AUTOMATIC CORRECTION ALGORITHM OF HYFROLOGY FEATURE ATTRIBUTE IN NATIONAL GEOGRAPHIC CENSUS

Chengming Li^a, PeipeiGuo^a, Xiaoli Liu^{a, *}

^a Chinese Academy of Surveying and Mapping, Haidian District, Beijing, China- (cmli, guopp, Liuxl)@casm.ac.cn

Commission VI, WG VI/4

KEY WORDS: National geographic census, Attribute correction, River system structuration, River selection

ABSTRACT:

A subset of the attributes of hydrologic features data in national geographic census are not clear, the current solution to this problem was through manual filling which is inefficient and liable to mistakes. So this paper proposes an automatic correction algorithm of hydrologic features attribute. Based on the analysis of the structure characteristics and topological relation, we put forward three basic principles of correction which include network proximity, structure robustness and topology ductility. Based on the WJ-III map workstation, we realize the automatic correction of hydrologic features. Finally, practical data is used to validate the method. The results show that our method is highly reasonable and efficient.

1. INTRODUCTION

The general survey of national geographic conditions in China has generated a great deal of data about landforms, land covers, and other important geographic features (Li et al. 2014, Chen 2012, He et al. 2013), which lay a foundation for the routine national geographic conditions monitoring. As one of the 12 types of level 1 surface geographic features, hydrology is among the main geographic feature type included in the general survey. The comprehensive and very up-to-date hydrologic data contain information about type, location, scope, area and other aspects of surface waters (Gao et al.2016, Chen et al. 2015, Wang et al. 2015). However, due to the complexity of data acquisition and the demand for hierarchical data management, the integrated acquisition of hydrologic data, which involves data collection in both indoor and field environments, can lead to undefined attributes of some linear rivers. This could impede further application of survey data in relevant areas such as hydrology and water industry, river management, municipal pipe network, and environmental protection (Li et al. 2014).

Currently, this problem is mostly addressed by manual correction (Xu et al. 2015, Song et al. 2014, Chen et al. 2012), but this method is very time-consuming and effortful. For this reason, an algorithm for automatically correcting attributes of hydrologic features in the general survey data was proposed in this study. After a systematic analysis of the characteristics of hydrologic data, hydrologic features with undetermined attributes were classified based on their locations and topographical relationships, and attribute correction principles were proposed for different types of hydrologic features. Next, an automatic attribute correction method was designed on the basis of the principles. The corrected data was then compared with real data and the results confirm the rationality and validity of the correction method proposed.

2. CHARACTERISTICS OF HYDROLOGIC DATA

Geographic features make up the bulk of the content of the general survey of national geographic conditions. According to the Content and Indicators for the General Survey of National Geographic Conditions (Liu 2013), the geographic features include 5 types of level 1 features, 16 types of level 2 features, and 47 types of level 3 features. Hydrology is one of the main feature types involved in the general survey. Hydrologic data collected during the general survey are stored in two layers: HYDA and HYDL. HYDA stores data of areal water bodies with high water levels, including lakes, reservoirs, double-line streams, marshes, main canals, etc. HYDL stores data about center lines and structural lines of these areal waters. For watercourses measuring wider than 20 m, data was collected across the polygons covered by these river courses and along their structural lines, and relevant attributes were assigned to the data collected. For watercourses narrower than 20 m, data was gathered along their center lines and relevant attributes were assigned (Wu et al. 2012). In urban regions, for watercourses wider than 5 m, data collection was carried out within corresponding polygons and along their structural lines, followed by attribute assignment; for those narrower than 5 m, data collection took place along their center lines and attributes were given to the data collected (Xu et al.2015).

Feature type	Data content	Data hierarchy	Geometry of data cube
Hydrology (H)	Hydrologic features	HYDA	Area
		HYDL	Line
	Hydrologic facilities	HFCA	Area
		HFCL	Line
		HFCP	Point

Table 1. Content of hydrologic data

* Corresponding author

In the hydrologic data obtained from the general survey, hydrologic features include single-line streams, double-line streams representing rivers, canals, ditches, seasonal streams and dry rivers, and areal water bodies through which single-ling or double-line streams pass, such as lakes, reservoirs, ponds, seasonal lakes, and dry lakes. A river network is made up of interconnecting structural lines (collected along double lines of streams or center lines of polygons) and single-line streams. While single-line streams have specific GB codes indicating their attributes, structural lines were designated "9999999" instead of a clear identification if there were data available about corresponding polygonal water bodies. Moreover, structural lines were added to connect sing-line tributaries flowing into double-line rivers (represented by center lines) to the structural lines of the corresponding double-line streams. Their extensions are regarded as structural lines of sing-line tributaries and also receive the GB code "999999".

As mentioned above, it is impossible to identify structural lines within polygons when organizing data into levels based on GB codes. Take perennial rivers for example. Only perennial river lines with GB code 230101002 are identifiable, while structural lines within polygons corresponding to areal water bodies are unidentifiable. As a result, the data of perennial rivers extracted could be incomplete, thus affecting the results of subsequent river system structuration, selection and generalization of rivers, and gradient effects. It is therefore necessary to assign attributes to structural lines of rivers whose attribute codes are undefined. At present, this work is done primarily by manual identification and processing, which is not only effortful and time-consuming, but is also unable to ensure data quality. This necessitates an algorithm that enables automatic correction of river attributes.

3. PRINCIPLES FOR AUTOMATIC ATTRIBUTE CORRECTION FOR LINEAR HYDROLOGIC FEATURES AND CORRECTION METHOD

3.1 Classification of structural lines of hydrologic features

Based on their locations and topological relationships, structural lines of hydrologic features undetermined GB attributes in HYDL were classified into the following four types:

(1) Inflow line: An inflow line represents a river that empties into a polygon along line's direction. It links to rivers' center lines outside this polygon (out-of-polygon center lines) or outflow lines coming from other polygons. Lines L2 and L6 in Figure 1 are examples of inflow lines.

(2) Outflow line: An outflow line indicates a river that flows out of a polygon along its direction. It connects to out-of-polygon center lines or inflow lines towards other polygons. Lines L4 and L7 are examples of outflow lines.

(3) Intermediate line: An intermediate line is entirely within a polygon, and its two ends join structural lines associated with this polygon. Line L3 in Figure 1 is an intermediate line.

(4) Independent line: An independent line is associated with one polygon that connects to only one structural line. In Figure 1, segments L10, L11 and L12 constitute an independent structural line.



Figure 1. Classification of structural lines of hydrologic features

3.2 Correction principles

As can be seen from the spatial characteristics of hydrologic features and topological relationships between them, structural lines with undetermined attributes typically connect to structural lines or polygons that have defined attributes. This allows us to assign attributes to structural lines with unclear attributes using the defined attributes of associated structural lines or polygons based on their spatial relationships (Ai et al. 2007). However, for a main river that runs through multiple areal features like reservoir and ponds, attribute assignment by simple spatial overlay of associated attributes will inevitably lead to varying attributes for this river. This could cause some new problems, such as discontinuity in river lines and poor connectivity between rivers. Therefore, attribute assignment should not only consider neighborhood correlation, but also follow some basic principles and requirements like the principles of structural robustness and topological extensibility (Ai et al. 2007, Li et al. 2008, Yu et al. 2014).

(1) The principle of structural robustness. During attribute assignment, it is necessary to keep river lines as smooth and complete as possible in order to ensure clearly identifiable river networks, and keep the prominent major center lines uninterrupted. Attribute assignment by manual operations depends on visual, qualitative analysis and tends to retain main river lines. In comparison, computer-based automatic processing relies primarily on the quantifiable structural and topological characteristics of river networks.

(2) The principle of topological extensibility. This principle requires rivers to be as topologically extensible as possible. During attribute assignment, adding some short lines, if necessary, can improve connectivity between rivers and thus their overall smoothness and extensibility. For example, a structural line can be added as a link between a single-line tributary (center line) and the structural line of the double-line stream fed by it. The spatial relationships between lines should certainly be clear and reasonable.

3.3 Correction method

As topological relationships between structural lines are related to their spatial distribution, polygons were selected as the basic correction units and for each correction unit, all structural lines associated with it were corrected. Besides, given the vast quantities of the hydrological data from the general survey, the spatial relationship of a structural line to an areal water body was determined according to the spatial relationship of the structural line's midpoint to that water body, in order to improve the efficiency of calculation. The flowchart in Figure 2 illustrates the attribute correction process implemented in this study.



Figure 2. Flowchart showing the attribute correction process

Topological pre-processing and establishment of 3.3.1 topological relationships: Establishing topological relationships requires topological pre-processing of the raw data, including correcting self-intersection of lines, node fitting, removing repeated lines, dummy nodes and undershoots, creating topological polygons, etc. Based on the topological characteristics revealed in Figure 1, a topological tree displaying the topological relationships between the nodes (Figure 3) and a topological structure showing the topological information of the points, lines, and polygons (Table 1) were constructed. The topological tree and topological information structure clearly reveal the topological relationships and connectivity between nodes and lines (Chen et al. 2006).

3.3.2 Structural line processing: In order to ensure the river network's integrity, the structural lines were divided into network-forming and non-network-forming structural lines, depending on whether the polygons that contain the intermediate lines associated with them are part of the river network. Network-forming structural lines are associated with polygons corresponding to water bodies like rivers and canals. Structural lines that are associated with polygons covered by waters like reservoirs and ponds fall into the other category. Next, the linear features with undetermined attributes were processed in accordance with the principles of neighbourhood correlation, structural robustness, and topological extensibility mentioned above. The correction method is detailed below:

(1) Inflow lines: Network-forming inflow lines: if inflow lines are associated with out-of-polygon center lines, the attributes of these center lines were assigned to the inflow lines; if they connect to within-polygon structural lines, the relevant polygons' attributes were assigned to the inflow lines.

Non-network-forming inflow lines: if a polygon has only one inflow line, the attributes of out-of-polygon center lines or structural lines were assigned to this inflow line and labeled as standard corrected attributes. If a polygon has multiple inflow lines, proper corrected attributes were determined based on the levels and topological extensibility of the out-of-polygon structural lines associated with these inflow lines and then marked as standard corrected attributes. (2) Intermediate lines: Network-forming intermediate lines: Network-forming intermediate lines were given the attributes of the polygons where they lie within.

Non-network-forming intermediate lines: these intermediate lines were given the standard corrected attributes determined for inflow lines.

(3) Outflow lines: Network-forming outflow lines: Network-forming outflow lines were assigned the attributes of corresponding polygons.

Non-network-forming outflow lines: Network-forming outflow lines were assigned the attributes of corresponding polygons. If a polygon has only one outflow line, this outflow line was given the standard corrected attributes determined for inflow lines. If a polygon connects to multiple outflow lines, a main outflow line was determined based on the levels and topological extensibility of the out-of-polygon structural lines associated with inflow lines and then the corresponding standard corrected attributes were assigned to the main outflow line. The other outflow lines received the attributes of out-of-polygon center lines or structural lines associated with them.

(4) Independent lines: Network-forming independent lines: Network-forming independent lines refer to intermediate lines that do not join any inflow or outflow line; they were assigned the attributes of corresponding polygons.

Non-network-forming independent lines: these independent lines were given the attributes of structural or center lines that connect to their starting points. If their starting points are not associated with any structural or center line, the attributes of structural lines associated with their ending points were assigned to these independent lines.

Figure 3 shows the results of attribute assignment to the linear hydrologic features in Figure 2.



Figure 3. Schematic of results of attribute assignment

4. EXPERIMENTAL ANALYSIS AND EVALUATION

With the support of the development interface provided by the NewMap WJ-III map workstation, the module for automatically correcting river attributes was generated. This study analyzed the hydrologic data from a city of Guangdong Province, including data of perennial rivers, dry rivers, overflowing main rivers, main canals, branch canals, reservoirs, and ponds. A correction experiment was conducted on the data of three groups that had different numbers of perennial rivers. The original HYDL layer encompassed a total of 751 structural and center lines of rivers; 445 lines had a GB code of 210101002 (i.e. perennial rivers), while 227 lines' GB code was 999999 (to be corrected). Figure 4a presents the original data.

The durations of the automatic data processing and corresponding results are shown in Table 2, Table 3, and Figure 4b. It is found that after processing, the HYDL layer contains 637 perennial rivers, a 112 increase from the original number. The processing of the 227 rivers that needed correction produced only one error, an accuracy rate of 99.56%. The average duration of the three experimental runs was calculated at 1.625 seconds, indicating that the automatic processing is much efficient than manual work. In the past, cartographers needed to frequently look up existing specific attributes when filling in attributes, which is inefficient, effortful and prone to error.

Experimental run	1	2	3	Average
Duration (s)	1.621	1.672	1.581	1.625
T 11		C		

Table 2. Durations of processing

Experiment	Count
Number of original river	751
Number of original perennial rivers	445
Number of rivers to be corrected	227
Number of perennial rivers after processing	637
Number of rivers processed error	1
Number of rivers needing correction after processing	0
Accuracy rate	99.56%

Table 3. Data after correction

Figure 4 compares the original HYDL and the corrected HYDL. The river system extracted from the original HYDL is noticeably incomplete due to missing data of perennial rivers. In contrast, the river system extracted from the automatically corrected HYDL exhibits a sound dendritic pattern, suggesting good results of the correction.



 (a) Original data
(b) Corrected data
Figure 4. Comparison of perennial river data before and after the automatic attribute assignment

5. CONCLUSIONS

When mapping the hydrologic data obtained from the general survey of national geographic conditions in China, cartographers find that the attributes of some linear hydrologic features are undetermined. In view of this problem, this study proposed an algorithm for automatic attribute correction. Three basic correction principles were put forward based on the structural characteristics of the data and topological relationships between them. Later, this algorithm was used to correct the perennial river data with unclear attributes from a city of Guangdong with the help of the WJ-III map workstation developed by the Chinese Academy of Surveying and Mapping. The experimental results show that this algorithm provided a high accuracy rate and generated a river network with high structural integrity and good topological connectivity. Future research will try to combine attribute assignment with determination of similarity between features in level, name, and other attributes, in order to further improve the reliability of attribute assignment.

ACKNOWLEDGEMENTS

This research was supported by the National SciTech Support Plan (2015BAJ06B01)

REFERENCES

Ai. T., *et al.*, 2007. The hierarchical watershed partitioning and generalization of river network. *Acta Geodaetica et Cartographica Sinica*, 36(2), pp. 231-236.

Chen. J., *et al.*, 2006. The refined calculation method of topological relationships between line objects. *Acta Geodaetica et Cartographica Sinica*, 35(3), pp. 255-260.

Chen. T., *et al.*, 2013. A method for extracting land-cover information from the data collected during the national geographic conditions monitoring. *Bulletin of Surveying and Mapping*, 8, pp. 84-86.

Chen. Y., *et al.*, 2012. Automated mapping based on NewMap DMP. *Science of Surveying and Mapping*, 37(10), pp. 1-4.

Chen. Y., *et al.*, 2015. Research and implementation of automated mapping based on data from the general survey of national geographic conditions. *Bulletin of Surveying and Mapping*, 7, pp.106-108.

Gao. H., *et al.*,2016. Creating maps of Gaoling County's geographic conditions. *Standardization of Surveying and Mapping*, 1, pp.47-48.

He. J., *et al.*, 2013. Design of atlas to map results of the national geographic conditions monitoring. *Geomatics Technology and Equipment*, 4, pp. 35-38.

Li,D., *et al.* 2014. Multilevel grid representation of national geographic conditions. *Geospatial Information*, 1, pp.1-5.

Chen. J., 2012. Study Notes on Geographic National Condition Monitoring. *Acta Geodaetica et Cartographica Sinica*, 41(5), pp.633-635.

Li. J., *et al.*, 2014. Feature extraction and change cause analysis of water system in geographical conditions monitoring. *Journal of Water Resources and Architectural Engineering*, 2, pp. 188-192.

Li. L., *et al.*, 2008. A method for plotting gradual change symbol of single-line stream. *Bulletin of Surveying and Mapping*, 11, pp.64-67.

Liu. P. 2013. Organization and management of data on basic municipal geographical conditions and data application system development. Thesis(PhD), Taiyuan University of Technology.

Song. Y., *et al.*, 2014. Application of data from the general survey of national geographic conditions. *Standardization of Surveying and Mapping*, 2, pp.5-7.

Wang. X., *et al.*, 2015. Quality examination of data about national geographic conditions. *Geospatial Information*, 2, pp.18-19.

Wu. W., *et al.*, 2012. On contents and classification of national geographic conditions monitoring. *Standardization of Surveying and Mapping*, 4, pp. 10-12.

Xu. Q., *et al.*, 2015. Quality control of information collection results of geographical national conditions. *Jiangxi Coal Science & Technology*, 4, pp.89-91.

Yu. L., *et al.*, 2014. Design and implementation of digital mapping in the river gradient method. *Journal of Shandong University of Technology (Natural Science Edition)*, 5, pp. 76-78.

Zhang. Q., *et al.*, 2012. Surveying and mapping and national geographic conditions monitoring. *Bulletin of Surveying and Mapping*, 11, pp. 78-80.