A novel interpolation square root model for multi-scale spatial database

Xianyong Gong¹*, Xiaojie Cui¹, Ruixing Xing¹, and Chengyi Liu¹
¹Information Engineering University, 450001 Zhengzhou Henan, China
* Corresponding author: gongxygis@whu.edu.cn

Abstract. The amount of map information is an important issue for multi-scale map making and spatial database producing in Cartography and Geographic Information System. Considering the multi-scale situation, a novel interpolation square root model for multi-scale spatial database is proposed. First, the limitations of the traditional square root model are analysed. Though some extensions were suggested, it is still incapable of guaranteeing the logic and quantity consistency in the multi-scale situation. Then, the improved novel interpolation square root model is proposed considering both the data at two-end neighbor source scales as control constraints. Finally, the experiments and comparisons on real multi-scale Topographic Database from the National Land Survey of Finland show that the proposed interpolation square root model is valid and outstanding in ensuring the logic and quantity consistency and continuity.

1. Introduction
Because the information capacity of the map is limited [1-4], it is impossible to completely represent all the geographic phenomena in the real world [2, 3]. The amount of map information means how many geographic objects to represent on the map. Therefore, it is an important issue for cartographers in multi-scale map making and spatial database producing in Cartography and Geographic Information System [1-4]. At present, commonly used models that indicate the relationship between map scale and the amount of map information include square root model, regression model, fractal model, and geometric progression model [2, 3]. Among them, the square root model is the most popular and widely accepted by the international cartography community [2, 4, 5].

2. Related works
This section will interpret the connotation of the square root model, including: why the map content or object number varying function with the map scales is satisfied with the square root model, what its basis is, and what it means.

The way multi-scale map making and spatial database producing can be divided into two categories [2].

● The first is the field surveying, where the coordinates are measured from the field measurement by surveying equipment, and then the map is formed by cartographers or computer-assisted professorial software. This kind of map making method brings out the large scale map with high accuracy, and generally has corresponding map specifications and standards.

● The second is map generalization, where small scale maps are derived and extracted based on the larger scale maps, together with spatial data integration and updating technologies.
Considering the role and mission, the map should represent the real geographic world as much as possible [2, 3]. However, due to the limitation of the map load, when the scale becomes smaller, the geographic objects on the source map cannot be represented as the infinitely same to the number on the derived map.

There is an empirical function or statistical relationship between the surveying location accuracy of the map data and the map scale, as defined below

\[ m = k \sqrt{M} \]  

(1)

Where \( m \) means the surveying location accuracy, \( M \) means the map scale, and \( k \) is an adjustment coefficient.

Considering this empirical statistical relationship, the famous cartographer scholar Töpffer pointed out that some principles are existed in multi-scale map making and spatial database producing [6]. By verifying a variety of multi-scale map productions, Töpffer holds the view that these principles can be expressed by the square root of the scales of the source and target map, hence the square root model comes. It is also termed as Töpffer radical law. Being simple and clear, the square root model is very easy to operate, and it better reflects the principle and relationship that how the map load and content vary with the map scales. Hence it has been widely accepted by the international cartography community. The square root model is defined as below.

\[ n_r = K \times n_s \times \left( \frac{m_s}{m_r} \right)^{1/p} \]  

(2)

where \( n_r \) is the map object number at the desired scale, \( m_s \) is the desired map scale, \( n_s \) is the object number at the source scale, \( m_r \) is the source map scale, \( K \) and \( p \) are adjustment coefficients. In normal square root model, \( K=1, p=2 \). Equation 2 is termed as traditional square root model (TSRM) in this paper.

The TSRM has been extended in different applications. Some extensions adjust the different coefficients \( K \) and \( p \) to fit the global scale range, while others adjust them by different scale range. Considering the influence of symbol size and the importance of object, Wang (2015) customized different \( K \) and \( p \) according to different features and scale ranges [3], with the aim to determine the selection number of polyline and polygon features in the scale from 1:25K to 1:4000K. When \( p=2 \), Wu (2012) pointed out that it is applicable only in the medium-scale range (1:250K-1:500K) [2]. However, the scale is not the only factor that determines the amount of map information. In [7, 8], the applicable scale range of the TSRM is jointly determined by cartographic region, geographic features, and map scale, suggesting that it is not limited to the medium-scales. Considering different geographic objects, geographic features, cartographic regions, and map load, the TSRM is modified in [7-9], so that it can be applied to the desired map generalization task.

However, all these extensions are incapable of guaranteeing the logic and quantity consistency in the multi-scale situation.

3. A novel interpolation square root model

Up to now, many countries have established the national fundamental multi-scale spatial database, such as the United States, the United Kingdom, Germany, France, and China. Deriving and extracting arbitrary scale spatial data products become one of the most pressing needs of map users. Note the national fundamental multi-scale spatial database as \{\( m_1, m_2, \ldots, m_q, \ldots, m_Q \}\), where \( m_q \) is the scale of the database. \( m_q \) and \( m_{q+1} \) are termed as neighbor scales, \( m_q \) is named as source scale and \( m_{q+1} \) is named as the next neighbor source scale.

The object number \( n \) in the desired arbitrary scale \( m_r \) (\( m_q \leq m_r \leq m_{q+1} \)) should be smoothly and continuously transformed from \( m_q \) to \( m_{q+1} \), which act as one of most challenge technologies in deriving arbitrary scale spatial data. The smooth and continuousness of object number between neighbor scale means that (1) when \( m_r=m_q, n_r=n_q \), and (2) \( m_r=m_{q+1}, n_r=n_{q+1} \).

However, although many improvements have been conducted on the TSRM, they only focus on the adjustment coefficients or application scale range. \( n_r \) is calculated based on \( n_s \) without the
consideration of $n_r$. When referring to $n_r$, the $n_r$ produced by TSRM may be more or less than $n_r$ when $m_s=m_r$, bringing information inconsistency and discontinuity. When the desired $n_r$ is less than $n_r$, the information on $m_r$ will be incomplete, leading the contradictory fact that some objects are selected on the smaller scale map but are deleted in the larger middle scale map generated by algorithms. When the desired $n_r$ is more than $n_r$, it may bring spatial conflicts, affect the quality of map publishing and reading, fail to achieve the abstraction nature of map, and increase the information discontinuity between the intermediate (larger) scale map $m_r$ generated by algorithms and the actual target scale map $m_r$.

In view of this, this paper proposes an interpolation square root model (ISRM) by using the source scale and the next neighbor base source scale as known two-end controlling scales. Referring to the TSRM, the ISRM proposed in this paper is shown in Equation 3.

$$n_r = n_s \times \left( \frac{m_s}{m_r} \right)^{\frac{1}{2}} \times \left( \frac{m_r - m_s}{m_r} \right) + n_s \times \left( \frac{m_s}{m_r} \right)^{\frac{1}{2}} \times \left( \frac{m_r - m_s}{m_r} \right)$$

When there is no next neighbor source scale as controlling, $m_s$ is considered infinity, and $n_r$ is infinitely close to 0. Equation 3 is transformed to Equation 2, indicating that ISRM is equivalent to TSRM. This indicates that the ISRM is more universal, and the TSRM is a special case of ISRM.

It can be proved mathematically that the ISRM meets the following two conditions:
- $n_r$ in Equation 3 is a monotonically decreasing function on $m_r$. It can be known from the common sense that the objects number $n_r$ on the map is decreasing as the scale $m_r$ is reduced.
- The range of $n_r$ is between $n_s$ and $n_r$, that is, when $m_s \in [m_s, m_r]$, $n_s \in [n_r, n_r]$, which will ensure the continuity of the map object number on the interpolation scale map, and avoid logic and number inconsistency caused the TSRM.

Therefore, the ISRM can controlling the selection quantity of the object number between the two controlling scale maps, ensuring that the selection quantity will never be more or less than the controlling scales, thereby avoiding the unreasonable phenomenon and contradictory fact that the selection number by TSRM is greater than or less than the object number on the next neighbor source scale in multi-scale spatial data producing.

4. Experiment and analysis

The experiments are divided into two groups. The first is to verify the difference between the ISRM and TSRM. It is intended to prove that the ISRM is more suitable for multi-scale spatial database. The second is to investigate the effect of the index $p$.

To verify the scientificity, the real multi-scale Topographic Database from the National Land Survey of Finland (delivered 06/2014) is used. The scales are 1:1000K, 1:2000K, 1:4500K, and 1:8000K. The technical way of this multi-scale topographic database producing is map generalization, where the database at larger scales is generalized to generate the smaller scale data. For example, the topographic database at 1:2000K is obtained by generalizing the data at 1:1000K.

Since there are few polygon features at these scales, this paper selects point and linear features as experimental objects, including three representative geography elements, namely settlement, road network and place name. Table 1 is the statistics of the object number. By calculating Equation 2, we can find that the map objects number between different scales does not strictly conform to TSRM.

| Table 1. Statistics of the multi-scale Topographic Database. |
|---------------------------------------------------------------|
| object number       | 1:1000K | 1:2000K | 1:4500K | 1:8000K |
| settlement           | 1711    | 454     | 143     | 19      |
| road                 | 14906   | 1656    | 925     | 603     |
| place name           | 2576    | 668     | 227     | 32      |
Table 2. Statistics of the object number at interpolation scales by different methods.

| $m$ | $n_1$ | $n_2$ | $n_3$ | $n_4$ | settlement | $n_1$ | $n_2$ | $n_3$ | $n_4$ | road | $n_1$ | $n_2$ | $n_3$ | $n_4$ | place name | $n_1$ | $n_2$ | $n_3$ | $n_4$ |
|-----|-------|-------|-------|-------|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|-------|-------|-------|-------|
| 100 | 1711  | 1711  | 1711  | 1711  | 14906    | 14906 | 14906 | 14906 | 2576  | 2576  | 2576  | 2576  |
| 125 | 1530  | 1291  | 1287  | 1116  | 13332    | 10522 | 10899 | 7347  | 2304  | 1939  | 1935  | 1668  |
| 150 | 1397  | 960   | 955   | 787   | 12170    | 7041  | 7464  | 4122  | 2103  | 1437  | 1433  | 1169  |
| 175 | 1293  | 687   | 684   | 586   | 11267    | 4144  | 4418  | 2528  | 1947  | 1022  | 1019  | 866   |
| 200 | 1209  | 454   | 454   | 454   | 10540    | 1656  | 1656  | 1655  | 1821  | 668   | 668   | 668   |
| 225 | 428   | 405   | 399   | 383   | 1561     | 1535  | 1507  | 1521  | 629   | 598   | 588   | 571   |
| 250 | 406   | 363   | 353   | 330   | 1481     | 1433  | 1390  | 1410  | 597   | 538   | 522   | 496   |
| 275 | 387   | 325   | 314   | 288   | 1412     | 1343  | 1294  | 1317  | 569   | 485   | 466   | 437   |
| 300 | 370   | 292   | 280   | 254   | 1352     | 1264  | 1215  | 1237  | 545   | 438   | 418   | 389   |
| 325 | 356   | 262   | 251   | 227   | 1299     | 1193  | 1148  | 1168  | 524   | 395   | 376   | 350   |
| 350 | 343   | 234   | 224   | 204   | 1251     | 1130  | 1091  | 1107  | 504   | 356   | 339   | 317   |
| 375 | 331   | 209   | 201   | 185   | 1209     | 1072  | 1041  | 1054  | 487   | 320   | 307   | 289   |
| 400 | 321   | 185   | 180   | 169   | 1170     | 1019  | 998   | 1006  | 472   | 287   | 277   | 265   |
| 425 | 311   | 163   | 160   | 155   | 1136     | 970   | 959   | 963   | 458   | 256   | 251   | 244   |
| 450 | 302   | 143   | 143   | 143   | 1104     | 925   | 925   | 925   | 445   | 227   | 227   | 227   |
| 475 | 139   | 131   | 132   | 118   | 900      | 891   | 885   | 888   | 220   | 208   | 210   | 188   |
| 500 | 135   | 119   | 122   | 98    | 877      | 861   | 850   | 855   | 215   | 190   | 193   | 158   |
| 525 | 132   | 109   | 112   | 83    | 856      | 832   | 818   | 824   | 210   | 173   | 178   | 134   |
| 550 | 129   | 98    | 102   | 70    | 836      | 805   | 789   | 796   | 205   | 157   | 163   | 114   |
| 575 | 126   | 89    | 93    | 60    | 818      | 780   | 763   | 770   | 200   | 142   | 148   | 98    |
| 600 | 123   | 80    | 84    | 52    | 801      | 756   | 739   | 746   | 196   | 128   | 134   | 85    |
| 625 | 121   | 71    | 75    | 45    | 784      | 733   | 717   | 724   | 192   | 114   | 120   | 74    |
| 650 | 118   | 63    | 66    | 39    | 769      | 712   | 697   | 703   | 188   | 101   | 106   | 64    |
| 675 | 116   | 54    | 58    | 34    | 755      | 691   | 678   | 684   | 185   | 88    | 93    | 57    |
| 700 | 114   | 47    | 50    | 30    | 741      | 672   | 661   | 665   | 182   | 76    | 80    | 50    |
| 725 | 112   | 39    | 42    | 26    | 728      | 653   | 645   | 648   | 178   | 64    | 68    | 44    |
| 750 | 110   | 32    | 34    | 23    | 716      | 636   | 630   | 632   | 175   | 53    | 55    | 39    |
| 775 | 108   | 25    | 26    | 21    | 704      | 619   | 616   | 617   | 172   | 42    | 43    | 35    |
| 800 | 107   | 19    | 19    | 18    | 693      | 603   | 603   | 603   | 170   | 32    | 32    | 32    |
| 800 | 19    | 19    | 19    | 19    | 603      | 603   | 603   | 603   | 32    | 32    | 32    | 32    |

4.1. Case 1

To further verify the superiority of the proposed ISRM, Jiang et al. (2010) is used as comparisons [7], noted as Jiang Model (JM). The JM does not need parameters. In the TSRM and ISRM, $K=1, p=2$.

The scale interpolation of the multi-scale Topographic Database is conducted with an equal intervals of 250K. The object numbers at each interpolated scale are shown in Table 2. $n_1$ denotes the TSRM with $p=2$. $n_2$ denotes the proposed ISRM with $p=2$. $n_3$ denotes the ISRM where $p$ is adaptive by JM. $n_4$ denotes JM where $p$ is adaptive.

The comparison of $n_3$ and $n_4$ aims to validate the availability of the ISRM and JM with the same $p$. The $n$ and $p$ in JM are calculated as following.

$$n_p = n_3 \times \left( \frac{m}{m_p} \right)^{\frac{1}{p}}$$  (4)
The bold rows in Table 2 are the actual object number of the existing multi-scale data, and others are the data calculated by different methods. Due to space limitations, only the plot graph of $n_1$ and $n_2$ are shown in Fig. 1. The following conclusions can be drawn.

1) For the object number, on the same scale, ISRM is similar to JM. The number of TSRM is more than ISRM and JM, as show in Table 2 and Fig. 1. For example, when the interpolation scale is 1:2000K, object numbers calculated by the three methods are 10540, 1656, and 1655, respectively. The result 10540 by TSRM has clearly exceeded the actual number 1656 on real map, which is obviously unreasonable. The result 1655 by JM are slightly different from the actual number 1656, while our ISRM is strictly consistent with the real data.

2) For the continuity of the object number at both controlling ends, the TSRM and JM are both of discontinuity. Since the TSRM does not consider the the next neighbor scale constraint, its discontinuity is much worse. The JM adopts the exponential function, even the upper bound is taken, there is still a small discontinuity. As shown underlined in Table 2, the roads number at the interpolated scale of 1:2000K is 1655, while the actual roads number is 1656; the settlements number at the interpolated scale of 1:8000K is 18, and the actual number is 19. However, the advantage of ISRM is that the logic and quantity consistency and continuity can be strictly guaranteed.

3) It can be seen from Fig. 1 that the ISRM is not only applicable to the scale range (more than 1:2000K) with a large map content change rate, but also the scale range (less than 1:2000K) with a small rate. That is to say, the proposed ISRM is more robust to scale range.

\[
\frac{1}{p} = \frac{\log n_i}{\log m_i}
\]
To further explore the relationship between the index $p$ and $n_r$, this experiment changes the index $p$ from 1 to 3 with a equal intervals of 0.2. The index $p$ in JM is automatically determined. As shown in Equation 5, if $p$ is fixed, it becomes TSRM, so it does not participate this experiment. However, the $p$ in ISRM and TSRM can be either manually set or automatically calculated by JM. Therefore, this experiment compares the influence of the index $p$ on ISRM and TSRM. Experiments are still performed using the data in Table 1. Due to space limitations, only the plot graphs of the interpolation results are shown in Fig. 2, where the dotted and solid lines are the results of TSRM and ISRM, respectively. The following conclusions can be drawn.
1) $p$ will affect the selection object number $n_r$. The larger $p$ is, the less $n_r$ is.
2) $p$ has a great influence on the TSRM, while has little effect on the proposed ISRM.
3) Regardless of the value of $p$, the proposed ISRM can guarantee the continuity of the selection object number on the interpolation scale.
4) Either the scale range where the map content changes rapidly or the scale range where the map content changes smoothly, the above three properties hold true. It shows that the proposed ISRM is more robust to scale range.

The above conclusions further prove that the proposed ISRM with strong robustness is not affected by both $p$ and scale range. The index $p$ has little effect on ISRM and does not affect the continuity of the selection number on the interpolation scale.

5. Conclusions
The advantages of ISRM include (1) taking the two-end neighbor source scales as control constraints; (2) being more robust to scale range; (3) less effectness by the exponential index $p$; and (4) ensuring the logic and quantity consistency and continuity in the multi-scale situation.

The mathematical theory derivation and the comparative experiments analysis collectively prove that the proposed ISRM can strictly guarantee the selection object number consistency in the multi-scale Topographic Database. Therefore, the ISRM is more suitable for the quantity determination in deriving and extracting arbitrary scale spatial data and multi-scale Topographic Database.
Acknowledgements
This study is supported by the National Natural Science Foundation of China (Grant No. 41471386 and 41801396).

References
[1] Z. Li. Algorithmic foundation of multi-scale spatial representation (Taylor & Francis, Bacon Raton, 2007)
[2] H. Wu. Basic models and algorithms of GIS geospatial generalization (Wuhan Uni. Press, Wuhan, 2012)
[3] J. Wang. MapMaking(Star map Press, Beijing,2015)
[4] F. Wu, X. Gong, J. Du. Overview of the research progress in automated map generalization, Acta Geodaetica et Cartographica Sinica, 46,1645-1664(2017)
[5] X. Gong, F. Wu. A typification method for linear pattern in urban building generalization. Geocarto Int. 33,189-207(2018)
[6] F. Töpfer, W. Pillewizer. The principles of selection, Cartogr J.3,10-16(1966)
[7] N. Jiang, X. Bai, Y. Cao, J.Wang. Establishment and application of multi-scale display model in basic electronic map, Geomat. Inf. Sci. Wuhan Univ, 35, 768-772(2010).
[8] N. Jiang, Y. Cao, J. Zhao, M. An. Foundation and application of electronic map multi-scale display models of different density areas, Geomat. Inf. Sci. Wuhan Univ, 38,465-469(2013)
[9] Q. Zhang, F. Wu, H. Qian, et al. Milestone Scales Oriented Spatial Data Multi-Representation Techniques, J. Geomat. Sci. Technol. 28, 383-386 (2011)