

# The Practice of Methods of Detecting On The Quality of Geographic Spatial Data

Zhang Jin-ming Wang Guang-xia Liu Fang  
(The Institute of Surveying and Mapping Zhengzhou 450052)

**Abstract** In this paper, we put forward five aspects of geography spatial data quality detecting according to a series of problems existing in the data of Digital Map Database. These five aspects are format detecting, topological relation detecting, position detecting and logic consistency detecting. We also put forward relevant detect method and arithmetic according to different detect aspects.

**Key Words** Geography spatial data, Data format detect templet, Topological relation, Position relation, Logic consistency, Digital Elevation Model.

## *Introduction*

All kinds of geographic data and attribute data which are stored and proceeded in the Digital Map Database can be indicated by graph. These two kinds of data are called geographic spatial data including spatial data and attribute data. With the development of theory of GIS, the reliability of conclusion about GIS analysis and the fundamental questions about the quality of the kinds of visual information products have been give more and more limelight. America is one of the countries which first put forward the standard of the geographic spatial data. *Federal Information Process Standard* lists five projects which must been evaluated the quality of data: lineage, position accuracy, attribute accuracy, completeness and logical consistency. Subsequently, Spatial Data Quality Committee of ICA add two projects: semantic accuracy and temporal information and detail these projects<sup>[1]</sup>.

In accordance with the rules of ICA and problems existing in the data of Digital Map Database, the authors deem the detecting on the quality of the geographic spatial data mainly manifest the following aspects:

- **Format Detecting** of geographic spatial data
- **Attribute Detecting** of geographic spatial data
- **Topological Relation Detecting** of geographic spatial data
- **Position Relation Detecting** of geographic spatial data
- **Logic Consistency Detecting** of geographic spatial data

## **1. Format Detecting of geographic spatial data**

Format detecting mainly detects whether the collecting data are complete and the format of files are right etc.

### *1.1 The Manage Pattern of Digital Map Database*

Digital Map Database manages in way of maps. Each map includes data files of 11 elements layers. They are survey control files (ssx , szb) ,single feature files (isx,izb) , residence files(rsx,rzb), traffic files(tsx,tzb), supply line and wall files(gsx,gzb), administrative boundary files(tsx,tzb), hydrographic net files(hsx,hzb), geomorphology files(csx,czb), control files(dsx,dzb), soil texture files(lsx,lzb), vegetation files(asx,azb), and one toponym file(ndm),one text file(wmd) and one public file(ogy). For one integral map (uninclude a map not having certain element). the above-mentioned files are absolutely necessary. Attribute files record the elements' attribute information; coordinate files record the elements' position information; toponym files record the position and attribute information of toponym etc; text files record the map margin information such as directional lines, table of adjacent maps etc; and public files record the map's name, number, date of finished, the production unit of original map and the information of data accuracy etc.

#### ***1.1.1 Format of Attribute Data***

Attribute data files record attribute data and are composed of first record point record, line record and area record, which record attribute information and topological relation information of spatial data. The detailed data format is in some reference book.

#### ***1.1.2 Format of Coordinate Data***

Coordinate files record the coordinate data, including one first record and lots of mid records which record the position information of spatial data. The detail data format is in reference book, Coordinate data and attribute data make up of a whole geographic spatial data.

#### ***1.1.3 Format of Toponym Data***

Toponym file records the toponym data, which includes first record and lots of mid records. The substance of toponym can be related by toponym pointer. the detailed data format is in some reference book.

### ***1.2 Format Detecting***

Format detecting includes integral detecting of files and format detecting of files.

#### ***1.2.1 Integral Detecting***

Integral detecting is whether the files are integral. According to the manage rules for military Digital Map Database we can detect the amount of the files. All program platform provide the function which can search whether the files exist. For example, in Visual C++, we use API:

```
...::SearchPath(lpszPath,lpszFile,lpszExtension,cchReturnBuffer,lpszReturnBuffer,plpszFilePart)[2]
```

to complete the detecting of integral.

Attribute data, coordinate data and toponym data in Digital Map Database all have certain format. Based on the format, the authors provide a new concept of “**Data Format Detect Templet**” to complete the format detecting of files.

Firstly make a “**Data Format Detect Templet**”. Secondly pick out a record form file and compare it with data format detect templet. And decide whether the format of file is right, such as if unnecessary spaces or record omission are existed etc. Finally

point out the error information in certain format.

Give an example of point record, its format as follows:

Number	Main Code	...	Direction Coordinate	...	Description Code Number
Description Code	...	...	...	...	Toponym Point

Every parameter are separated by space. Then we can make data format detect templete:

%6ld	[]	%6ld	[]	%6ld	[]	%6ld	[]	%6ld	[]	%6ld	[]	...
------	----	------	----	------	----	------	----	------	----	------	----	-----

According to “**Data Format Detect Templet**”, we can check the format of data files and find the error information and show it in table.

Another advantage of “**Data Format Detect Templet**” is that data format detect templete is changeful, that is to say, we can make different templet in view of different data format. Thus flexibility of data format is enhanced.

## ***2. Attribute Detecting of Geographic Spatial Data***

In standard format rules of Digital Map Database, the attribute and code of map element are closely connected with each other. The rules stipulate the attribute codes are made up of main code and sub-main code. Sub-main code may divide into identify code, description code and parameter code. In order to describe the logic record of element, we must use one main code, one identify code, some description codes and some parameter codes. The rules detailedly stipulate main code, spot code, description codes and parameter code such as the range they are founded and some special rules. The stipulations above consist the basis of theory for attribute detecting.

Attribute detecting of geographic spatial data includes two aspects: range detecting and symbolizing detecting.

### ***2.1 Range Detecting***

The theory foundation in checking rightness of map elements’ attribute is the series of range which the main code, identify code, description code and parameter code are founded in rules. We particularly point out that for parameter code detecting, the program can’t distinguish in which situation the parameter code is right. It is necessary to arrange the collection of parameter code. For example, bridge has three parameter: the first is length parameter, the second is width parameter, the third is carrying capability parameter.

### ***2.2 Symbolizing Detecting***

“Special effect of the visual technical use in quality detect of spatial data” which academician Gao Jun advanced in book *The visualizing of geographic spatial data* will be paid more attention to.

If we symbolizing in attribute detecting of map elements, there must be symbolic library to be supported. For those new elements we can renew at any time.

## ***3. Topological Relation Detecting of Geographic Spatial Data***

Vector data, as the expression of GIS, receives much attention all the while in GIS.

It results from that vector data has different character compare with grid data. The vector data with topological information is the data resource which are used in net analysis and spatial analysis. It is far-reaching to perform topology detecting to the data that has topology relation. The algorithm of automatic topology relation building has become more and more mature in recent years. The left-turning algorithm's(right-turning algorithm) being put forward has made it possible to search polygons automatically. Cheng Chun, Zhang Shu-wen, Xu Gui-fen make a further elaboration on the maths basis of topology relation in “*The Basis for Generation of Topologic Information Of Polygons in GIS*”<sup>[3]</sup>; Bai Ling prompts the nine relation of automatic topology relation building in “*The Topologic Relationship of Spatial Data and its Auto-establishment*”<sup>[4]</sup>. Yang Hai-hong, Li Tian-hong, Xue An's *Double Neighboring Point Judgement Algorithm*<sup>[5]</sup>, Qi Hua, Liu Wen-xi's *QI Algorithm*<sup>[6]</sup>, Zhang Jin-ming's *Partition Algorithm*<sup>[7]</sup>. The above algorithms improve the efficiency of automatic topology relation building and prompt it. But there're scarcely papers about topology relation detecting.

The authors believe during automatic building the topological relation, we can set certain condition and give some error information at any time to complete the detecting of topological relation and build the right topological relation.

The authors think the topological relation errors mainly display as follows:

**Exist suspense line:** node matching forms “table of node and arc-line” to save topological node information during the automatic constructing topological relation. If a some node only has one relative arc-line, the arc-line is the suspense line choice of matching limit result in the arc-line should have be matched.

**Exist bridge:** For a certain bridge, it is called bridge when its left-area number equal the right-area number. The existing of bridge make the errors of topological relation in logical **that is one polygon can't be connected by one line.**

**Lack on surplus of point inward:** collecting more on less point inward in one polygon may cause the topological relation error.

**The wrong number of left or right area:** It is most important in topological relation to make polygon topological relation. And polygon topological form the relation form the relation of arc-line and point inward. If there are wrong numbers of left or right area in an arc-line, the automatically searching polygon will be fantastic.

#### **4. Position Detecting of Geographic Spatial Data**

Map is the main carrier of spatial data and visual. Its image error has strict limit. And it can be checked by seeing directly. The errors of surveying are under control of every working procedure<sup>[1]</sup>. With the appearance of digital map, there must be man-made errors, machine errors and technical errors. When the paper map covered into digital map, three results to position errors of geographic spatial data. So, we must detected the position of geographic spatial data.

The traditional way of detecting position is mapping the digital map one paper and

compared it with the original map. The search the errors and lacks. Although its final effect is pretty well, it wastes lots of labors and material resources.

In accordance with those question and combined with technique of proceeding digital image, we can superimpose the vector map on the grid map to check. The authors advance the “*vector-grid integration*”:

- If the bitmap is colorful, it can be transformed into black and white bitmap. If the bitmap is black and white, it is unnecessary to do like this;
- Match the bitmap map and vector map making use of method of minimal power.
- Using Algorithm Bresenham, convert the line elements in vector map into grid and expend it so as to widen the grid line.

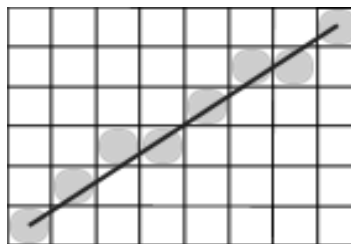


Figure 1 The position of pels using bresenham arithmetic

- Write the grid data received from 3 into the bitmap and proceed XOR. If there is one spot whose pels value is 1 on the bitmap. When the grid data has the same pels value at the same spot, this spot one pels map will be evaluated 1 or retain it .



Figure 2 the sample of “XOR”

- Full scan: if the pels value of all the eight elements around one element is 1, record the position, it indicates that there’s error, we display it in high brightness.

## 5. Logic Consistency Detecting of Geographic Spatial Data.

Logic consistency of geographic spatial data considers all kinds of logic relation of elements themselves or one another elements. For example, there are altitude relations among contours , altitude relations between toponym pointer and toponym entity in the Map Database.

The authors expatiate logic consistency by three examples: detecting the altitude of contours, detecting the contraction between the altitude points and contours and detecting toponym pointer.

### 5.1 Detecting the Altitude of Contours

Contours are principal elements and they are resources which are used to build up DEM. The rightness of contours will directly affect the reliability of DEM.

Detecting the altitude of contours includes color distinguish check and model check. Each method has advantages in general we use them together and they supplement each other.

### **5.1.1 Color Distinguish**

Color distinguish is that calculate every color value corresponding the altitude of contours according to the isometry and the rules of colors to decide the rightness of altitude. The methods of color distinguish consist core color every contours, one color every five contours and one color every delimitation of layers.

One color every contours, make one contours a certain color and the adjacent five contours as one group. Repeat above five colors with the rise of elevation. This method has good effect for those mistakes whose altitude error is within five contours.

One color every five contours, make five contours a certain color and the adjacent 25 contours as one group. This method is similar with method 1. But it widen the color from five contours to 25 contours. One color every delimitation of layers we prefer to here is not whole meaning of it. We interpolate colors in view of elevation for every contours we set different color. This method will control the mistakes of elevations in whole, such as whether the contours are given the elevation or not.

### **5.1.2 Model Check**

DTM is the order value array while describes the spatial distribution of ground characters. In general, the ground character is elevation  $Z$  whose spatial distribution may be described not only by  $x, y$  rectangular plane coordinate, but also by longitude  $\lambda$  and latitude  $\phi$ . In latest books, in order to distinguish the DTM describing other ground characters, the DTM of elevation or altitude distribution is called DEM.

In general, contours are plane figure. Convert the plane contours into the visual DEM and use the visual method to aim at checking. We usually use the add-right arithmetic to set up DEM. Show the model with the visual way (such as OpenGL) and the user can clearly learn where the mistake exists.

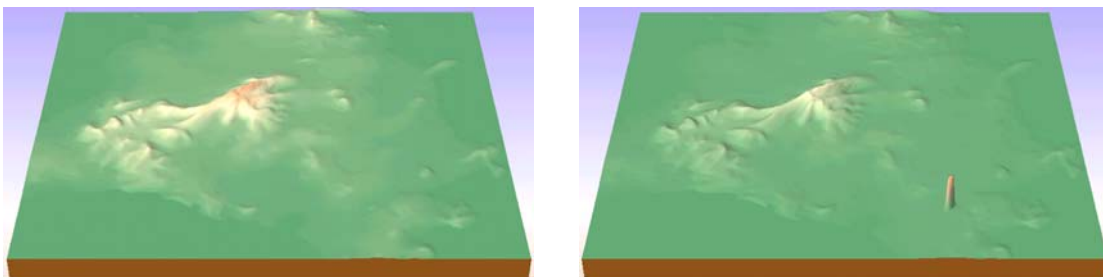


Figure 3 Mode Visual of contours Detecting

(The left is the right DEM, the right is the wrong DEM)

## **5.2 Detecting the Contradiction between the Altitude Points and Contours**

The contradiction between the altitude points and contours is inconsistency of altitude point and elevation. For the contours which surround the altitude point, the elevation difference between them must be in one contours, otherwise the logical mistake will exist between them. During the building DEM, the altitude will be calculated as an important parameter. Thus the mistake must be found.

The authors use the “*Get Point Of Intersection In Four Or Eight Directions*” to decide the relation of them. Give the example of “*Get Point Of Intersection In Four Directions*”, the steps as follows:

- Take the altitude point’s x coordinate as benchmark and intersect it with contours. Then record the elevation and x, y coordinate of the intersections.
- Order the elevation and pick out the first and the last point with least distance with altitude point and make them condition 1.
- Take the altitude point’s y coordinate as benchmark and intersect it with contours then record the elevation and x,y coordinate of the intersections.
- Order the elevations and pick out the first and the last point with least distance with altitude point and make them condition 2.
- Judge condition 1 and 2, if the elevation difference of three in two conditions is all less than one contour this altitude point is reasonable. Otherwise, record the error information.

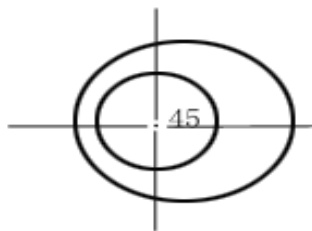


Figure 4 Get the Point of Intersection on Four directions

We may use more benchmark of intersecting such as eight directions and sixteen directions to judge the contradiction between the altitude point and contours. The more benchmark used the more accurate result is.

### ***5.3 Detecting Toponym Pointer***

There exists a toponym file to mark the note on map besides attribute files and coordinate files in the Military Digital Map Database. Database use the toponym pointer connect to relate the substance of toponym with toponym and reduce the surplus data and enhance the structure of data. But there may be some logical mistakes. Relate the A’s toponym pointer with substance B for example. We may connect the toponym with substance and make it visualize to aim at checking because the relations between substance of toponym and toponym is inside and not easily to be directed.

It arithmetic as follows:

- Searching the substance corresponding the toponym pointer.

- Connect the substance and the toponym with line. The lines usually exist around the toponym and show radial.
- Decide if there are errors in view of the line.

## 6. Conclusion

The authors practice according to the arithmetic advance aim at the detection of five aspects and get good effect. But there are still many aspects to be enhanced for many reasons. For instance, the rector-grid integration exists ambiguity because the bitmap is made of scatter data. With the visual technical applied in mapping and surveying, it become the main force of spatial data detecting, such as logic consistency detecting this paper mentioned.

## Reference

- 1 . Gao Jun. *Visualization in Geo-Spatial Data*. Engineering of Surveying and Mapping. 2000. 8
- 2 . Microsoft Corporation. *Microsoft Win32 Programmer's Reference*. Tsinghua University Press. 1995. 5
- 3 . Cheng Chun, Zhang Shu-wen, Xu Gui-fen. *The Basis for Generation of Topologic Information Of Polygons in GIS*. ACTA GEODAETICA et CARTOGRAPHICA SINICA. 1996(11)
- 4 . Bai Ling. *The Topologic Relationship of Spatial Data and its Auto-establishment*. Journal of Institute of Surveying and Mapping. 1993(4)
- 5 . Yang Hai-hong, Li Tian-hong, Xue An. *Optimization on Automatic Topological Relation Generating Algorithm with Double Neighboring Point Judgement Method*. Journal of Image and Graphics. 1998(7)
- 6 . Qi Hua, Liu Wen-xi. *The Method of QI on Automatic Topological Relation Generating Algorithm*. ACTA GEODAETICA et CARTOGRAPHICA SINICA. 1996(8)
- 7 . Zhang Jin-ming. *Optimization on Automatic Topological Relation Generating Algorithm with Data Partition Method*. Journal of Institute of Surveying and Mapping. 2000. 6