

Study on Method of Building a Regional Digital Elevation Model Based on Accuracy Evaluation by Cross Validation

Chen Keyun

Guangzhou Urban Planning & Design Survey Research Institute, Guangzhou, Guangdong, 510060 China

Keywords: digital elevation model; spatial interpolation; cross validation

Abstract: With the rapid social and economic development, the demand for application of surveying and mapping geographic information has been ever-increasing. Every year, a large amount of elevation point data will be added or updated, and it is in urgent need of building an authoritative, unified and high-precision digital elevation model to improve the guaranteed service capability of surveying and mapping geographic information and assist the urban planning and management departments in making decisions. This paper builds a digital elevation model for Nansha District through 41159 actually measured elevation points respectively using the global polynomial interpolation, local polynomial interpolation, inverse distance weighted interpolation and radial basis function interpolation. Cross validation shows that, the global polynomial interpolation is suitable for building a trend surface, and the local polynomial interpolation, inverse distance weighted interpolation and radial basis function interpolation are suitable for building a digital elevation model, where the radial basis function interpolation has the highest precision for building the regional digital elevation model for this research area.

1. Introduction

Digital elevation model (DEM for short) is the digital expression of surface configuration, and contains abundant topographic and geomorphologic information necessary for application analysis of geoscience^[1]. The digital elevation model is one of the most important national basic geographic information data, and its theory, method and application have always been a hot research topic of surveying and mapping and geographic information science at present. Digital elevation model and terrain factors derived thereof (such as: gradient, slope direction, etc.) are widely applied in geoanalysis, such as hydrology, geology, geomorphology, soil, meteorology, remote sensing, urban planning and 3D modeling. For example, it helps to analyze static hydrological characteristics and simulate dynamic hydrological process in terms of hydrology^[2]. It helps to assess landslide risks in terms of geological disaster prevention and control^[3]. In terms of geomorphology, it contributes to water system analysis, tectonic analysis^[4], geomorphologic shape and spatial pattern^[5], geomorphologic evolution and development^[6], etc. In terms of soil, it helps to study soil sampling strategy, soil erosion, soil properties, etc.^[2]. In terms of meteorology, it contributes to building a meteorological factor model^[7]. In terms of remote sensing application, it helps to correct remote sensing images and interpret remote sensing information. In terms of urban planning, it helps to analyze sunlight, intervisibility and the like. In terms of 3D modeling, it contributes to visualization and simulation of 3D terrain, etc.

In recent years, the rapid social and economic development and the rapid development of information technology have resulted in ever-increasing demand for application of surveying and mapping geographic information. Every year, a lot of elevation point data will be added or updated. Therefore, it is in urgent need of fully managing and making good use of the existing information resources, and building an authoritative, unified and high-precision digital elevation model, which shall be provided in accordance with the law as required to avoid redundant construction. This is of great significance for improving the guaranteed service capability of surveying and mapping geographic information, assisting the urban planning and management departments in making decisions, and meeting the needs for economic and social development.

2. Data sources and methods

2.1 Overview of the research area

This paper selects 41159 elevation points in the mid-east area of Guangzhou Nansha District, and explores the optimal method of building a digital elevation model for Nansha District by comparative analysis. Guangzhou Nansha District is located in the geometrical center of the Pearl River estuary and the great Pearl River Delta, is the junction of Xijiang River, Beijiang River and Dongjiang River, is a hub connecting urban agglomerations of the Pearl River estuary, and is the only sea route of Guangzhou City. This paper selects Tianshan in the mid-east area of Nansha District and surrounding areas thereof as the research area with a total area of 13.46 km². Nansha District is close to the estuary with gentle terrain. The research area includes small hills, farmland, residence areas, etc., and has certain representativeness.

2.2 Measurement and pretreatment

This research actually measures the control points using RTK, actually measures detail points using total station on this basis, and totally obtains 41159 elevation points, as shown in the figure below. Except for a small part of areas inconvenient for measure, other elevation points are uniformly distributed in the research area. Moreover, detail points are more densely covered and the measuring interval is shortened where the terrain changes relatively greatly (such as Tianshan), so as to ensure the accuracy of the digital elevation model.

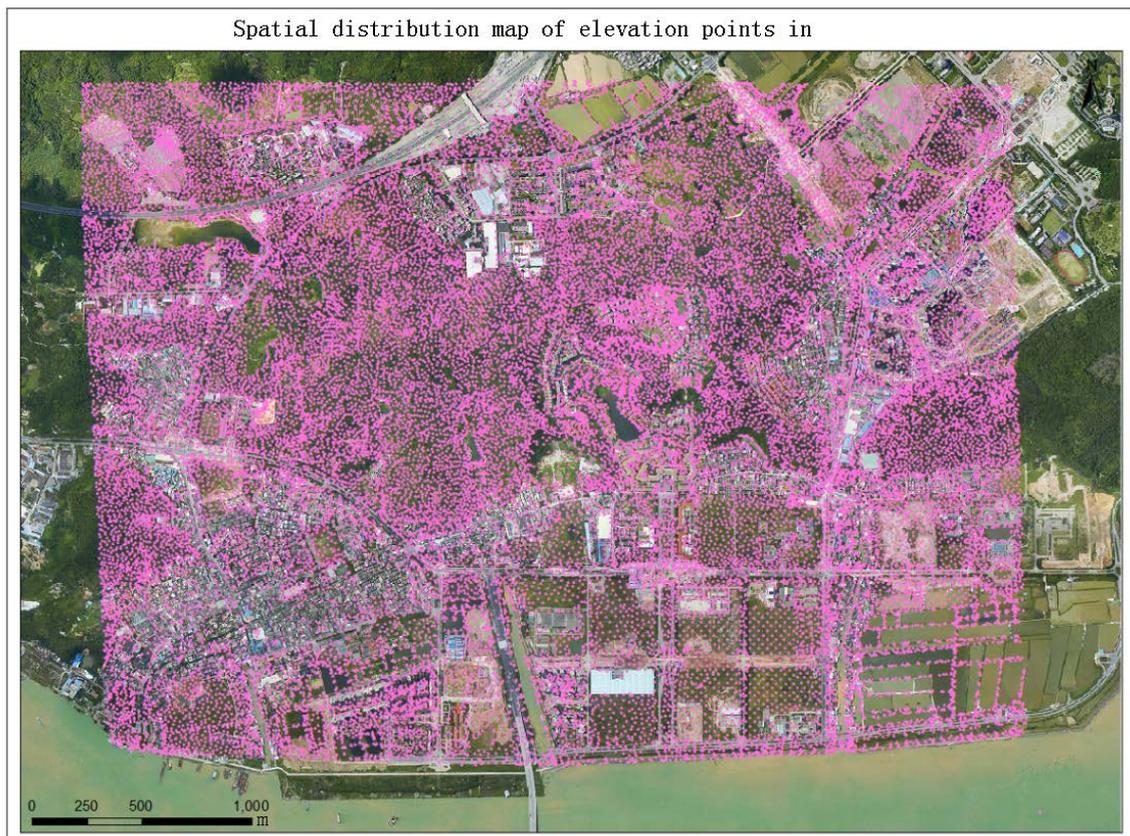


Fig. 1 Spatial Distribution Map of Elevation Points in Nansha Research Area

2.3 Interpolation method

Spatial interpolation method is used to analyze and predict the values associated with spatial or spatiotemporal phenomena. The commonly used spatial interpolation methods include: global polynomial interpolation (GPI), local polynomial interpolation (LPI), inverse distance weighted interpolation (IDW), radial basis function (RBF) interpolation, kriging interpolation, geostatistical simulation, etc.

The global polynomial interpolation fits out a smooth surface defined by a mathematical function

(polynomial) based on an inputted sampling point, and the global polynomial surface will gradually change and capture a coarse scale model in the data. Compared with the global polynomial interpolation that fits an entire surface using a polynomial, the local polynomial interpolation uses a plurality of polynomials in a specified overlapped neighborhood for fitting, and the neighborhood search can be defined using size, shape, number of adjacent factors and sector configuration. The inverse distance weighted interpolation assumes that each measuring point has local influence, which will decrease with the increase of distance, and the interpolation effect depends on the power value selection and the neighborhood search strategy. The radial basis function interpolation is the combination of a series of precise interpolation methods, and is a function that changes with the distance from a particular location, and its interpolation surface must pass through each of the measured sampling values. The kriging interpolation means to estimate the values of regionalized variables of unsampled points using the structural characteristics of known sample data and variation functions in the finite neighborhood of to-be-estimated points^[8]. Geostatistical simulation is the spatial extension method of Monte Carlo experiment, and maintains a certain spatial correlation. The experimental information has the same covariance function or variation function^[9].

ArcGIS10.2 is used herein as the software for building the digital elevation model.

2.4 Accuracy test method

This paper evaluates the accuracy of the digital elevation model by cross validation. Cross validation removes one data value each time, predicts the currently removed data value using remaining data values, then compares the predicted value with the actual value of the removed point position, and then repeats the process for the second point. The rest can be done in the same manner. Therefore, cross validation will compare the measured values and predicted values of all points.

Prediction accuracy evaluation indexes include mean error (ME) and root-mean-square error (RMSE), and the specific evaluation criteria are as follows:

$$ME = \frac{1}{n} \sum_{i=1}^n [z(x_i) - z^*(x_i)] \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [z(x_i) - z^*(x_i)]^2} \quad (2)$$

where $z(x_i)$ is the measured value, $z^*(x_i)$ is the predicted value, σ^2 is the predicted variance of the predicted value, and n is the number of test points. The closer the ME is to 0, the higher the prediction accuracy is; and the smaller the RMSE is, the higher the prediction accuracy is.

3. Analysis

3.1 Exploratory data analysis

According to the regulation, if the data follows a normal distribution, the mean is close to the median value, the skewness should be close to zero, and the kurtosis should be close to 3. As shown in the table below, there are 41159 elevation points in the research area with a minimum value of 0 and a maximum value of 180.81. The skewness is 1.54, is greater than 0, and belongs to a right skewed state, indicating that the data on the right side of the mean is more disperse. The kurtosis is 4.70, and is greater than 3, indicating that the distribution peak is steep. The median is 12.94, and is less than the mean 30.01, while 3/4 fractile is 43.28, and is far less than the maximum value. At the same time, according to the histogram, the data points are skewed to the left and concentrated in small numerical value segments. According to the normal QQ plot, the data points deviate from the 45-degree reference line, and don't follow a normal distribution. Therefore, in subsequent analysis, a spatial interpolation method that is not based on the normal distribution assumption of data should

be selected, such as: local polynomial interpolation, inverse distance weighted interpolation and radial basis function interpolation. In addition, normal transformation of elevation point data in this research area is attempted using SPSS. However, the results after transformation still can't pass the normality test. Therefore, this paper builds the elevation model using an interpolation method (such as: ordinary kriging interpolation, simple kriging interpolation and universal kriging interpolation) that does not request the data to follow a normal distribution.

Table 1 Descriptive Statistics of Elevation Points in Nansha District

Total number	Min. value	Max. value	Mean	Standard deviation	skewness	kurtosis	1/4 fractile	median	3/4 fractile
41159	0	180.81	30.01	31.83	1.54	4.70	7.86	12.94	43.28

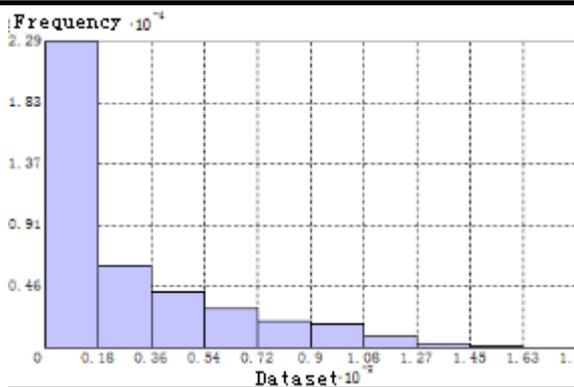


Fig. 2 Histogram

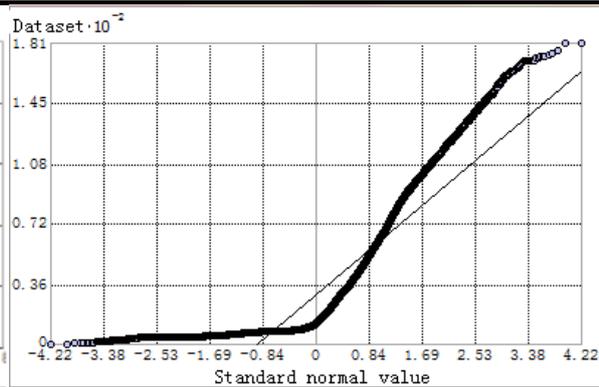


Fig. 3 Normal QQ plot

3.2 Spatial interpolation analysis

This paper builds a digital elevation model for Nansha District respectively using global polynomial interpolation, local polynomial interpolation, inverse distance weighted interpolation and radial basis function interpolation.

In Fig. 4 to Fig. 7, colors are applied to the digital elevation model using the same hypsometric tinting, where the elevation values in red areas are highest, successively followed by the elevation values in yellow areas and the elevation values in blue areas. According to the distribution of color bands, the prediction results of local polynomial interpolation, inverse distance weighted interpolation and radial basis function interpolation are similar, where the line of Central Tianshan-Lushan- Feie Mountain- Nansha Tianhou Palace has high terrain, and presents red or orange; and the southeast and southern areas are farmlands and residence areas, present yellow or blue, and comply with the actual situation. The global polynomial interpolation obtains a smooth surface that can show an area of interest with a gradual surface trend, mainly shows the terrain change situation in the research area, neglects a lot of details, and is not suitable for building the digital elevation model.

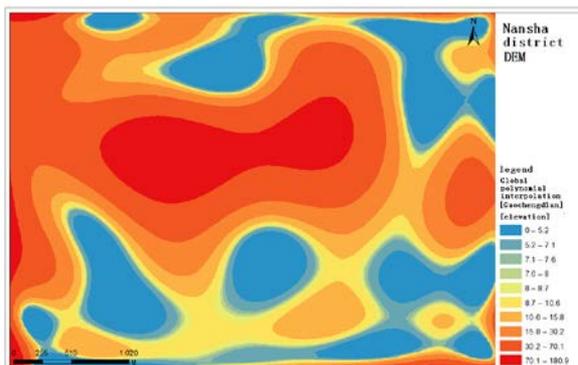


Fig. 4 Global polynomial interpolation

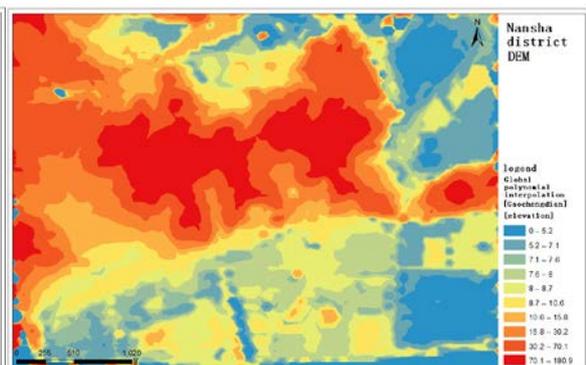


Fig. 5 Local polynomial interpolation

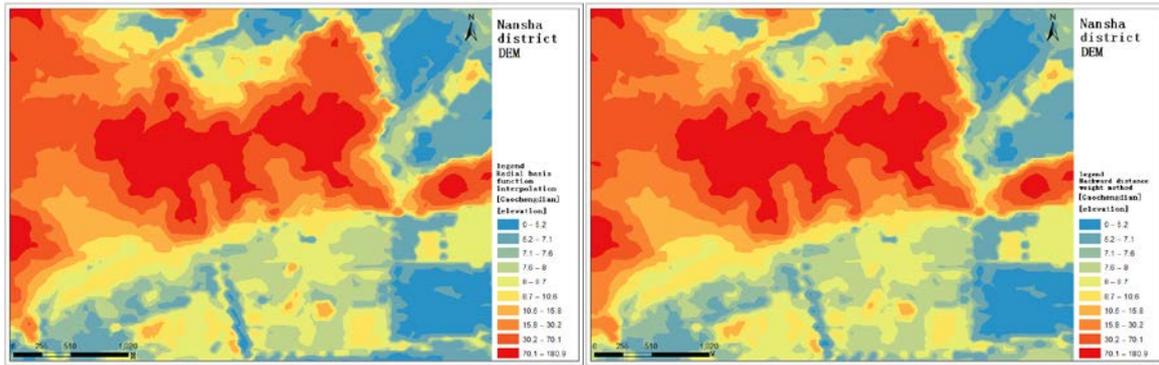


Fig. 6 Inverse distance weighted interpolation Fig. 7 Radial basis function interpolation

The optimal model should have a mean error closest to 0, and the minimum root mean square error. As can be seen from the table below, the mean error of the four interpolation methods are close to zero; and the root mean square error of the interpolation methods except the global polynomial interpolation is less than 2, showing that the results of the three interpolation methods are better. In addition, the radial basis function interpolation produces the mean error closest to 0, and the minimum root mean square error in this research. Therefore, the radial basis function interpolation is most suitable for building the digital elevation model in this research area.

Table 2 Building Accuracy of Four Interpolation Methods

Method	Mean error (ME)	Root mean square error (RMSE)
Global polynomial interpolation	0.000550	15.251545
Local polynomial interpolation	0.00710	1.58246
Inverse distance weighted interpolation	-0.01678	1.92836
Radial basis function interpolation	-0.00123	1.52232

4. Conclusion

(1) Before building a digital elevation model, it is necessary to verify whether the elevation data follow a normal distribution through exploratory data analysis, so as to select a proper interpolation method.

(2) The global polynomial interpolation is suitable for the trend surface building, and the local polynomial interpolation, inverse distance weighted interpolation and radial basis function interpolation are suitable for building a digital elevation model, where for the digital elevation model for this research area, the radial basis function interpolation has the highest building accuracy.

References

- [1] Tang Guoan, Li Zhiyuan and Liu Xuejun. *Teaching Course of Digital Elevation Model*. Beijing: Science Press, 2010.
- [2] Tang Guoan. Analysis and Research Progress of Digital Elevation Model and Digital Terrain in China [J]. *Acta Geographica Sinica*, 2014, 69(9): 1305-1325.
- [3] Zhu Axing, Pei Tao, Qiao Jianping, et al. A Landslide Susceptibility Mapping Approach Using Expert Knowledge and Fuzzy Logic Under GIS [J]. *Progress in Geography*, 2006, 25(4): 1-12.
- [4] Ji Yapeng, Gao Hongshan, Pan Baotian, et al. Implication of Active Structure in the Upper Reaches of Weihe River Basin from Stream Length-Gradient Index (SL index) and Hack Profile [J]. *Journal of Lanzhou University (Natural Sciences)*, 2011, (4): 1-6.
- [5] Li Zhiyuan, Tang Guoan, Jia YiNi, et al. Slope Spectrum Information Entropy Scale Effect and Spatial Differentiation [J]. *Geo-Information Science*, 2007, 9 (4): 13-18.

- [6] Xiong Liyang, Tang Guoan, Yuan Baoyin, et al. DEM based Research on the Inheritance of Geomorphic Evolution in the Loess Plateau (Main Soil Erosion Area) [J]. *Science China: Earth Sciences*, 2014, 57 (2): 313-321.
- [7] Pan Yaozhong, Gong Daoyi, Deng Lei, et al. Smart Distance Searching-based and DEM-informed Interpolation of Surface Air Temperature in China [J]. *Acta Geographica Sinica*, 2004, 59(3): 366-374.
- [8] Wang Zhengquan. *Geostatistics and its Application in Ecology* [M]. Beijing: Science Press, 1999.
- [9] Chen Yaxin, Shi Haibin and Wei Zhenmin. *Prediction Theory and Conditional Simulation of Spatial Variability of Soil Water and Salt Information* [M]. Beijing: Science Press, 2005.