

# An Algorithm of Automatic Administrative Region Index Map Generation for National-level Map Production

Ruolei Wang<sup>a, \*</sup>, Jianjun Liu<sup>a</sup>, Zhao Li<sup>a</sup>, Jun Zhang<sup>a</sup>

<sup>a</sup> National Geomatics Center of China, Beijing, China; wangruolei1007@163.com; liujianjun@ngcc.cn; lzhn@nsdi.gov.cn; junzhang@ngcc.cn

## Abstract:

Since the project of dynamic updating of national fundamental geographic information database was launched, National Geomatics Center of China has published the newest version of multi-scale databases, derived databases and subdivision maps once a year. As one of the derived cartographic databases, administrative region index database allows users to find maps covering their regions of interest quickly by input the name or number of the relevant map. Due to the boundaries or attributes of some administrative regions may often change, updating of administrative region index database is necessary. To improve efficiency and quality in the national-level map production and database construction, an algorithm of administrative region index map automatic generalization has been proposed in this paper to replace the traditional manual production lines, which are time-consuming and uncontrollable in quality. The method goes in depth with handling coordinate recognition error accumulation and residue lines caused by the clip.

**Keywords:** China, map production, administrative region index database, index map, automatic

## 1. Introduction

In order to satisfy the user's high requirements for the timeliness and content richness of geographic information data, the project of dynamic updating of national fundamental geographic information database has been launched for seven years. National Geomatics Center of China has published the newest version of multi-scale topographic databases (1:50,000, 1:250,000 and 1:1,000,000), derived databases (cartographic database, administrative region index database, DEM database, etc.) and subdivision maps once a year (Wang & Liu, 2015).

Despite the popularity of database query functions, as one of the derived databases, the administrative region index database is still relatively more convenient and fast for users to find their interest parts from a geographic database with large amounts of data. The index database usually works as a grid covering the whole geographic database (Jensen, 2004).

As a big map production project covering China's national territory, the number of maps with a scale of 1:50000 to be produced exceeds 24 thousand. Referring to the core idea of this project, linkage update, the number of maps involving changes in administrative boundaries or attributes that need to be updated each year is still close to 10 thousand. These maps are also the basis for the annual establishment of administrative region index database. For such a huge amount of data, manual work will bring high production cost and will almost

unavoidably lead to various human errors lowering map production quality. This problem can be relieved by the algorithm of automatic administrative region index map generation that presented in this paper.

## 2. Methodology

According to the concept of linkage update in this project, as shown in figure 2, the main idea of the method is to update the boundary line, polygon features and their attributes of the administrative region index map from that of main map section in a subdivision map, and then the administrative region index database is built from all the subdivision maps.

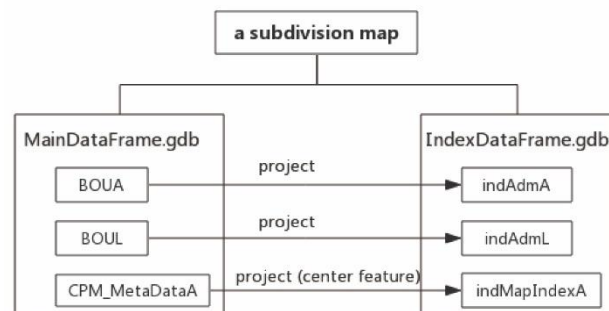


Figure 1. Figure placement and numbering: the famous ICA logo.

As the project designed, each subdivision map is composed of two geodatabases, MainDataFrame.gdb and IndexDataFrame.gdb. The latter is projected from the former and their spatial data attributes are consistent.

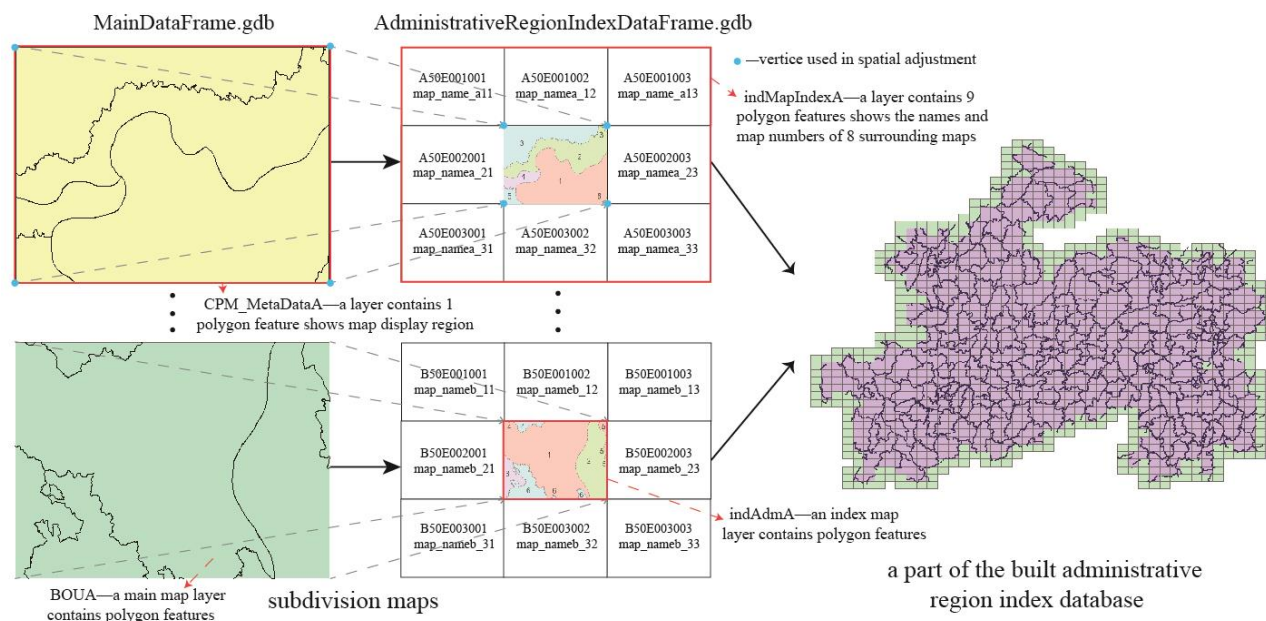


Figure 2: The main idea of the index map and database automatic generation algorithm

The corresponding projection relationships of boundary area, boundary line and frame scope layers of the above two databases are shown in figure 1.

Taking polygon features as an example, the detailed algorithm workflow is shown in Figure 3. Processing of the boundary line features is similar to the polygon ones with several steps omitted. Description of some used layers can be found in Figure 2. For each subdivision map, the first step is importing the layer “BOUA” from the main database into the index database, which makes it convenient for subsequent processes. Then the imported features in the new target layer should be clipped by the scope layer “CPM\_MetaDataA” to avoid repeat importing features into the final administrative region index database. The following sequence of processes, including merging and exploding, are necessary to prevent one feature from corresponding to multiple

separate polygons, which may cause errors when importing the spatial data into Oracle database in the last step. It is simpler in terms of boundary lines that the processes mentioned above could be omitted.

The key step for index map automatic generalization, spatial adjustment, should be done here by associating four sets of vertices from two corresponding layers, “CPM\_MetaDataA” and “indMapIndexA” (the center feature). Coordinates of the selected vertices involved in the calculation are automatically recognized by ArcGIS Engine tools. Before copying the adjusted features to the final target layer “indAdmA”, the existing old elements need to be deleted in advance. And a four-color map representation should be added subsequently to the adjusted features in layer “AREA”. If the features are boundary lines, changing the process here to adding linear symbolization. The final step is to fill in the

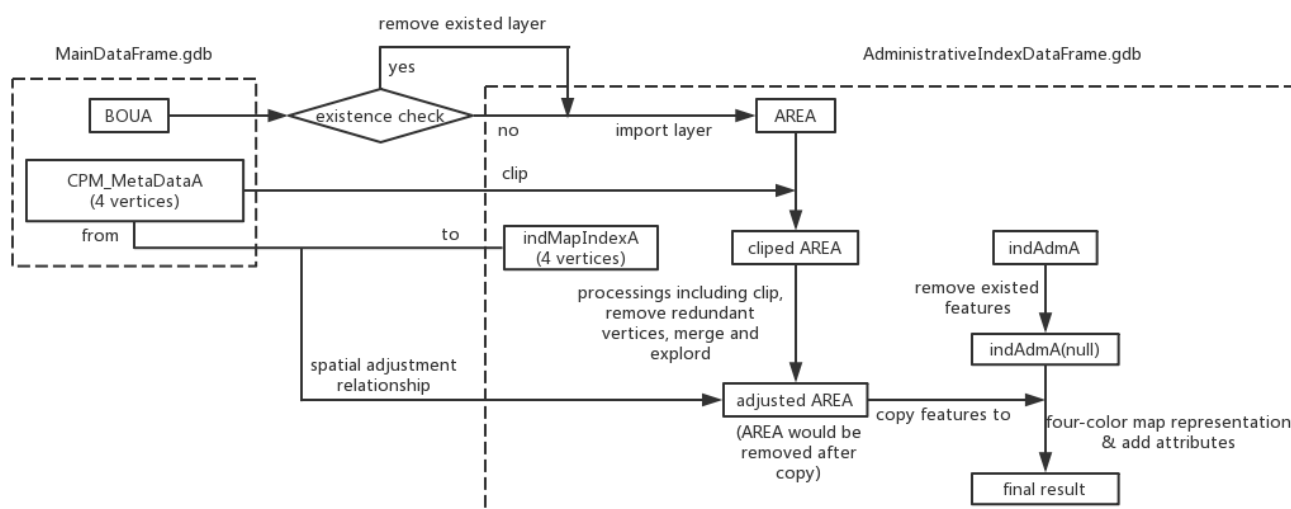


Figure 3: A detail workflow of the algorithm realization

missing feature attributes in the process of clipping, copying, pasting, etc. So far, the whole automatic generation process of index map for a subdivision map is basically completed. In the above processes, we mainly focus on two problems: error accumulation after successive spatial adjustment and dealing with the residue lines after clipping.

## 2.1 Spatial adjustment optimization

As mentioned above, before the annual geodatabase construction, the amount of data in map production is very large and China needs more than 24,000 subdivision maps with a scale of 1:50,000 to cover the whole country. In practical production, occasionally, it might be found that after the automatic generation of nearly one thousand maps, both of boundary line and polygon index map features might show spatial position deviation, and the deviation will continue to expand with the increase of the number of maps processed. An obvious spatial deviation example is presented in figure 4.

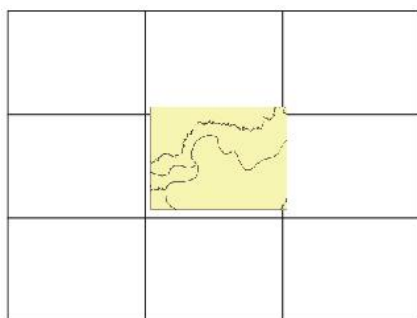


Figure 4: A spatial deviation after spatial adjustment

Due to the possible deviation do not only exist in scattered subdivision maps, but it has been there ever since it began and it has become more and more serious. Therefore, it can be inferred that the problem should not be resulted from the source and target layers but caused by coordinate recognition. Comparing with the vertex coordinates obtained manually, we realized that the decline in quality results from the accumulation of errors in the automatic recognition of vertex coordinates of layers (CPM\_MetaDataA and indMapIndexA).

Based on the initial consideration of improving operational efficiency, all subdivision map processing is done in the same workspace in computer memory. After many attempts, releasing memory and rebuilding workspace is found to be the easiest way to eliminate error accumulation. To solve this problem, we analyzed the mathematical relationship between the number of processed maps and spatial coordinate deviation and estimated the number of maps whose error accumulation is greater than the acceptable error threshold. Resetting the workspace in time after reaching the processed map number can effectively improve higher generation accuracy and maintain operation efficiency.

## 2.2 Clip process optimization

Another challenge is removing the residue lines caused by clip which are not allowed to appear. They cannot be found and repaired by geometry and topological checks. Because they belong to the clipped polygon features, spatial selection also does not work. By zooming in as much as possible, it can be found that residue lines are actually tiny polygons or narrow gaps that are hard to see with the naked eye, and they are usually caused by tolerance for vector data clipping.

The parameter, *xy\_tolerance*, determines the minimum distance separating all feature coordinates (nodes and vertices) as well as the distance a coordinate can move in X or Y, or both. And it is intended to integrate line and boundaries during the topological process, which means coordinates that fall within small distances of one another will be integrated. According to the instruction, although the default value for the parameter that is 10 times the default x,y resolution is recommended for most cases, users can set the value to be higher for data that has less coordinate accuracy and lower for datasets with extremely high accuracy.

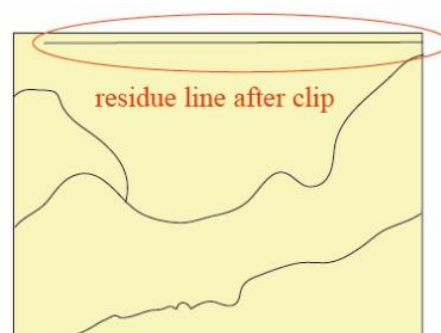


Figure 5: A residue line after clip process

Referring to the above methods, we have carried out many experiments. Because of the high accuracy of the data and the resolution of 13 decimal points, we adopt smaller tolerance, and the result of clipping has not been improved. When the value is large, the edge of the element becomes rough and the resolution decreases. From all the above, it can be found that the problem cannot be solved by simple adjusting *xy\_tolerance* ("ArcGIS Desktop Help," 2010).

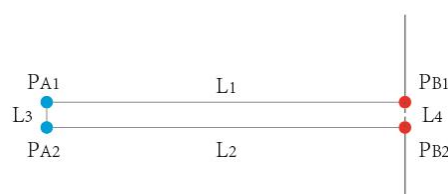


Figure 6: Enlarged diagram of residue line after clip

Generally, the problem arises from the polygon features rather than the line features. Therefore, we can get the coordinates of all the nodes in the "AREA" layer and project them to the Y-axis, which is to add the Y

coordinates of the points with the same X coordinates. In order to avoid the influence of irregular administrative boundary, the Y coordinate values of nodes on the corresponding boundary line are subtracted. In addition to the vertices of the upper and lower boundaries, the points whose y coordinates are greater than zero are redundant nodes that need to be removed. Figure 6 presents the enlarged diagram of the residue line after the clip. After removing point  $P_{A1}$  and  $P_{A2}$ , the residue lines  $L_1$ ,  $L_2$  and  $L_3$  will automatically disappear. The final task is to reconnect the two points,  $P_{B1}$  and  $P_{B2}$ , so that the feature can be a closed polygon.

### 3. Evaluation

Based on the proposed algorithm, we developed a tool for automatic administrative region index map Generation and applied it to practical cartographic map production. Meanwhile, an evaluation of this algorithm has been done according to the actual effect in production.

We operated on a total of 2980 subdivision maps at one time, which took about 21.67 hours to run on a graphics workstation without any unexpected pause or crash occurred. The average processing time of each map was 26.5 seconds. 300 maps, 10% of the total, were randomly selected for quality inspection. It was found that there were no unacceptable spatial position deviation and redundant lines. The cartographic representation (polygon colour or linear representation) and attributes are also correct, which has a positive impact on the subsequent construction of the administrative region index database. Therefore, we believe that the algorithm can not only meet the project design requirements but also improve the efficiency of mass map production.

### 4. Conclusion

In order to improve the efficiency and quality of map production and database construction in the project of building the national basic geographic information database, an automatic boundary updating algorithm is proposed. This paper introduces the specific process of the algorithm and focuses on solving two problems encountered in practical production, error accumulation elimination and clip process optimization. After the evaluation of the production results, the algorithm not only guarantees the quality of map production but also significantly improves the production efficiency, which has a certain reference value for the engineering of mass map production. The reason for the unexpected error coordinate recognition is still not clear, accordingly, the future work will focus on how to control error form the source.

### 5. References

- ArcGIS Desktop Help. (2010). Retrieved from [http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Feature\\_class\\_basics&anchor=dw\\_xytolerance](http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Feature_class_basics&anchor=dw_xytolerance)
- Jensen, K. L. (2004). Index maps for the digital age. *Information Technology and Libraries*, 23(2), 81–87.

Wang, D., & Liu, J. (2015). Key Techniques for Dynamic Updating of National Fundamental Geographic Information Database. *Acta Geodaetica et Cartographica Sinica*, 44(7), 822-825.