Accuracy Loss Analysis in the Process of Cultivated Land Quality Data Gridding

Lingling Sang, Dehai Zhu, Chao Zhang, Wenju Yun

HAL Id: hal-01220848
https://inria.hal.science/hal-01220848

Submitted on 27 Oct 2015

Accuracy loss analysis in the process of cultivated land quality data gridding

Lingling Sang[a,c], Dehai Zhu[a,c*], Chao Zhang[a,c], Wenju Yun[b,c]

[a] China Agricultural University, Beijing, 100083, P. R. China

[b] Land Consolidation and Rehabilitation Center, Ministry of Land and Resources, Beijing 100035, P. R. China

[c.] Key Laboratory of Land Quality, Ministry of Land and Resources, Beijing 100035, P. R. China

ABSTRACT. Spatial data gridding is one of the effective methods to solve the multi-source data fusion. In view of the current problems in the process of comprehensive analysis between the cultivated land quality data and other multi-source data, This paper, by adopting the method of the Rule of Maximum Area (RMA), converted the cultivated land quality data to the grid scale and analyzed the accuracy loss in the process of cultivated land quality data gridding in 6 grid scales (10M $\times$ 10M, 5M $\times$ 5M, 3M $\times$ 3M, 2M $\times$ 2M, 1M $\times$ 1M, 30S $\times$ 30S). Some conclusions have been reached. (1)The use of gridding methods will have assigned any analysis units to the specified data grid scales, and it provides a basis for spatial data integration, comprehensive analysis and spatial models construction;(2) Grid scale accuracy is higher, the original figure segmentation of cultivated land quality data is more serious, and grid results is more accurate, but grid computing time is increased step by step;(3) Through the study of the multistage of cultivated land quality data grid, the smaller the grid scale, the smaller the loss area of cultivated land quality, such as 10M $\times$ 10M gridding results lead to the most loss area, and Each grade area loss curve has a certain regularity;(4)From accuracy and computational efficiency, the most appropriate grid scale choice is 1M $\times$ 1M grid of 1:10000 cultivated land quality data of Daxing for the gridding processing.

∗ Corresponding address: College of Information and Electrical Engineering, China Agricultural University, 17 Qinghua East Road, 100083, PR China.
E-mail address: zhudehai@263.net .

1. Introduction

Cultivated land quality grading results, namely agricultural land classification results, is under the unified national standard farming system, and are based on solar and temperature potential productivty of appointment crop.There includes cultivated land natural quality grade, utilization quality grade and economic quality composite grade.Cultivated land natural quality is assessed by the natural conditions of grading units to calculate the theoretical yield of crop; and on the basis of cultivated land natural quality, there are assessed to get the grades of cultivated land utilization quality and the economic quality through the correcting step by step of the land use status and input and output level[1].The assessed results is mainly in view of the natural attribute, utilization, the social economic of cultivated land[2], and is quality evaluation in a specific time focusing on the capacity. However, the main parts of quality evaluation finally are the patches of cultivated land, and cannot achieve full coverage of the cultivated land quality assessment. In the process of practical application, the quantitative analysis which combined land quality data with land use data, soil data, environmental data, and other socioe conomic statistical data, must be the basis of administrative units to extract the relevant data. The work which the analytical methods are mostly correlation analysis is not only larger, but also is not conducive to developing mathematical models.

There are often different statistical units for model construction of multi-source data. This does not facilitate comparison and correlation analysis of multi-source data, and cannot fully play value of the existing data [3]. Spatial data gridding is one of the effective methods to solve this problem. It can not only improve the efficiency of management, but also brings about dynamic rule analysis under the support of GIS technology. The data are including the grid of the processed data and the derived results data. They could be formed on the space gradient between grid data which play the research of spatial differentiation , and formed in a grid-based data sequence which was based on the different units. Simultaneously, the gridded data have many advantages to match and integrate with multi-source data, especially they are suitable for spatial model construction to implement and express [3]. It is the basis work of a graphic rendering, scientific computing and space model implementation.

The study of spatial data gridding was carried out both at home and abroad. GIS data stored in the form of a polygon data, could also use the above method for grid transformation. There were different conversion methods between different data sources and grid data. The current gridding algorithms, most of which were regular (or irregular) the spatial distribution of point data. Spatial interpolation could achieve the conversion of point data from statistical units to the grid units. The traditional method was weighted average method to obtain the value of a grid point according to finding several nearest points in some rules spatially [4]. Interpolation methods could be used to deal with socio-economic data from the irregular grid to a regular grid [5]. In processing of meteorological data, the gridded results could better reflect the characteristics of the discrete data by comparing Kriging method and inverse distance interpolation method[6].Interpolation method of eight faceted search methods could ensure the accuracy of the grid interpolation for contour data[7].As well as, GIS spatial overlay between the layer files and the grid files could achieve data connection by the public fieled, and finally calculated attribute value of each grid cell using the weighted average method.In order to make remote sensing data can be matched with the grid scale data,we had to converse raster to vector data, and then used the method of grid processing.Of course, remote sensing data could also be obtained by resampling the grid cell size to match results. Moreover, Yang Cunjian and Cheng Jiehai discussed the accuracy of area loss during the conversion from the various land use types with different grid sizes [8,9,10].

Currently, there are no literature data to discuss the gridding methods and precision analysis for cultivated land quality. This paper chooses the Daxing district as a study area to analyze the accuracy of area loss during the conversion from 1:10000 cultivated land quality data to different grid scales.

2. Study Area and Data Sources

Daxing district, as the study area, is located in the south of Beijing, neighboring Tongzhou district in the east, west of Fangshan district across the Yongding River and Gu'an and Bazhou in Hebei Province in the south. The longitude is from 116 °13' to 116 °43 ', and the latitude is from 39 °26' to 39 °51'. There are 14 townships and 527 villages with a total area of 103 595.39 hectares, accounting for 40.18% which is the largest. There are not much available land area for development and utilization, the potential of land development is very limited.

The paper took 1:10000 cultivated land quality data in 2010 for the works, and the patches of existing cultivated land where the main object of the study. The nature quality of cultivated land in the study area was relatively homogeneous, of which the indicator differences in topsoil texture, profile configuration, salinization, organic matter content and other natural factors were not very significant. From the field investigation, most of the drainage condition indicators were 1-2 levels, and the irrigation rate indicators were" fully satisfied". According to the statistics of 1:10 000 cultivated land quality data in 2010 of Daxing district, cultivated land natural quality indexes were between 2 100 to 2 680, and the natural quality grade was from 11 to 14. Cultivated land utilization indexes were between 1 300 to 2 000, and the utilization quality grade was from 14 to 20, of which the area ratio of cultivated land of the maximum grade and the minimum grade was 4.9% and 6.9%. The grade distribution of cultivated land utilization index in Daxing district was put up a gradually decreasing tendency from the northeast to the southwest.

3. Gridding method of cultivated land quality data

3.1 Selection of grid cell

The grid data is generated in the ArcGIS9.3.The methods are as follows:

(1) Defining the geographic coordinate system.

Xi 'an 80 coordinate is one of the current commonly used coordinate system, so GCS_xian_1980 is defined as the geographic coordinate system in the study.

(2) Selecting the starting point and grid interval.

The starting point is the intersection point with the equator and 0 °longitude [11,12]. Firstly, the basic grid of the gird division of the China is based on the fixed longitude and latitude interval. Then the basic grid is divided into the next grid using the fixed longitude and latitude interval.

(3) First-level gridding results of the Chinese mainland

According to China's latitude and longitude range, the most Western longitude is 73.66E, the most eastern longitude is 135.08E, the most northern latitude is 53.52N and the most southern latitude is 4.00N. The latitudes from north and south are nearly 50 °, and the longitudes from west and east are nearly 60 °. The gridding results are as follows:

☞ There are divided into 64 zones (73°-136°) from the starting longitude73° along the longitude direction based on the longitude interval (1°);

☞ There are divided into 51 zones (4°-54°) from the starting latitude 4° along the latitude direction based on the latitude interval (1°);

☞ At this moment, the earth is divided into 64 × 51 grid cells, which constitutes the first-level subdivision units of the grid system, and the span of each grid is 1°×1°. Combined with the location of Chinese mainland, there are identified 1148 grids, each of which is 1°×1° grid of an area of approximately 0.739 ×10^4~1. 227 ×10^{4}km^2.

(4) Determining the first-level grid of the study area

Using spatial overlay analysis in ArcGIS, we can pick out the first-level grid of the study area which are based on the above gridding results.

(5) Subdivision

The starting point is based on the first-level grid in the study area, then start to subdivide in latitude and longitude. There would establish 10M×10M, 5M×5M, 3M×3M, 2M×2M, 1M×1M, 30S×30S multi-level grids which preparing for gridding process and data analysis of cultivated land quality (Figure 1).

| 10M×10M | 5M×5M | 3M×3M |

| 2M×2M | 1M×1M | 30S×30S |
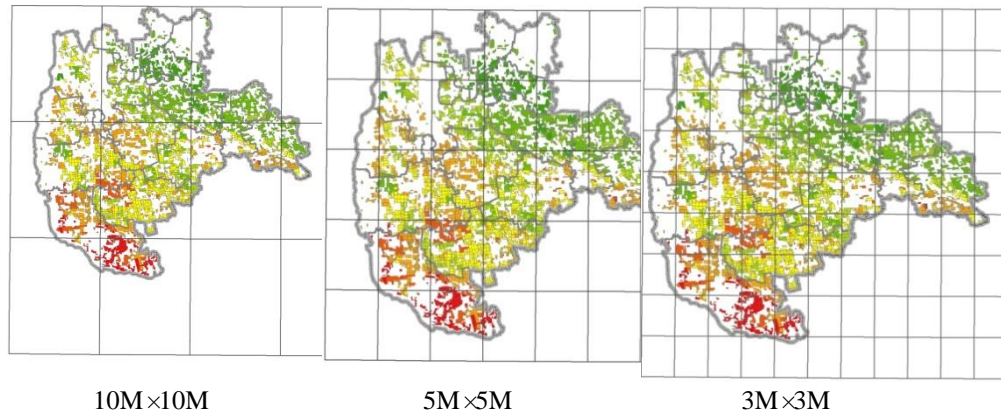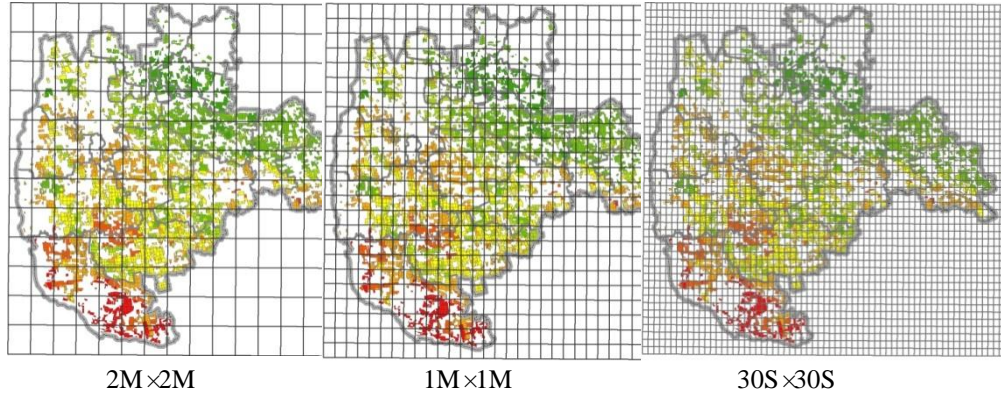
**Figure 1** Subdivision in study area (Note: M—minute；S—second; 10M×10M—10 minute gird; 5M×5M—5 minute gird; 3M×3M—3minute gird; 2M×2M—2 minute gird; 1M×1M—1 minute gird; 30S×30S—30 second gird)
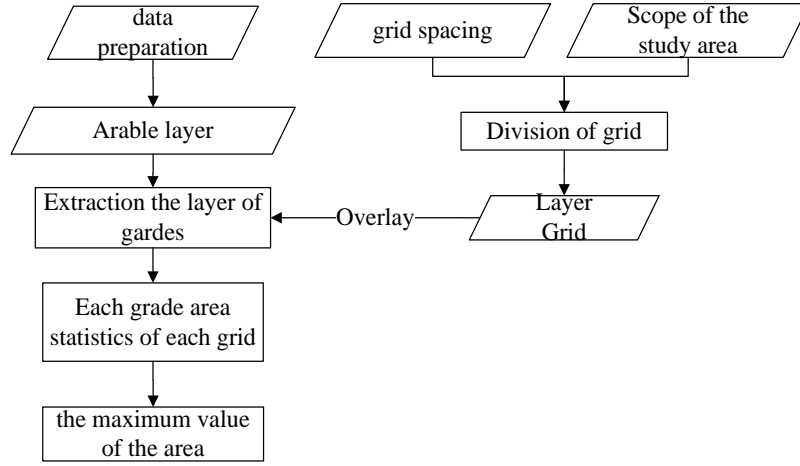
3.2 Selection of gridding method

Gridding is not the meaning of rasterization. Traditional rasterization can also be obtained by resampling the grid cell size to match results , and it is a process along with attribute information loss [13]. The reason is that the original grid exists in mixed types. This paper focuses on the different vector data which was gridded , and try to achieve the purposes of less attribute information loss.
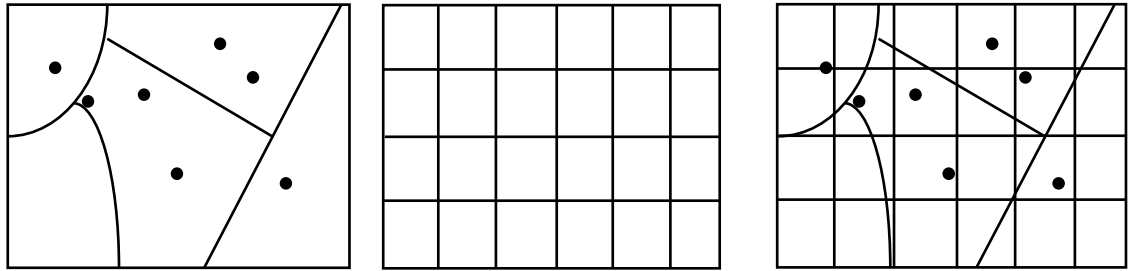
Figure 2 Multi-source data into a regular grid

According to the conversion of spatial data to the grid cell, the common methods are including the maximum value of the area method (Rule of Maximum Area, RMA) [14], the central attribute value (Rule of centric cell, RCC) [15] and the simple area weighted method. Among them, The RMA method is most commonly used that the value of a grid cell is the largest type of attribute values. If there are two or more dominant types, there can randomly select one of them as the output of the grid cell. The process is shown in Figure 3.

```
┌─────────────┐       ┌─────────────┐   ┌─────────────┐
│    data     │       │grid spacing │   │ Scope of the│
│ preparation │       │             │   │  study area │
└──────┬──────┘       └──────┬──────┘   └──────┬──────┘
       │                     └─────────┬───────┘
       ▼                               ▼
┌─────────────┐              ┌─────────────────┐
│ Arable layer│              │ Division of grid│
└──────┬──────┘              └────────┬────────┘
       │                              │
       ▼                              ▼
┌─────────────────┐              ┌─────────┐
│Extraction the   │◄──Overlay────│  Layer  │
│layer of gardes  │              │  Grid   │
└────────┬────────┘              └─────────┘
         │
         ▼
┌─────────────────┐
│  Each grade area│
│statistics of    │
│  each grid      │
└────────┬────────┘
         │
         ▼
┌─────────────────┐
│the maximum value│
│  of the area    │
└─────────────────┘
```

Figure 3 Rule of maximum area

3.3 Accuracy loss analysis

In the research on gridding process and analysis of cultivated land quality data, firstly, selecting each grade area of cultivated land before the conversion as the reference data; then comparing the gridding data and the reference data of cultivated land quality data, there would achieve the patch number and area change results of cultivated land quality in different grid scales; finally, calculating such other area of precision loss with grid scale changes. Calculating formula is as follows:

$$E_i = A_{gi} - A_{bi} \tag{1}$$

$$L = \sum_{i=1}^{n} |E_i| \tag{2}$$

Where E_i is the area loss of grade i, the positive value is larger than the reference area, the negative value is smaller than the reference area; A_{gi} is the area of grade i after gridding; A_{bi} is the reference area of grade i; L is the overall area loss which is equal to the sum of the absolute value of area loss of each grade.

4. Results and analysis

## 4.1 Changes of spatial distribution of each grade

Cultivated land quality data in Daxing district is processed by the above gridding method. In order to compare with the results, there are given 6 gridding results map with different grid scales in Figure 4.

From Figure 4, it can be seen that as the grid scale becomes smaller, the gridded results of cultivated land quality data are closer to the real data. However, there will increase the amount of computation. The disadvantages are as follows by comparing cultivated land quality raw data and the gridded data:

(1) Cultivated land quality raw data was based on the cultivated land patches as analysis units, which it was not only factored on the administrative unit instability, but it would lose any meaning of data monitoring for cultivated land quality;

(2) The quantitative analysis which combined land quality data with land use data, soil data, environmental data, and other socio-economic statistical data, must be the basis of administrative units to extract relevant data. The work which the analytical methods are mostly correlation analysis is not only larger, but also is not conducive to developing mathematical models.

Above all, the gridding method presented in the paper can not only maintain the basic characteristics of the raw data, but also it is superior than cultivated land quality raw data in multi-source spatial data integration and comprehensive analysis.

| 10M×10M | 5M×5M | 3M×3M |

<center>2M×2M   1M×1M   30S×30S</center>
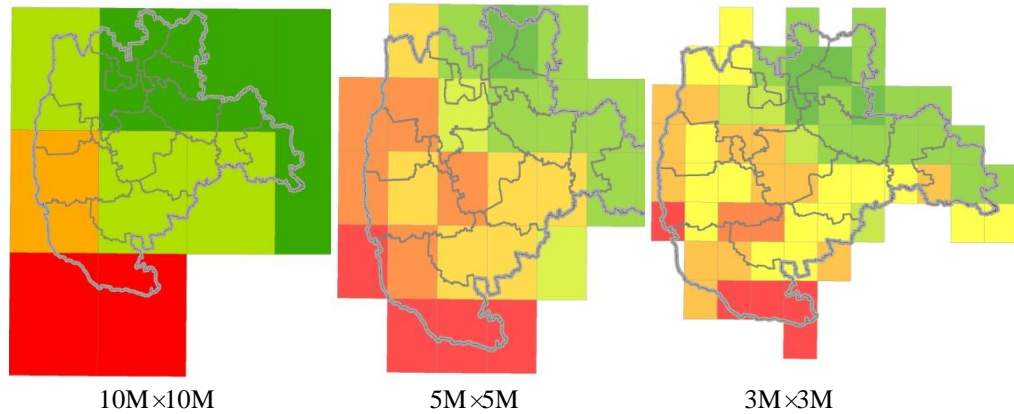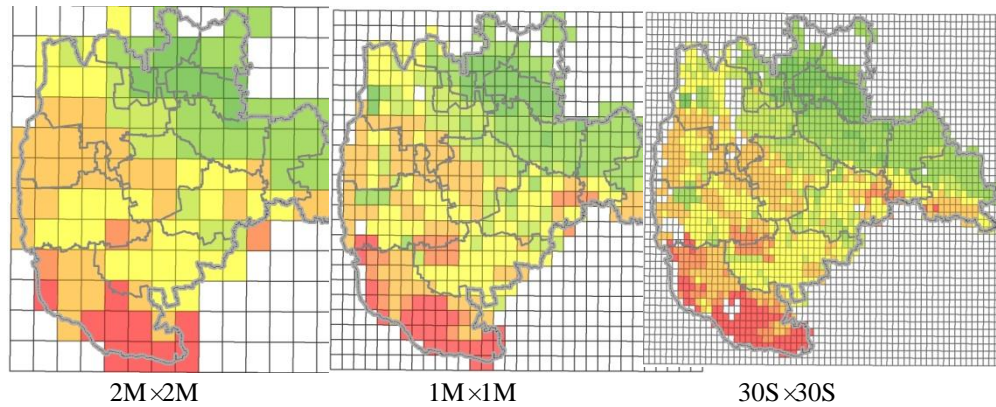
Figure 4 Gridding results map

4.2 Analysis of cultivated land quality data

According to the statistics of 1:10000 cultivated land quality data in 2010 of Daxing district, there were 5191 patches of cultivated land. Through grid processing (10M × 10M, 5M × 5M, 3M × 3M, 2M × 2M, 1M × 1M, 30S × 30S), the original patches were divided and the grid scale cultivated land quality data information was shown in Table 1, where the 30S×30S grid division of the original data was most serious, eventually reached 10120 patches.

Table 1 Statistical analysis of cultivated land quality data based on grids

Grid	10M_Grid	05M_Grid	03M_Grid	02M_Grid	01M_Grid	30S_Grid
Number of patches	5388	5609	5915	6313	7432	10120
Area of the SD	439.47	155.55	65.63	32.08	10.02	3.85

In addition, there were some differences of each grade sequence integrity through grid processing of the original data in Table 2. The 10M × 10M gridding results were in loss of grades 15,18,20; The 5M ×5M gridding results were in loss of grades 15.The other girdding results were no loss of any grades, but their area loss was different.

Table 2。 Grade sequence integrity after gridding (RMA)

|	RMA_10M	RMA_5M	RMA_3M	RMA_2M	RMA_1M	RMA_30S
14	√	√	√	√	√	√
15	×	×	√	√	√	√
16	√	√	√	√	√	√
17	√	√	√	√	√	√
18	×	√	√	√	√	√
19	√	√	√	√	√	√
20	×	√	√	√	√	√

4.3 Accuracy loss analysis of each grade

Selecting tools from Arctoolbox\Data Management Tools\Generalization\Dissolve in ArcGIS, there can add up the cultivated land area of each grade after gridding, and compare with the reference data in Figure5 and Figure 6. Some conclusions have been reached.

(1) The smaller the grid scale , the less area loss of cultivated land quality;

(2) The 10M × 10M gridding results lead to the most area loss which is 280.01km^2 ; The 30S × 30S gridding results lead to the lest area loss which is 20.63km^2 ;

(3) Area loss curve of each grade has a certain regularity, which 10M ×10M, 5M × 5M, 3M × 3M, 2M × 2M gridding results were more area loss in the grades of 16 ,17,18,19, and less area loss in the grades of 14,15;1M × 1M, 30S × 30S gridding results were less area loss in all the grades.
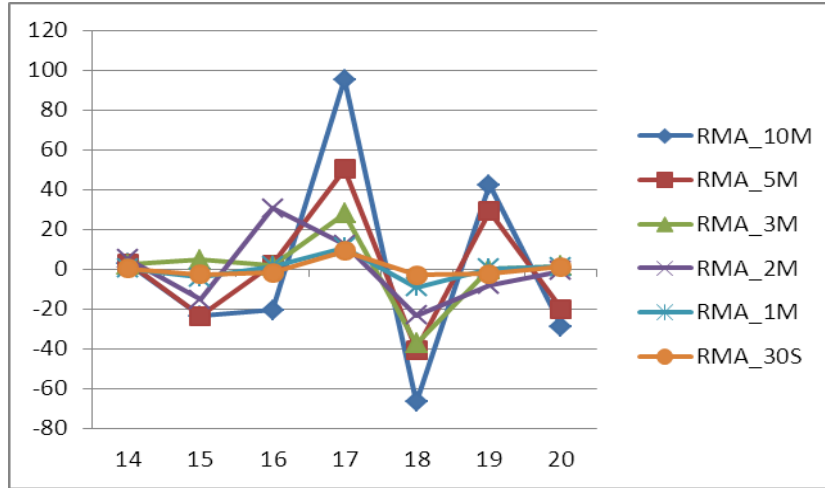
Figure 5 Area loss of each grade in different girdding scales (km^2)
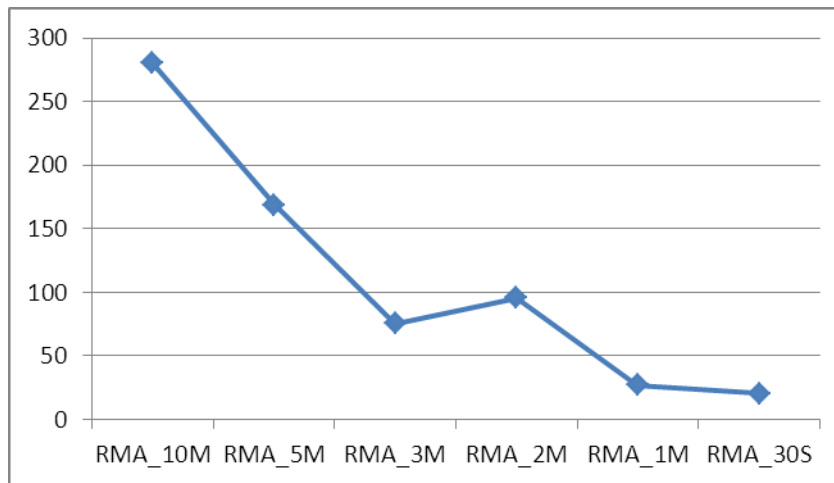
Figure 6 The overall area loss of of each grade in different girdding scales (km^2)

5. Conclusion and Discussion

This paper chose Daxing district as a study area and adopted the method of the Rule of Maximum Area (RMA) to analyze the accuracy of area loss during the conversion from 1:10000 cultivated land quality data to different grid scales and analyzed the accuracy loss in the process of cultivated land quality data gridding in 6 grid scales (10M ×10M, 5M ×5M, 3M ×3M, 2M ×2M, 1M ×1M, 30S ×30S). The

foregoing analysis could be drawn：

(1) Grid scale accuracy was higher, the original figure segmentation of cultivated land quality data was more serious, and grid results were more accurate, but grid computing time was increased step by step;

(2) Through grid processing (10M ×10M, 5M ×5M, 3M ×3M, 2M ×2M, 1M × 1M, 30S ×30S) and studying the multistage of cultivated land quality data grid, the smaller the grid scale was, the smaller the area loss of cultivated land quality was, such as the 10M × 10M gridding results led to the most loss areas, and Each grade area loss curve had a certain regularity;

(3) The 1M × 1M and 30S × 30S girdding results were no loss of any grades which maintaining the grade sequence integrity after gridding. The area loss of cultivated land quality was less, and the number of patches by the 30S × 30S grid were 26.6% less than 1M ×1M grid. From accuracy and computational efficiency, the most appropriate grid scale was 1M ×1M grid of 1:10 000 cultivated land quality data of Daxing for the gridding processing.

The use of gridding methods would have assigned any analysis units to the specified data grid scales, and it provides a basis for spatial data integration, comprehensive analysis and spatial models construction. They can provide relevant researchers valuable references for spatial data processing and analysis. In addition, there are large differences in different scales and different research areas, such as the choice of the grid cell, the choice of the gridding methods and so on. It is to be carried out in-depth research in the future.

Acknowledgements

References

[1] GB/T 28407-2012.Regulation for gradation on agriculture land quality [S]. 2012.
[2] Sang Lingling, Zhang Xiaopei, Yang Jianyu, et al. Construction and empirical analysis of green productivity evaluation system of agricultural land [J]. Transactions of the CSAE, 2010, 26 (7): 235－239.

[3] Fan YiDa, Shi PeiJun, Gu Zhihui, et al, A Method of Data Gridding from Administration Cell to Gridding Cell [J]. Scientia geographica sinica, 2004，24 (1) ：105-108．

[4] Bai Yan, Liao Shunbao. Implement Method of Vector Data Rasterization That without Attribute Information Loss: A Case Study of 1：250 000 Land Cover Data of China [J]. Journal of Geo-information science, 2010,12 (3): 385-391.

[5] Li Deren，Zhu Xinyan，Gong Jianya．From Digital Map to Spatial Information Multi-grid [C]．IEEE International Geoscience and Remote Sensing Symposium，Anchorage，USA，2004．

[6] Li Deren, Peng Mingjun. Transformation between urban spatial information irregular grid and regular grid [J]. Geomatics and Information Science of Wuhan University, 2007, 32 (2): 95－101.

[7] Foster I，Lman C，Tuecke S．The Anatomy of the Grid：Enabling Scalable Virtual Organizations [J]．International Journal Supercomputer Applications，2001，15 (3) ：200-222．

[8] Yang Cunjian, Zhang Zengxiang, Models of accuracy loss during rasterizing land use vector data with multi2scale grid size [J]. Geographical research, 2001，20 (4) ：416-422．

[9] Cheng Jiehai, Wei Fengyuan, Xue Huanzhu. Method for Spatial Data Scale Conversion [J]. Geospatial information, 2008，6 (4) ：13-15．

[10] Yang Cunjian, Liu Jiyuan, Zhang Zengxiang, et al, Analysis of Accuracy Loss During Rasterizing Vector Data with Different Grid Size [J]. Journal of mountain science, 2001,19 (3): 258-264.

[11] FGDC-STD-011-2001 United States National Grid [S]. 2001.

[12] GB/T 12409-2009 Geographic grid [S]. 2009.

[13] Ma Ting，Zhou Chenghu，Xie Yichun，et al．A discrete square global grid system based on the parallels plane projection [J]．International Journal of Geographical Information Science，2009，23 (10) ：1297-1313．

[14] Sahr K，Denis W，Kimerling J．Geodesic Discrete Global Grid Systems [J]．Cartography and Geographic Information Science，2003，30 (2) ：121-134．

[15] Ye Jing, Yang Xiaohuan, Jiang Dong, The grid scale effect analysis on town leveled population Statistical Data Spatialization [J]. Journal of geo-information science, 2010,12 (1): 2010，12 (1) ：40-47．