing) reaches 0.95, and the value of R^2 is equal to 0.91 while the model was validated with samples from the whole nation. On the other hand, the R^2 of a model established based on nationwide samples reaches 0.96, and R^2 is equal to 0.91 while it was validated with the samples in Beijing. These models depict well the relationships between rasterization errors and their affecting factors (DA, DP and SG). The analyzing method established in this study can be applied to effectively predicting rasterization errors in other cases as well.

Keywords: vector-to-raster conversion; rasterization; error prediction; map load; cell size

Citation: Liao Shunbao, Bai Zhongqiang, Bai Yan, 2012. Errors prediction for vector-to-raster conversion based on map load and cell size. *Chinese Geographical Science*, 22(6): 695–704. doi: 10.1007/s11769-012-0544-y

1 Introduction

There are two basic formats of geospatial data used in Geographical Information Systems (GIS): vector and raster (Maguire *et al.*, 1991). Theoretically, vector data have higher accuracy than raster data. However, raster data are more suitable for spatial modeling and analysis than vector data (Bates and De Roo, 2000; Demers, 2002). Therefore, it is customary to convert the data formats between the two, that is, vector-to-raster which is the so-called rasterization, and raster-to-vector, the so-called vectorization.

As the technology of remote sensing advances forward, raster data have become more and more popular than vector data. Raster is now the most dominant format of data sources. The researches related to raster data processing, storage, analysis and application have once again drawn a lot of attention in the GIS field (Chen, 2002). Data products saved in raster format are increasing rapidly. Rasterization is beneficial for performing comprehensive analysis of GIS data in different sources and structures.

In recent years, rasterization has become a generalized practice, not only in GIS but also in Spatial Data Infrastructures (SDI), in 3D modeling (Settgast *et al.*, 2004) and in 'scenes' (Roth *et al.*, 2004). Moreover, rasterization methods have been widely applied to different cartographic studies. For example, Su *et al.* (1997; 1998) proposed a map generalization method, by which users apply first the rasterization process and then return

Received date: 2011-08-08; accepted date: 2012-01-11
Foundation item: Under the auspices of Strategic Priority Research Program of Chinese Academy of Sciences (No. XDA05050000), Special Program for Informatization of Chinese Academy of Sciences (No. INF0-115-C01-SDB3-02)
Corresponding author: LIAO Shunbao. E-mail: liaosb@igsnrr.ac.cn

to the vector representative model. This method was discussed in detail by Peter and Weibel (1999).

However, rasterization always suffers a loss of information with regard to its spatial accuracy. The loss of information inherited to rasterization is usually defined by a change in the area of the raster model. For the rasterization of polygon features, errors are found in area, perimeter, shape, structure, position and attributes (Carvers and Brunsdon, 1994). These errors vary as data source, rasterization algorithm and data structure change (Burrough, 1986). Given the same input, the outcome may still be different because a variety of rasterization algorithms have been used (Knaap, 1992). Shortridge (2004) concluded that rasterization errors are sensitive to cell size, polygonal shape and structure. Using limited samples, Shortridge (2004) investigated the influence of different polygonal areas in ordinal class raster images. Bregt *et al.* (1991) proposed an error analysis method that combines dual conversion and boundary index. Zhou *et al.* (2006) presented an Equal Area Conversion (EAC) model for rasterization, which minimizes the error in area but at the cost of losing boundary features. Wu *et al.* (2009) proposed a winding number algorithm based on the rotation angle theory in computational geometry. Wang *et al.* (2006) proposed an optimization algorithm which aims to minimize the rasterization errors in area. These studies demonstrate that error evaluation is an important component of rasterization.

Generally, there are two ways to present rasterization errors. The first is to predict the error before applying the rasterization process and the second is to determine the final error after applying the conversion.

Quite a few studies have attempted at predicting rasterization errors in the past. Frolov and Maling (1969) presented a probability statistical model that estimates rasterization errors based on the 'Bufon's needle' theory. They used the average length of polygon edges (the number of perimeter cells) as the variable of their model. Goodchild (1980) and Goodchild *et al.* (1996) revised the work of Frolov and Maling. They conducted a total of 50 000 simulations and obtained a solution similar to Frolov and Maling. Lawrence and Ripple (1996) calculated the perimeter cells of polygons using an approximation with simple geometric shapes. Crapper (1980; 1981) focused on the calculation of the number of perimeter cells. He added two new parameters (perimeter contortion and sinuosity) to the approach of Frolov and Maling, and established the mean contortion. Carvers and Brunsdon (1994) studied the dependence of error area in quadtrees on the complexity of the polygon perimeter and the depth of the tree structure. The results showed that the error depends on the polygon complexity, which is in turn determined as a function of the parameter of angular complexity (mean angle among segments) and the root level of the quadtree.

Up to now, the prediction of rasterization errors has relied exclusively on theoretical models. They are difficult to be applied in practical evaluation of rasterization errors. For this reason, a few pragmatic methods have been developed for assessing rasterization errors in recent years. Yang and Zhang (2001) took the land-use data at a scale of 1：100 000 as an example to investigate the degree of precision losses for all land categories with respect to different grid-cell sizes. They established the relationship model among precision loss, the average size of polygon patch and grid cell size. Liu *et al.* (2001) concluded that the relative error of rasterization is in proportion to the complexity of the land cover within a given region through analyzing the rasterization errors of land-cover data of China at a scale of 1：100 000. The more complex the region is, the larger the error becomes.

Conventional methods used for evaluating rasterization errors are based on the entire region rather than on each grid cell. They focused on the total error in the entire study region and ignored the spatial distribution of errors, which often increases in one cell but decreases in another. As a consequence, for most cases rasterization errors are greatly underestimated (Chen *et al.*, 2007). Total error can not reflect the irregularity of errors in the entire region. Ignoring local errors can cause the loss of important information and even lead to wrong conclusions (Lunetta *et al.*, 1991). To mitigate this problem, Liao and Bai (2010) presented a new error evaluation method based on grid cells (EEM-BGC). The method follows four steps. First, the area of each land category inside a square is represented in a vector format. The size and location of the square are exactly the same as those of a grid cell that is to be generated by rasterization. Second, the area is treated as the attribute of the grid cell. Vector data are rasterized into n grids, where n is the number of land categories. Then, the relative area error resulting from rasterization for each land category in the grid cell is calculated in raster format. Lastly, the average of the relative area error for all land categories in the grid cell is computed with the area of a land category as weight.

Compared with the existing methods of theoretically predicting errors before rasterization, errors evaluation after the conversion is more practical. However, sometimes people also want to know how much the errors are before the conversion is carried out so that an appropriate rasterization scheme can be made, for instance, by selecting a suitable size of grid cells. Therefore, the best approach for assessing rasterization errors is to combine the technologies of errors evaluation after the conversion, together with those of errors prediction before rasterization.

In this study, the method of EEM-BGC was used to evaluate the errors and take the rasterization of land cover data of China at a scale of 1：250 000 as an example to analyze the relationships between rasterization errors and the density of arcs length (DA), the density of polygons (DP) and the size of grid cells (SG), and to express these relations in mathematical terms. Using these equations, rasterization errors can be predicted with given DA, DP and SG.

2 Data and Method

2.1 Data

The data used in this study include: 1) vector data of land cover of China at a scale of 1：250 000; 2) raster data of land cover of China at a resolution of 1 km by 1 km; 3) vector data of provincial administrative boundary of China composed of 32 units (provinces/municipalities/autonomous regions; Macao and Hong Kong are included to Guangdong Province); 4) raster data of land cover of Beijing, China at the resolutions of 0.1 km by 0.1 km, 0.2 km by 0.2 km, 0.5 km by 0.5 km, 1 km by 1 km, 2 km by 2 km, 5 km by 5 km, and 10 km by 10 km, respectively; 5) vector data of county level administrative boundary of Beijing composed of 11 units; 6) raster data of land cover of Hebei Province at the resolutions of 0.5 km by 0.5 km, 1 km by 1 km, 2 km by 2 km, 4 km by 4 km, 8 km by 8 km, 16 km by 16 km and 32 km by 32 km, respectively; and 7) raster data of land cover of Sichuan Province at the resolutions of 0.6 km by 0.6 km, 0.7 km by 0.7 km, 0.8 km by 0.8 km, 0.9 km by 0.9 km, 1 km by 1 km, 2 km by 2 km, 3 km by 3 km, 4 km by 4 km, 5 km by 5 km, 6 km by 6 km, 7 km by 7 km, 8 km by 8 km, 9 km by 9 km, 10 km by 10 km, 20 km by 20 km and 30 km by 30 km, respectively.

Among these data, those of Beijing were used to analyze and establish a model, and those of Hebei and Sichuan were used for validating the model. The reason that Beijing was selected as a modeling area is that there is complex terrain and diverse land covers in Beijing. It is surrounded by mountains in the west, north and northeast, and occupied by a plain in the east. It consists of mountains, hills, plains and valleys. It is representative as a modeling area even though its area is small.

The spatial distribution of regions used for establishing and validating the model is shown in Fig. 1. Figure 2 is the map of administrative divisions of Beijing at the county level.

The database of land cover of China at the scale of 1：250 000 categorized the land covers into 6 level-1 classes and 25 level-2 classes (Zhang *et al.*, 2009). In this study, there are only 15 level-2 classes out of the 25 total level-2 classes in Beijing, 20 out of the 25 total in Hebei, and 23 out of the total 25 in Sichuan. Needless to say, the smaller sample size of classes should not bear any impact on the application of the data in this study.

2.2 Method

2.2.1 Evaluation index of errors

In order to better depict the irregularity of rasterization errors in the entire region, Error Evaluation Method Based on Grid Cells (EEM-BGC) was opted to used as the evaluation index (Liao and Bai, 2010). The principles and evaluation steps of EEM-BGC are described as follows.

Suppose that vector data are composed of n polygons of land categories inside a given grid cell. They are labeled as P_1, P_2, …, and P_n, respectively. Their corresponding areas are A_1, A_2, …, A_n, and the total area is A, that is,

$$A = \sum_{i=1}^{n} A_i \tag{1}$$

It is further assumed that polygon P_n has the largest area among all the polygons (Fig. 3). Based on the largest area theorem, category attribute of the grid cell is replaced by the attribute of polygon P_n after it is rasterized. This results in area gain or loss for each polygon of the land category inside the grid cell. The details are given below.

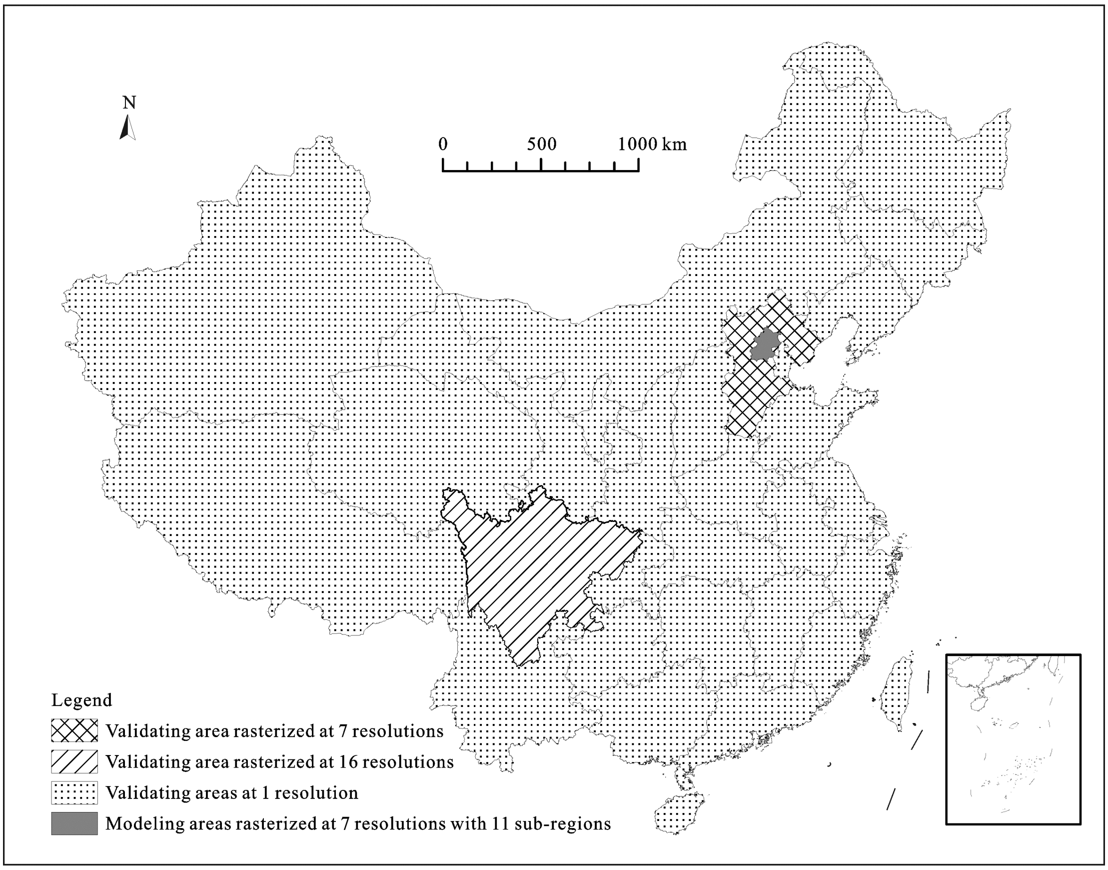


Fig. 1 Distribution map of modeling area and validating areas

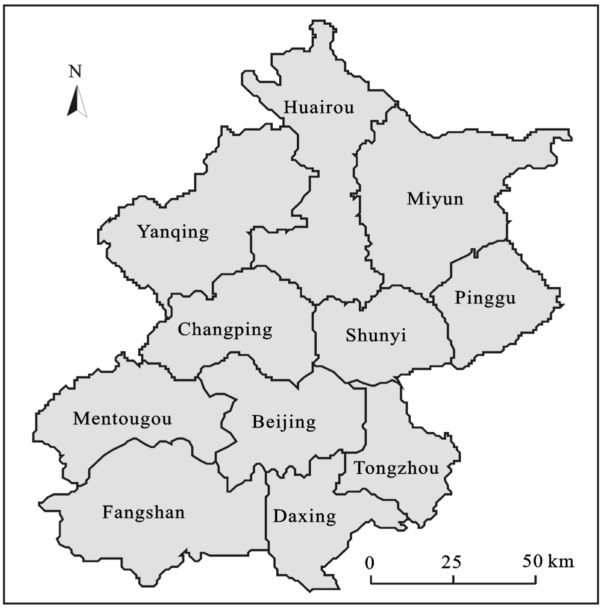


Fig. 2 Administrative division map of Beijing at county level

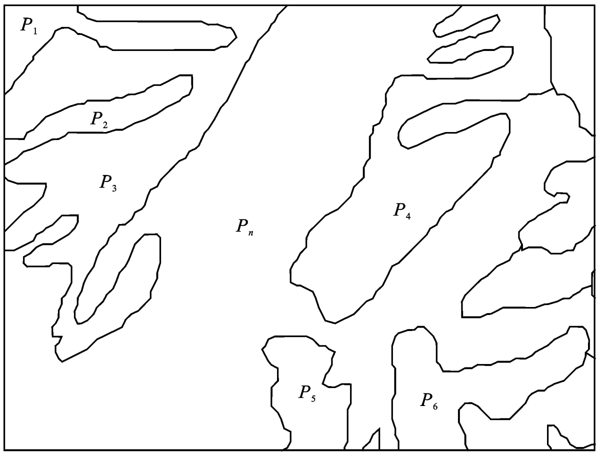


Fig. 3 A grid cell with n land categories and category n having largest area within grid

(1) The areas of polygons P_1, P_2, …, and P_{n-1} are A_1, A_2, …, A_{n-1} before rasterization. They are all reduced to zero after rasterization. The values of their absolute errors are exactly A_1, A_2, …, A_{n-1}, respectively. The relative errors are all 100%.

(2) The area of polygon P_n is A_n before rasterization and becomes A (total area of the grid cell) after rasterization. The absolute error is $(A - A_n)$ and relative error is $(A - A_n) / A_n \times 100\%$.

The relative error is more scientifically meaningful and reasonable than the absolute error. Polygons of different patch sizes contribute differently to the general error even if they have the same relative errors. Similarly, given the same relative error, the larger area a polygon has, the larger a contribution to the general relative error of a grid cell it makes. Therefore, it is justifiable to use polygon area as weight to calculate the General Relative Error of a grid cell (*GRE*), that is,

$$GRE = \sum_{i=1}^{n-1} \left[1 \times \left(A_i / A\right)\right] + \left[\left(A - A_n\right) / A_n\right] \times \left(A_n / A\right) \quad (2)$$

From Equation (1), $\sum_{i=1}^{n-1} A_i = A - A_n$ can be got. Therefore, Equation (2) can be rewritten as follows.

$$GRE = 2\left(A - A_n\right) / A \quad (3)$$

It is easy to show that *GRE* is in the range of [0, 2]. To make it more consistent with the commonly used relative error that is in the range of [0, 1], *GRE* should be divided by a facor of 2, that is,

$$GRE = \left(A - A_n\right) / A \quad (4)$$

The result of Equation (4) is the general relative error based on a grid cell. A general relative error for entire region composed of the grid cells can be got by calculating average value of relative errors of all grid cells in the region. The general relative errors will be used as evaluation index to analyze relationship between the errors and their affecting factors. It should be noted that one vector coverage should be rasterized into multiple layers of grid data in order to meet Equation (4). The number of the layers is equal to the number of land cover categories. The value of a grid cell represents the percentage of area of a specific land cover category in the cell.

2.2.2 Error prediction method

Rasterization errors have close relationship not only with the complexity of the polygon perimeter but also with the size of grid cells. In this study, DA and DP in a region were chosen as the complexity index of the region to establish quantitative expressions between the errors and DAs, DPs and SGs by means of multivariate regression analysis. If the expressions can fit well with the relationship among errors and DAs, DPs and SGs, they can be used to predict rasterization errors effectively. Because it is easy to calculate DAs and DPs in a GIS software, in practice it takes much less time to calculate DA and DP than to perform vector-to-raster conversion. Different DA, DP and SG lead to different errors. When the quantitative relationship is known, it would be able to be decided how much the resolution should be set before rasterization. If the rasterization error is too high in a certain resolution, the error can be reduced by changing the resolution of grid cells. In this way, the magnitude of errors before rasterization can not only be estimated approximately but also qualified raster data products can be output at the end.

3 Results

3.1 Preliminary analysis of data

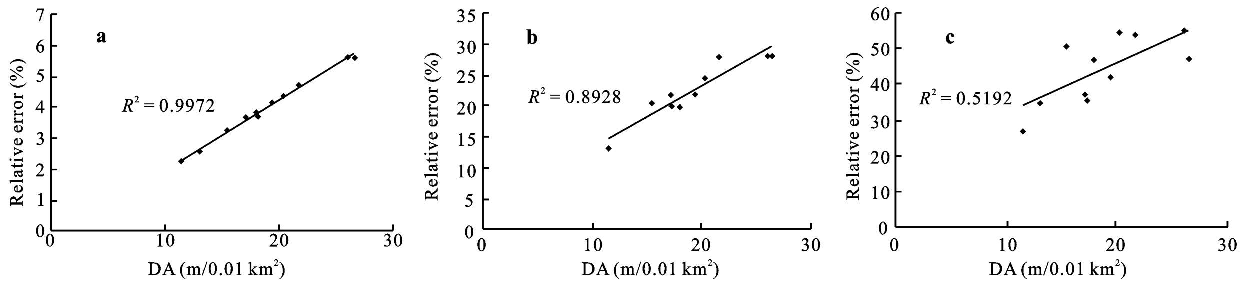
Beijing was chosen as the area for modeling. Vector data of land cover of Beijing were rasterized at seven different resolutions. Therefore, seven grid files were generated. First, Equation (4) was used to calculate *GRE* of each grid cell for all the seven resolutions. Then general rasterization errors, DAs and DPs for all the 11 county units at the seven resolutions were calculated by using the ARCGIS software. Lastly, the relationships between the general rasterization errors and DAs, DPs and SGs were analyzed, respectively.

The relationship between the general rasterization error and DA is shown in Fig. 4. It should be noted that only three of seven resolutions in Fig. 4 were presented to save space. It can be seen from Fig. 4 that the errors have significant linear correlation with DA though the correlation coefficient decreases with increase of grid cell size. The same characteristics were found in the relationship between rasterization error and DP (Fig. 5).

The relationship between the general rasterization error and SG is shown in Fig. 6. As in Fig. 4 only three out of the total 11 regions are shown to save space. It can be seen that the errors have strong logarithmic correlation with the size of grid cells. The correlation coefficients from different regions have no obvious differences from each other.

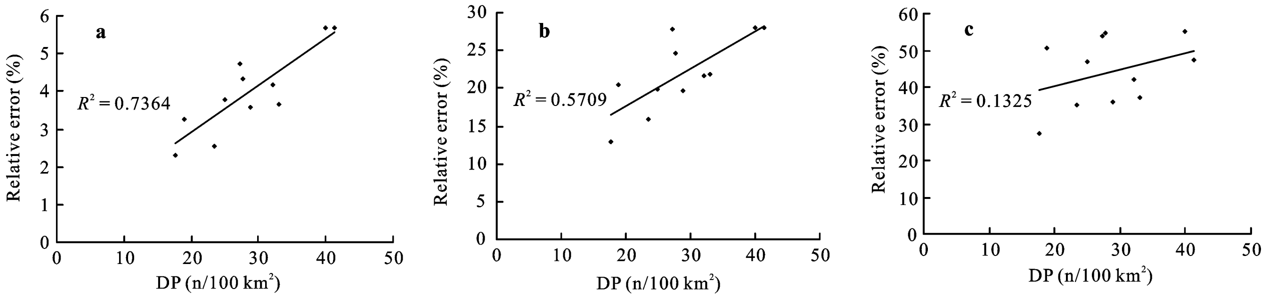
3.2 Modeling for predicting rasterization error

It can be concluded that the errors resulting from polygonal vector-to-raster conversion have significant linear correlations with DA and DP, and a strong logarithmic correlation with SG from the above analysis. Based



Resolutions for a, b, and c are 0.1 km,1 km and 10 km, respectively

Fig. 4 Relationship between rasterization error and density of arcs length (DA) in Beijing



Resolutions for a, b, and c are 0.1 km,1 km and 10 km, respectively; *n* represents number of polygons

Fig. 5 Relationship between rasterization error and density of polygons (DP) in Beijing

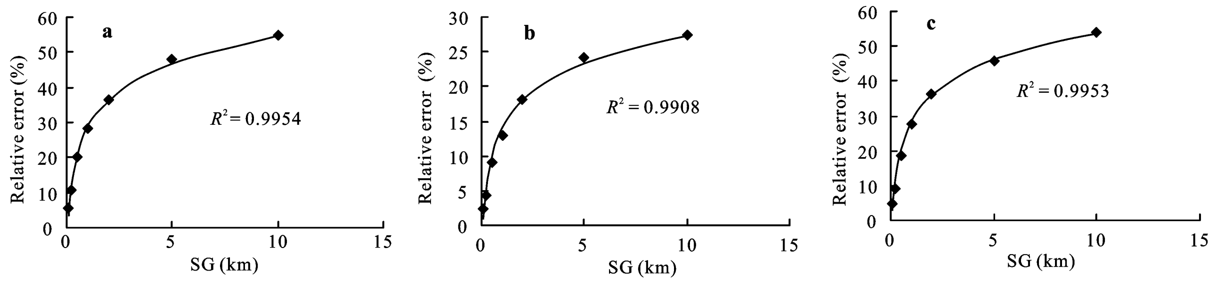


Fig. 6 Relationship between rasterization error and size of grid cells (SG) in Miyun (a), Beijing proper (b), and Fangshan (c)

on this finding, in the following 77 samples (11 regions multiplied by 7 resolutions) collected in Beijing were used to establish a quantitative expression between rasterization errors and DAs, DPs and SGs.

A multivariate linear regression analysis was conducted with the error as the dependent variable, and DA, DP and the logarithm of SG as independent variables. The following expression was derived,

$$E = -55.4333 + 1.552d_a - 0.444d_p + 8.912\ln(s_g) \quad (5)$$

where E represents the general relative error resulting from rasterization in a given region, d_a represents DA, d_p represents DP and s_g represents SG.

The relationship between actual error and that predicted by Equation (5) for the 77 samples was shown in Fig. 7. It can be seen that the errors calculated by Equation (5) are consistent with the actual error. The R^2 reaches 0.9462 and the slope of the regression is close to 1 (0.9463). This indicates that Equation (5) is capable of depicting well the dependences of rasterization error on its affecting factors. Before this model can be used as a general tool for predicting rasterization errors in all regions, it has to be kept in mind that the samples used for constructing the model were collected in Beijing, which only covers about 1/600 of the total area of China. Therefore, it can not be applied in nationwide for predicting rasterization errors until it is verified.

3.3 Model validation

3.3.1 Representativeness analysis of validation areas

The prediction model as represented by Equation (5) is validated at two spatial scales, i.e., the nationwide and

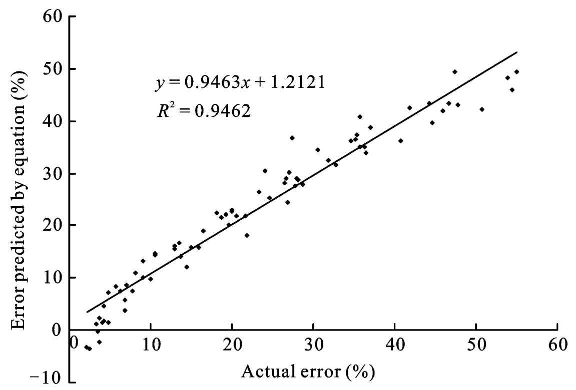


Fig. 7 Relationship between actual error and that predicted by Equation (5) in Beijing proper

the regional. China has an area of 9.6 × 10^6 km^2. It has complicated natural conditions and diverse types of land cover. Therefore, it would be convincing to use nationwide data to verify the model. At the regional scale, the validation is carried out using samples collected in Hebei Province in North China and Sichuan Province in Southwest China (Fig. 1). Hebei Province has high terrains in the northwest where the Yanshan Mountains are located, and low flat topography in the southeast which is a part of the North China Plain. The terrain of the province consists of plateaus, mountains, hills, plains, and basins surrounded by mountains. There are 20 out of the total 25 land cover categories. Therefore, the land cover categories are typical for model validation. Natural conditions in Sichuan Province are also complicated. There exist hills and mountains in the east and a great plain, the famous Chengdu Plain in the center. The west is adjacent to the Qinghai-Tibet Plateau. So land cover in the province is also representative.

The nationwide data were rasterized at a resolution of 1 km by 1 km. As a result, the influence of SG on rasterization error was not validated at the nationwide scale. Nevertheless, the influences of DA and DP on the error were carefully validated in nationwide. Through rasterization, 32 provincial units of grid data products were obtained. Hong Kong and Macao were included in Guangdong Province. Islands and enclaves were excluded from the calculation. Each product consists of 25 layers of data because there are 25 categories of land cover in China. The values of grid cells represent the percentages of a specific type of land cover in the cell. Equation (4) was first used to calculate *GRE*s of each cell and then the general relative errors for each unit were derived. Thirty-two general relative errors corresponding to 32 provincial units in the same resolution were got.

At the regional scale, the data collected in Hebei Province were rasterized at seven different resolutions and those in Sichuan at 16 resolutions. Therefore, the influence of SG on the error was fully considered though the influences of DA and DP on the error were not validated at the regional scale for DA and DP remain unchanged at the different resolutions. Following the same procedure that was used to process the nationwide data, 23 general relative errors corresponding to the 23 different resolutions in two regions, i.e., Hebei and Sichuan, were obtained. Up to now, 55 samples have been obtained, each of which includes 4 parameters: the actual error, DA, DP and SG. The samples are then used to validate the model (Equation (5)). It should be kept in mind that the cases of Hebei and Sichuan at the resolution of 1 km by 1 km have been counted twice since they were included in both the nationwide and regional scales. Therefore, the cases of Hebei and Sichuan at 1 km by 1 km should be removed. The remaining 53 samples were then used for verifying the model.

3.3.2 Validation

Equation (5) was used to calculate the errors of all samples with DA, DP and SG as input parameters, and the resulting errors were compared with the actual error that was generated in Section 3.3.1. The results are presented in Fig. 8. It shows clearly that there exists a close correlation between the actual errors and the predicted errors by Equation (5). The linear correlation coefficient (R^2) reaches 0.9109 and the slope of the regression is close to 1. The correlation is statistically significant at the 0.001

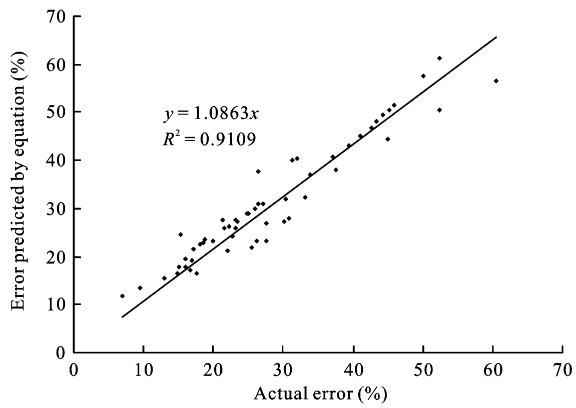


Fig. 8 Relationship between actual error and that predicted by Equation (5) at nationwide scale

level (R < 0.4433; number of samples is 53). It can be concluded that although the model was established based on samples collected in Beijing, it describes very well the relationship between rasterization error and DA, DP and SG for all other regions. It can be applied to predicting rasterization errors at nationwide scale.

3.3.3 Reversed consideration

In the preceding sections, the prediction model was established based on samples collected in Beijing and was verified with data collected in the entire country. A natural question to ask is that whether or not the prediction model can be established using samples collected nationwide, and be validated in smaller regions, for instance, in Beijing. The following calculation confirms that this process is reversible.

Following the same process that was used to construct the model Equation (5) in Beijing, the 53 samples in nationwide were used to establish a national model. The result is expressed as follows,

$$E = -58.499 + 0.418 d_a + 0.311 d_p + 9.456 \ln(s_g) \quad (6)$$

in which the variables have the same meanings as those in Equation (5). The relationship between the actual error and that derived from Equation (6) is shown in Fig. 9.

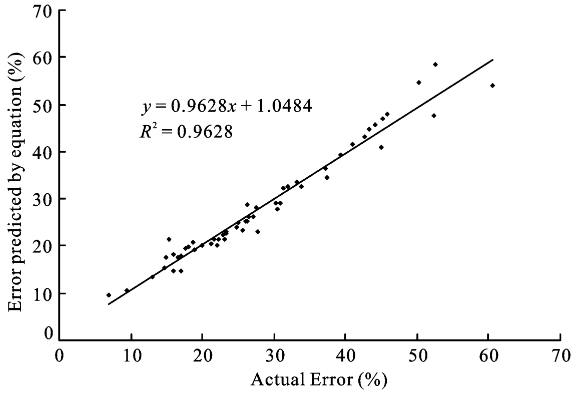


Fig. 9 Relationship between actual error and that derived from Equation (6) at nationwide scale

Then, the 77 samples collected in Beijing were used to validate the national model (Equation (6)). The relationship between the actual error and that derived from Equation (6) was shown in Fig. 10.

The correlation is statistically significant at the 0.001 level (R < 0.3799; number of samples is 77). It demonstrates that the error prediction model based on samples collected at large scales (nationwide) can be similarly applied to predicting errors at small scales (Beijing). Both approaches lead to the same conclusion that it is an effective and feasible way to predict rasterization errors with DA, DP and SG as input parameters.

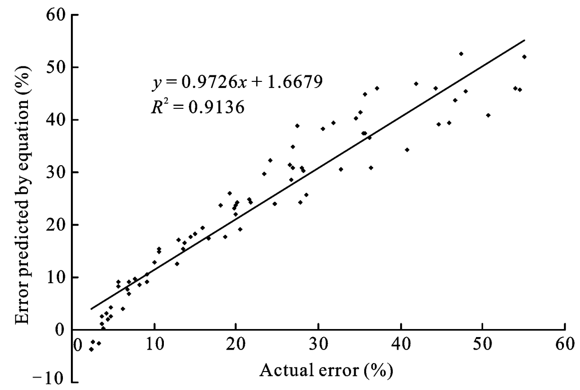


Fig. 10 Relationship between actual errors and those derived from Equation (6) in Beijing

4 Discussion

Vector-to-raster conversion is a process that is inevitably accompanied with errors. The errors have close relationship with the complexity of the region in consideration and the size of grid cells. Using models to predict rasterization errors is beneficial to make technical planning of rasterization projects and to producing rasterized products with high quality. In this study, land cover data of China at a scale of 1∶250 000 was taken as an example to construct errors prediction models of vector-to-raster conversion. The model built based on a small region (Beijing) could be verified in a large region (provinces and nationwide), and vice versa. It is demonstrated that the method used in this study is reasonable and feasible. However, the following two points need to be discussed and explained further.

(1) Coefficients of regression equations. Both the model expressed with Equation (5) and that with Equation (6) can be used for predicting rasterization errors, but they have different regression coefficients, especially for those of DP and DA. From the statistical point of view, for the same governing equation, different samples will inevitably lead to different regression coefficients. Therefore, one model is not necessarily better than the other. The main purpose of this study is to introduce the method of predicting rasterization errors rather than establishing the specific equations.

(2) Application of method. Although the dataset of land cover of China at the scale of 1∶250 000 was used

as an example to analyze the relationship between rasterization errors and their affecting factors, the method described in this study can also be applied to establishing other regression models with different thematic elements, for instance, soil or landform, and with different scales, for instance, at 1∶50 000 or 1∶10 000. The reason is that no matter what the thematic elements are and how much the scale is, it is the DA, DP and SG that directly affect rasterization errors. That is, it is by way of DA, DP and SG that different thematic elements and scales affect rasterization errors.

5 Conclusions

In this study, DA and DP were taken as indexes to measure the complexity of land cover in a given region. The influences of DA, DP and SG on rasterization errors were analyzed. The following findings can be made from this study.

(1) The rasterization errors are determined jointly by DA, DP and SG. The errors are linearly correlated with DA and DP, logarithmically with SG. The more complex (larger value of DA or DP) the region is, the larger the error becomes at a given resolution. The larger the size of a grid cell is, the larger the error also becomes when the complexity of region is fixed.

(2) The relationship between the errors and the DA, DP and SG can be represented by a multi-variant regression model. The model can be applied to predict the error before rasterization is conducted. The predicted error can be used as references for optimizing a rasterization scheme and for producing rasterization products with small errors.

(3) The two models established based on samples collected either in a small region (Beijing) or in the whole nation were both cross-validated. They all passed the test of statistical significance at the 0.001 level. It demonstrates that these models depict very well the relationship between rasterization errors and their affecting factors. They can be used to predict rasterization errors at regional scale and national scale as well.

(4) The method used in this paper can also be applied in predicting rasterization errors of other types of vector data.

References

Bates P D, De Roo A P J, 2000. A simple raster-based model for flood inundation simulation. *Journal of Hydrology*, 236(1–2): 54–77. doi: 10.1016/S0022-1694(00)00278-X

Bregt A K, Denneboom J, Gesink H J *et al.*, 1991. Determination of rastering error: A case study with the soil map of the Netherlands. *International Journal of Geographical Information Science*, 5(3): 361–367. doi: 10.1080/02693799108927861

Burrough P A, 1986. *Principles of Geographical Information Systems for Land Resources Assessment*. Oxford: Clarendon Press.

Carvers S J, Brunsdon C F, 1994. Vector to raster conversion error and feature complexity: An empirical study using simulated data. *International Journal of Geographical Information Systems*, 8(3): 261–270. doi: 10.1080/02693799408901999

Chen Jianjun, Zhou Chenghu, Cheng Weiming, 2007. Area error analysis of vector to raster conversion of areal feature in GIS. *Acta Geodaetica et Cartographica Sinica*, 36(3): 344–350. (in Chinese)

Chen Shupeng, 2002. Grid mapping and grid computing. *Science of Surveying and Mapping*, 27(4): 1–6. (in Chinese)

Crapper P F, 1980. Errors incurred in estimating an area of uniform land cover using Landsat. *Photogrammetric Engineering & Remote Sensing*, 46(10): 1295–1301.

Crapper P F, 1981. Geometric properties of regions with homogeneous biophysical characteristics. *Australian Geographical Studies*, 19(1): 117–124. doi: 10.1111/j.1467-8470.1981.tb00380.x

Demers M N, 2002. *GIS Modeling in Raster*. New York: John Wiley & Sons, Inc.

Frolov Y S, Maling D H, 1969. The accuracy of area measurement by point counting techniques. *Cartographic Journal*, 6(1): 21–35.

Goodchild M F, 1980. Fractals and the accuracy of geographical measures. *Mathematical Geology*, 12(2): 85–98. doi: 10.1007/BF01035241

Goodchild M F, Steyaert L T, Parks B O *et al.*, 1996. *GIS and Environmental Modeling*: *Progress and Research Issues*. New York: John Wiley and Sons.

Knaap Van Der, 1992. The vector to raster conversion: (Mis) use in geographical information systems. *International Journal of Geographical Information Systems*, 6(2): 159–170. doi: 10.1080/02693799208901902

Lawrence R L, Ripple W J, 1996. Determining patch perimeters in raster image processing and geographic information system. *International Journal of Remote Sensing*, 17(6): 1255–1259. doi: 10.1080/01431169608949084

Liao S B, Bai Y, 2010. A new grid-cell based method for error evaluation of vector-to-raster conversion. *Computational Geosciences*, 14(4): 539–549. doi: 10.1007/s10596-009-9169-3

Liu Mingliang, Tang Xianming, Liu Jiyuan *et al.*, 2001. Research on scaling effect based on 1 km grid cell data. *Journal of Remote Sensing*, 5(3): 183–190. (in Chinese)

Lunetta R S, Gongalton R G, Fenstermaker L K *et al.*, 1991. Remote sensing and geographic information data integration: Error sources and research issues. *Photogrammetric Engineering*

& Remote Sensing, 57(6): 677–687.

Maguire D A, Goodchild M F, Rhind D W, 1991. *Geographical information systems*, *Volume 1*: *Principle*. Essex: Longman Group UK Limited.

Peter B, Weibel R, 1999. Using vector and raster-based techniques in categorical map generalization. *Third ICA Workshop on Progress in Automated Map Generalization*. Ottawa, Canada, 1–14. Available at: http://aci.ign.fr/BDpubli/ottawa1999/papers/ peter_weibel.pdf.

Roth M, Voss G, Reiners D, 2004. Multi-threading and clustering for scene graph systems. *Computers & Graphics*, 28(1): 63–66.

Settgast V, Müller K, Fünfzig C *et al.*, 2004. Adaptive tesselation of subdivision surfaces. *Computers & Graphics*, 28(1): 73–78.

Shortridge A M, 2004. Geometric variability of raster cell class assignment. *International Journal of Geographical Information Science*, 18(6): 539–558. doi: 10.1080/13658810410001702012

Su B, Li Z, Lodwick G *et al.*, 1997. Algebraic models for the aggregation of area features based upon morphological operators. *International Journal of Geographic Information Science*, 11(3): 233–246. doi: 10.1080/136588197242374

Su B, Li Z, Lodwick G, 1998. Morphological models for the collapse of area features in digital map generalization. *Geoinformatica*, 2(4): 359–382. doi: 10.1023/A:1009757422454

Wang Xiaoli, Sun Qinghui, Jiang Chengshun, 2006. An optimization algorithm for transferring vector to raster data based on minimized area error. *Acta Geodaetica et Cartographica Sinica*, 35(3): 273–277. (in Chinese)

Wu Guangchen, Zuo Jianzhang, Liu Yan *et al.*, 2009. A high-effective algorithm for rasterization of vector data: Winding number algorithm. *Science of Surveying and Mapping*, 34(1): 50–52. (in Chinese)

Yang Cunjian, Zhang Zengxiang, 2001. Models of accuracy loss during rasterizing land use vector data with multiscale grid size. *Geographical Research*, 20(4): 416–422. (in Chinese)

Zhang Zengxiang, Wang Xiao, Wang Changyao *et al.*, 2009. National land cover mapping by remote sensing under the control of interpreted data. *Journal of Geo-information science*, 11(2): 216–224. (in Chinese)

Zhou Chenghu, Ou Yang, Yang Liao *et al.*, 2006. Equal-area-conversion from polygonal vector to raster. *Science in China* (*Series D Earth Sciences*), 36(supp. 2): 157–163. doi: 10.1007/s11430-007-5013-6