A STUDY ON METHODS FOR QUALITY CONTROL OF SPATIAL DATA IN THE POYANG LAKE AREA

Xiaosheng Liu [a,b,*], Weibo Wu [b], Xiaoli Zhu [c]

[a] Civil Engineering Institute, Tongji University, Shanghai, China - Lxs9103@163.com
[b] School of Environmental and Architectural Engineering, JiangXi University of Science and Technology, Ganzhou, China - (Lxs9103,wuweibo1999)@163.com
[c] College of Environment and Planning , Henan University, Kaifeng, China - xlzhu1981@163.com

KEY WORDS: Classification, Data, Floods, GIS, Metadata, Spatial

ABSTRACT:

The Poyang Lake is the largest fresh water lake in China, of which the vast area is about 39955 km^2. The Poyang Lake Area, on the other hand, is the place where floods often occur. In order to reduce the impact and damage of floods on the lives and properties of people living in the Poyang Lake Area, the authors have developed a Geographical Information System for preventing and controlling floods in the Poyang Lake Area with the strong assist of the Science and Technology Office and the Water Conservancy Office in Jiangxi province. The system blended various techniques to obtain, transmit and process spatial data. Various techniques produced a large quantity of spatial data, which provided abundant and thorough information support for the fast and accurate decision on preventing and controlling floods, but a higher request to control the quality of spatial data was brought forward to us at the same time. Therefore, the authors studied the problem of quality control of spatial data while developing the Geographical Information System for preventing and controlling floods in the Poyang Lake Area. According to the occurring sequence of quality problems of spatial data, the authors classified these problems, analyzed their reasons and proposed the control methods of them. This paper depicts and discusses the whole process of the study, at five stages of obtaining, storage, processing, transition and application of spatial data.

1. INTRODUCTION

The main purpose of the study was to control the quality of spatial data, and offer high accuracy, improved timeliness and multi-sources spatial data for the GIS for preventing and controlling floods in the Poyang Lake Area. The study aimed at finding out methods of quality control of spatial data and controlling the quality of spatial data in the GIS for preventing and controlling floods in the Poyang Lake Area.

The authors analysed and researched the problems according to the occurring sequence of quality problems of spatial data in the course of constructing the system. In this study, the authors classified the quality problems of spatial data, analyzed the reasons and proposed the control methods of them at five stages of obtaining, storage, processing, transition and application of spatial data. First of all, the rational methods to control the quality of spatial data aiming at the reasons for them

2. METHODS AND MATERIALS

The quality control of spatial data involved in the whole course of constructing engineering, because the construction of spatial data ran through the whole course of constructing the GIS. At five stages of obtaining, storage, processing, transition and application of spatial data, we analyzed the reasons of the problems and explained the methods of controlling the quality of spatial data used in the GIS for preventing and controlling floods in the Poyang Lake Area one by one.

2.1 Obtaining and Quality Control of Spatial Data

In the GIS for preventing and controlling floods in the Poyang Lake Area, there were mainly three ways of obtaining spatial data that were used in the course of constructing engineering: field survey, digitizing paper map and application of Remote Sensing data. In the study, we gave the classification of the errors which were produced in the course of obtaining spatial data in these three ways, analyzed the reasons of these above-mentioned errors and gave their treatment methods one after another.

2.1.1 Errors of Field Survey and Treatment Methods: The errors of field survey were mainly as follows, (1) The artificial errors: errors of centring, errors of reading, errors of calculating, etc.; (2) Errors of instruments: the collimation axis was not perpendicular to the cross-axle, scores of level circle were asymmetry, instruments were not collimated, etc.;(3) Impacts of environmental elements: climate, atmospheric pressure, temperature, magnetic field, transparency, lateral refraction, etc..(WANG Ying, 2002)

All the above errors of field survey were strictly limited by the country's relevant norm and rules in the surveying and mapping standardization of China. So, we controlled these errors in the course of obtaining spatial data in field survey according to the norm and rules.

2.1.2 Errors of Digitizing Paper Map and Treatment Methods: There were three kinds of errors in the course of digitizing paper map: (1) Errors of the original paper map: They were some errors in the paper map used to digitize, including errors of control points, error of projection, errors of compiling, errors of fair drawing, errors of duplicating, etc.; (2) Distortion errors: They were inflation and shrink of the paper map because of the environmental conditions such as humidity and temperature were changed. Generally under the temperature-

resistant situation, if the humidity increases to 25%, then the size of the paper map may change 1.6%. In the course of printing, the paper map changes widely with the rising of temperature first, but shrinks if temperature cools again, the net growth in the long and width is about 1.25% and 2.5%, finally, the range of the distortion errors is 0.24-0.48 mm; (3) Errors of vectoring: For vectoring, we scanned the topographic map, transformed it to raster bitmap, and changed it into vector diagram by using the software of interactive vectorization. Factors influencing the quality of vectoring the map were as follows, resolution ratio of the scanner, precision of the scanner, the picture quality of the original map, human factor at the time of mutual vector, etc. (HAN Litao, 2003)

In order to avoid introducing and enlarging various kinds of above-mentioned errors, the authors chosen reliable quality and improved timeliness original map and with the high-accuracy scanner based on engineering standard firstly. When the map was artificially digitized, we checked figure part of the map through eyes or type the result of digitizing on the transparent picture to superpose with original map; and the attribute data of the map were checked according to the original map one by one. The errors of straight line and independent features should be smaller than 0.2 mm; errors of curve feature and river system were generally smaller than 0.3 mm; and errors of features which had fuzzy border should be smaller than 0.5 mm. When the error of matching edge was smaller than 0.3mm, we changed a feature among them to make the two features to be identical; when the error of matching edge was 0.3-0.6 mm, each of the two features was changed half; when the error was greater than 0.6 mm, we found out the reasons for them at fist, and took corresponding treatments.

2.1.3 Errors of the Application of Remote Sensing Data and Treatment Methods: The accumulating course of errors of the application of Remote Sensing data was divided into: data collection, data processing, data analysis, data transition and artificial interpretation.

However, in the GIS for preventing and controlling floods in Poyang Lake Area, Remote Sensing images were only used as the reference background. Therefore, errors of the application of Remote Sensing data were mainly the errors of artificial interpretation. We avoided this kind of errors depending on our own professional experience.

2.2 Storage and Quality Control of Spatial Data

The spatial data were stored in background database of the SQL Server by using ArcSDE (ArcSDE is a server software product used to access massively large multi-user geographic databases stored in relational database management systems) as middle component in the GIS for preventing and controlling floods in the Poyang Lake Area. In the course of depositing spatial data, errors were produced when the precision of number value did not enough or the spatial precision of images was too low.

In order to avoid this kind of errors, we designed the spatial database of the GIS for preventing and controlling floods in the Poyang Lake Area before the spatial data were stored into the spatial database. First of all, we did not adopt the systematic acquiescence precision, but figured out the storage units according to the precision demand of region in the Poyang Lake Area. The quality of spatial data, which were stored with the storage units that we had figured out, was greatly improved. Secondly, we calculated and established the domain of every attribute of spatial data to avoid storing invalid data.

2.3 Processing and Quality Control of Spatial Data

In the course of processing spatial data, errors were produced because of unreasonable classification, interpolation of value, integrating and synergy of multi-sources data and so on.

In order to avoid the classification error, we made a series of rules to guarantee the suitable granularity of classification and the consistency of the meaning of the concept, basing on the international standard of classification.

The key factors of interpolation methods were the choice of interpolation function and the building of interpolation model. As to features that changed at a long distance such as terrain, we adopted the method of least squares to interpolate their trend surface. Moreover, when we exported parts of the map, we adopted Kriging interpolation model, and chose the appropriate shape and size which indicated spatial relationship for them to improve the part-estimation value of points or areas in calculating.

When the multi-sources data were merged, we mainly used the method of geographical relation, which was the way to judge the logic mistake of the spatial data, according to the spatial distribution rules and interrelation of the geographical elements. Certainly, the method depended on the professional experiences of experts and professors. For example, when we overlaid the two layers data of river and contour lines (Figure 1.), there was at least quality problem in one layer of the two layers data if the river was not on the connection line of protruding points in the contour. At that moment, we overlaid one layer of the two layers data with another reliable layers, did further analysis to find out errors and corrected them.

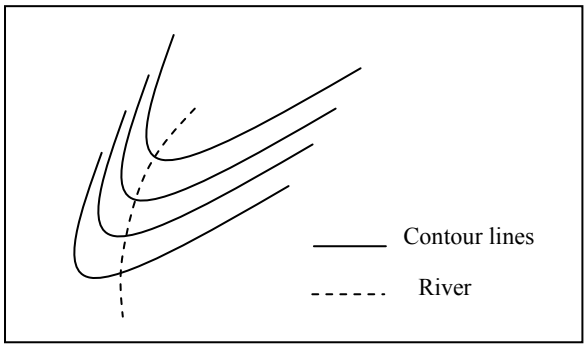

Figure 1.Overlaying River and Contour lines Layer

2.4 Transmission and Quality Control of Spatial Data

In the course of data transmission, errors would be produced due to the unstable stability of transmission media. For instance, real-time data on rain and water regime were mainly acquired through joining the long-range database in the GIS for preventing and controlling floods in the Poyang Lake Area. The data were transmitted with the special communication network for preventing and controlling floods in the Poyang Lake Area. However, in the course of transmission, the discontinuous phenomenon such as discontinuity points and intermittence lines emerged.

For these errors we mainly made use of metadata method, which had two types: explaining metadata and controlling metadata. The explaining metadata described contents, quantity, condition and the background information of data, while the controlling metadata was mainly used for the control of operating process (HU Jinlin, 2004). When the receipted data were not accord with their metadata, we applied techniques of numerical interpolation and numerical fit. After interpolated and fitted, line of rainfall process and profile of river were continuous, further more, errors of them were minimized.

2.5 Application and Quality Control of Spatial Data

When the spatial data were used to do various spatial analyses, errors would occur in two cases of overplaying layers of spatial data and introducing application models.

2.5.1 Errors of Overlaying Layers of Spatial Data and Control Methods: At the moment of overplaying layers of spatial data, redundant polygons were often produced (Servingne S.A, 2000). In order to dispel the influence of redundant polygons on space analysis, we setup the limit rate of relative area of polygons overplayed in the procedure of space analysis. When the area of polygons was smaller than certain limit value, they were filtered in the spatial analysis program.

2.5.2 Errors of Introducing Application Models and Control Methods: In the GIS for preventing and controlling floods in the Poyang Lake Area, there were several kinds of application models to be used, such as flood predicting model, reservoir dispatching model, flood progress model, flooded area model and the best withdraw route model of victims. In the following we took reservoir dispatching model which was most frequently used in preventing floods for example to discuss the choice and setting-up of models

Many reservoir dispatching models were built through nearly 40 years research. According to the characteristic of input parameter and objective function, these reservoir dispatching models were divided into four big classes: determinacy reservoir dispatching model, randomness reservoir dispatching model, fuzzy reservoir dispatching model, multi-goals decision theory model. We selected multi-goals decision theory model for use, of which the goal function and restraining condition were the minimum loss and dependability restriction. Although the calculating amount increased, errors of introducing model were greatly reduced, because random factors what the model involved were less than other models.

3. RESULTS AND DISCUSSIONS

The key tasks of the quality control of spatial data were how to check out quality problems rapidly and how to revise them in time. The authors put forward many methods of checking and processing the errors of spatial data, which occurred in the practical work of constructing the GIS for preventing and controlling floods in the Poyang Lake Area. However, there were a large number of problems which need to be resolved, due to the inherent complexity, randomness and dynamic change of research problems. In order to keep the timeliness of data, we need to introduce new access technology of spatial data. While raising the efficiency of data access, the new technology gave us a higher demand for controlling the quality of spatial data. Therefore, we need to do further research in the future.

4. CONCLUSIONS

By above-mentioned various means, we effectively controlled the quality of spatial data used in the GIS for preventing and controlling floods in the Poyang Lake Area, subsequently supplied high precision , timeliness and multi-sources spatial data for the construction and run of the system. The study on quality control of spatial data made the system run more steadily and provide better services for preventing and controlling floods in the Poyang Lake Area.

REFERENCES

HAN Litao, 2003. Approach to related problems regarding the spatial date quality. *Northeast Surveying and Mapping*, 26(1), pp. 11-14.

HU Jinlin, 2004. The extended metadata method of spatial data quality management. *Modern Surveying and Mapping*, 27(3), pp. 21-24.

Servingne S.A, 2000. Methodology for Spatial Consistency Improvement of Geographic Databases. *GeoInformatica*, 4(1), pp. 7-34.

WANG Ying, 2002. Discuss about sources of errors of GIS data and methods to correct them. *Hydrographical Surveying and Charting*, 22(1), pp. 18-20.