Errors and Their Scale Effect for Spatialization of Air Temperature Data

Shunbao Liao[1,2], Sai Zhang[2]
[1] College of Environment and Planning, Henan University, Kaifeng 475004, China
[2] Institute of Geographic Sciences and Natural Resources Research (IGSNRR), Chinese Academy of Sciences, Beijing 100101, China

***Abstract*—Spatialization of attribute data is a way to output grid data products from vector data. It is beneficial to integrated analysis of geosciences data from various sources and in different formats. However it is also a process companied with errors, and the errors are closely related to density of data sources, spatializing models and resolution of grid cells. In this paper, seven levels of density of meteorological stations, five spatializing models and nineteen levels of resolutions of grid cells were used to analyze the relationships between the errors for spatialization of air temperature and these factors. It was found that reduction of density of meteorological stations led to increasing of errors. Of the five models, Adjusted IDW, Regression and ANUSPIN had higher accuracy than IDW and Kriging. And the accuracy generally decreases with increasing of size of grid cells. Of the three factors affecting accuracy of spatialization, the models had the greatest impact on the accuracy, the resolution of grid cells second and the density of meteorological stations the lowest.**

Keywords-air temperature; spatialization; errors; scale effect

I. INTRODUCTION

Attribute data, vector data and raster data are three basic data types in geosciences. With development and application of technologies of remote sensing, geographical information system (GIS), data integration and data confusion in the field of geosciences, conversions between various data types have become routine activities for geosciences data processing. It has become more and more often and necessary for scientists to produce new data sets in format of raster using observed data from observatories or statistical data based on administrative divisions, which are in format of vector. This process is called Spatialization of Attribute Data (SOAD). SOAD was defined as a process through which the attribute data for point, linear or polygonal objects, such as precipitation from meteorological stations or population at county level, were converted to regular grid cells (for example one kilometer by one kilometer) from tabular structure in light of relevant models or formulas [1].

SOAD are mainly employed to meet the demands of multiple-type data based integrated analysis, interpolation of data in places without observatories and enhance of spatial resolutions of data in the field of geosciences. GIS based spatial analysis and modeling often use data in grid format [2-3]. Landscape, regional and global ecosystem models for global change, for example MT-CLIM [4] and FOREST-BGC [5] need spatialized air temperature, precipitation and solar radiation as input parameters. In China, there are more than two thousand meteorological stations. However, only data from more than six hundred stations can be shared in China and those from about two hundred stations can be exchanged internationally. Most of data from other stations cannot be shared. This also raises the demand of spatialization of observed data based on sites.

Over the last decade, many national and global grid meteorological databases were established unceasingly and related computer software was developed. The main grid meteorological databases include PRISM based spatial meteorological data information systems for United States, Canada, China, Mongolia and Europe [6], ANUSPLIN based systems for Australia and South Africa [7-8], and DAYMET based United States biological meteorological data information system [9]. In the meantime, some regional grid climate data sets at various resolutions were developed one after another. They include biological climate data sets at 30 meters resolution for Catalonian, Spain [10], data sets for Karnataka, India, data sets for VIC model in United States and Canada, and data sets for Vegetation-Ecosystem Modeling and Analysis Project [11].

In China, studies on spatialization of climate data have also been carried out extensively. Liao *et al.* spatialized 30-year mean air temperature (1951-1980) by means of multi-dimension regression plus interpolation of residual [12], made comparisons between different interpolation methods including IDW, Kriging, Spline and Trend [13], and discussed zoning on spatialization of accumulated air temperature [14]. Cai [15] and Peng [16] also conducted researches on spatialization of air temperature in nation-wide and Xinjiang respectively. Yu *et al.*, Liu *et al.* and He *et al.* have made intensive researches into spatialization of air temperature, precipitation and solar radiation [17-19].

Spatialization of air temperature is a process companied with errors, and value of the errors depends on density of observatories, spatializing methods and size of grid cells. Different density of observatories, spatializing methods or size of grid cells result in different error values. Therefore, it is significant to study scale effect of errors for spatialization to improve accuracy of spatialization. So far little study on scale effect of errors for spatialization has been carried out. However, some methodologies used in study on scale effect of errors for rasterization of land use data can be referenced [20-21].

II. DATA USED IN THIS STUDY

The data used in this study included:

(a) 30-year average air temperature data (1971-2000) from 698 meteorological stations in mainland of China, which were obtained from China Meteorological Administration.

(b) Air temperature data of Hong Kong and Macao (1971-2000), and those from 98 meteorological stations in Taiwan province and the countries surrounding China (1961-1990), which came from Hong Kong Observatory, http://www.hko.gov.hk/wxinfo/climat/world/chi/asia/asia_c.htm.

(c) DEM of China at 30-second resolution, which came from Earth Resources Observation and Science (EROS) Center, http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html.

(d) Administrative division map of China at scale of 1 to 4,000,000, which came from State Bureau of Surveying and Mapping of China, http://sms.webmap.cn/.

(e) Air temperature data from 56 meteorological stations which distribute evenly in nation-wide. The data were used to assess errors of spatialization and therefore were not used in process of spatialization.

Fig. 1 is distribution map of meteorological stations in China and the areas surrounding it.

III. DATA ANALYSIS

A. *Settings for Analysis*

1) Number of meteorological stations

The number of meteorological stations was set to seven levels in light of principle of evenly decreasing to analyze influence of density of meteorological stations on spatialization accuracy. They were level 1 with 743 stations, level 2 with 593, level 3 with 493, level 4 with 393, level 5 with 293, level 6 with 193 and level 7 with 93.

2) Size of grid cells

The size of grid cells was set to nineteen levels to analyze influence of size of grid cells on spatialization accuracy. They were level 1 at a resolution of 1km by 1km, level 2 at 2km by 2km, level 3 at 3km by 3km, level 4 at 4km by 4km, level 5 at 5km by 5km, level 6 at 6km by 6km, level 7 at 7km by 7km, level 8 at 8km by 8km, level 9 at 9km by 9km, level 10 at 10km by 10km, level 11 at 20km by 20km, level 12 at 30km by 30km, level 13 at 40km by 40km, level 14 at 50km by 50km, level 15 at 60km by 60km, level 16 at 70km by 70km, level 17 at 80km by 80km, level 18 at 90km by 90km and level 19 at 100km by 100km.

3) Selection of spatializing methods

Five methods were selected to spatialize air temperature data to analyze influence of the methods on spatialization accuracy. They were IDW, Kriging, Adjusted IDW, Regression and ANUSPLIN.

B. *Results from Analysis*

1) The relation between errors and density of meteorological stations

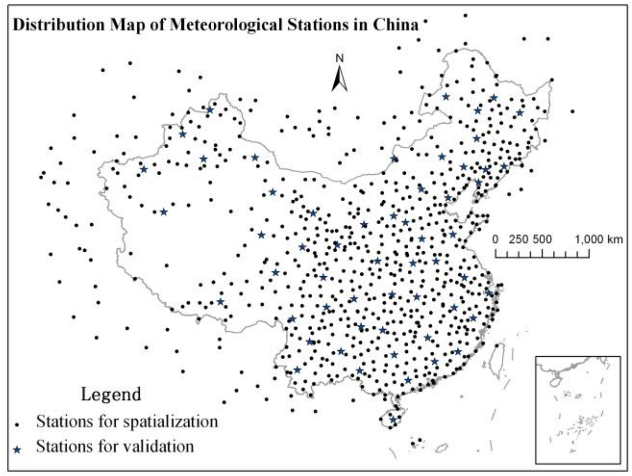

Figure 1. Distribution of meteorological stations in China and its surrounding areas

In this study, average absolute errors (AAE) based on 56 verification stations were taken as accuracy assessment index. Data from the verification stations were not used for spatialization to enhance credibility of verification. For each level of number of meteorological stations, each level of size of grid cells and each spatializing method, an AAE could be got. Fig. 2 shows the relationship between the AAEs and densities of meteorological stations in the case of the grid resolution of 1km by 1km.

It can be seen from Fig. 2 that AAEs increase generally while the number of meteorological stations used for spatialization decreases though the increasing trend is not significant.

2) The relation between errors and resolutions of grid cells

Fig. 3 shows the variation of AAEs with resolutions of grid cells for five methods in the case that data from 743 meteorological stations were used for spatialization.

It shows that variation of resolution of grid cells did not lead to significant change of AAEs for IDW and Kriging, which means resolution of grid cells did not have obvious affection for accuracy of spatialization for the two methods. However, the errors increased obviously when size of grid cells became larger and larger for Adjusted IDW, Regression and ANUSPLIN. For example the errors were larger than 1℃ when the resolution of grid cells exceeded 20 km.

3) Quantitative relation between errors, resolution of grid cells and density of observatories

It could be seen from Fig. 2 and Fig. 3 that Adjusted IDW had least errors in five interpolation methods. This paper focused on errors and their scale effect. So we only chose Adjusted IDW to calculate AAEs in all cases of different size of grid cells and different number of meteorological stations. The results were shown in Table 1.

Binary linear regression was conducted with AAE as dependent variable, number of meteorological stations and size of grid cells as independent variables. A linear regression equation was established as (1).

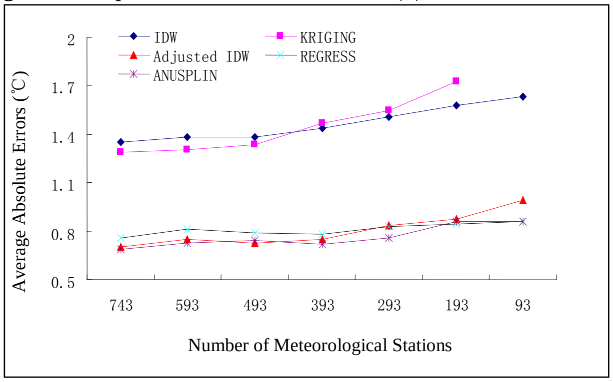

Figure 2. The relationship between AAEs and number of meteorological stations

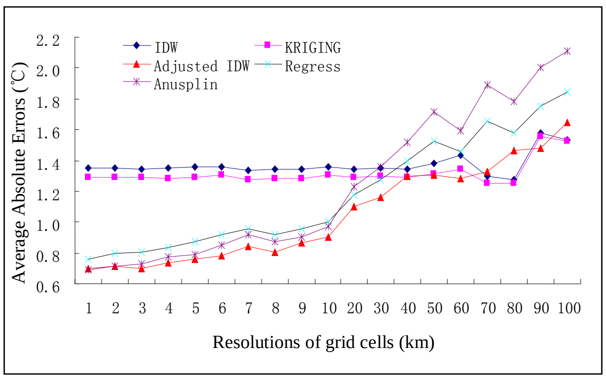

Figure 3. The variation of AAEs with resolutions of grid cells

$$y' = 1.096 - 0.00046x_1 + 0.008038x_2 \qquad (1)$$

Where x_1 represents number of meteorological stations, x_2 represents size of grid cells (unit: km), and y' represents AAE(unit: ℃).

When predictive errors by the regression equation and actual errors were fitted, (2) was drawn.

$$y = 1.23\ln(y') + 1.01 \qquad (2)$$

Where y' represents predictive AAE(unit: ℃) and y represents actual AAE(unit: ℃), $ln()$ represents natural logarithm. The relation between y *and* y' can be shown in Fig. 4.

TABLE 1. AVERAGE ABSOLUTE ERRORS FOR DIFFERENT SIZE OF GRID CELLS AND DIFFERENT NUMBER OF METEOROLOGICAL STATIONS

Average absolute errors(℃)		Number of meteorological stations used for spatialization						
		743	***593***	***493***	***393***	***293***	***193***	***93***
Size of	***1***	0.70	0.75	0.73	0.75	0.83	0.88	0.99
	2	0.71	0.77	0.75	0.78	0.87	0.91	1.05
	3	0.70	0.76	0.73	0.79	0.86	0.89	1.03
	4	0.74	0.80	0.77	0.81	0.91	0.97	1.12
	5	0.76	0.81	0.78	0.86	0.93	1.00	1.17
	6	0.78	0.86	0.82	0.91	0.99	1.04	1.19
	7	0.85	0.94	0.91	0.99	1.06	1.09	1.23
	8	0.81	0.90	0.88	0.94	1.06	1.11	1.26
	9	0.86	0.94	0.90	0.96	1.04	1.09	1.21
	10	0.91	1.00	0.96	1.04	1.14	1.17	1.30
	20	1.10	1.19	1.13	1.20	1.24	1.27	1.46
	30	1.16	1.23	1.20	1.22	1.24	1.28	1.38
	40	1.29	1.41	1.36	1.38	1.45	1.50	1.64
	50	1.31	1.38	1.30	1.31	1.28	1.38	1.59
	60	1.28	1.41	1.35	1.37	1.44	1.50	1.59
	70	1.33	1.40	1.39	1.40	1.45	1.47	1.63
	80	1.47	1.48	1.45	1.51	1.64	1.70	1.80
	90	1.48	1.51	1.46	1.49	1.54	1.55	1.72
	100	1.65	1.60	1.59	1.52	1.53	1.64	1.95

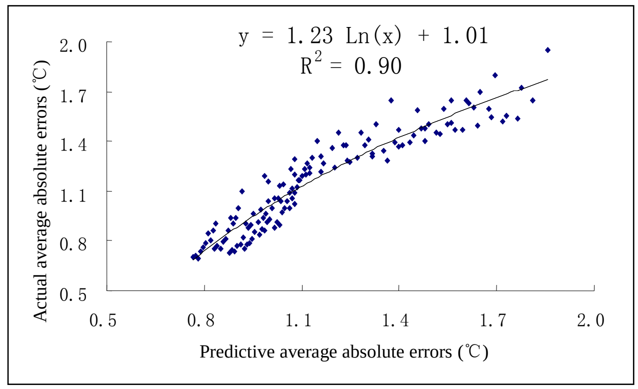

Figure 4. The relationship between predictive and actual AAEs

We could derive (3) by combining (1) and (2). It is a quantitative expression for the relation between errors, resolution of grid cells and density of meteorological stations.

$$y = 1.23\ln(1.096 - 0.00046x_1 + 0.008038x_2) + 1.01 \qquad (3)$$

Where y, x_1, x_2 and $ln()$ represent the same meanings as what they do in (1) and (2).

A conclusion could be drawn by comparison of coefficients of x_1 and x_2 that an error for spatialization of air temperature was more sensitive to size of grid cells than to number of meteorological stations through it was affected by both of them.

IV. CONCLUSIONS

In this paper, seven levels of densities of meteorological stations distribution, five spatialization models and nineteen levels of resolutions of grid cells were used to analyze the relations between errors resulting from spatialization and them. The following conclusions were drawn.

(a) Density of meteorological stations had affection on spatialization accuracy to a certain extent. Reduction of density of meteorological stations led to increasing of errors, but the trend of increasing was not significant.

(b) Models had significant influence on spatialization accuracy. The models in the study were classified into two groups in light of errors and their change trend. Adjusted IDW, Regression and ANUSPIN had higher accuracy than IDW and Kriging for the factors related to air temperature, for example altitude, were considered in the formers.

(c) The resolution of grid cells also affected spatialization accuracy. Generally the accuracy decreased with increasing of size of grid cells.

(d) Of the three factors mentioned above, the models had the greatest impact on the accuracy, the resolution of grid cells second and the density of meteorological stations the lowest.

(e) Adjusted IDW, Regression or ANUSPLIN method should be used and the resolution of grid cells should not exceed ten kilometers in order to assure AAE of spatialized air temperature products lower than 1℃ in nation-wide of China.

ACKNOWLEDGMENT

This research was supported by the State Key Laboratory of Resources and Environment Information System (Grant O88RA106SA) and the Institute of Geographic Sciences and Natural Resources Research (Grant O66U0309SZ) of the Chinese Academy of Sciences. Their supports are gratefully acknowledged.

REFERENCES

[1] Liao Shunbao and Zhang Sai, “Study on error evaluating index for spatialisation of attribute data,” Journal of Geo-Information Science, vol. 11, 2009, pp.176-182.

[2] Michael N. Demers. “GIS Modeling in Raster,” New York: John Wiley & Sons, Inc, 2002.

[3] Bates P. D, Deroo A. P. J., “A simple raster-based model for flood inundation simulation,” Journal of Hydrology, vol. 236, pp. 54-77.

[4] Running S. W., “Extrapolation of synoptic meteorological data in mountainous terrain and its use in simulating forest evapo-transpiration rate and photo synthese,” Canadian Journal of Forest Research, vol.17, 1987, pp. 472-483.

[5] Band L. E., “Forest ecosystem processes at the watershed scale: basis for distributed simulation,” Ecological Modeling, vol. 56, 1991, pp.171-196.

[6] Daly C., G. H. Taylor and W. P. Gibson, “The PRISM approach to mapping precipitation and temperature,” In: Amer. Meteor. Soc. Proc. 10th AMS Conf. on Applied Climatolog Meteorogical Soc., Reno, NV. Boston, Mass: Amer. Meteor. Soc., 1997, pp.10-12.

[7] Hutchinson M. F., “Interpolating mean rainfall thin plate smoothing splines,” INT J Geographical Information Systems, vol. 9, 1995, pp. 385-403.

[8] Stephen J. Jeffrey and John O. Cater, “Using spatial interpolation to construct a comprehensive archive of Australian climate data,” Environmental Modeling & Software, vol. 16, 2001, pp.309-330.

[9] Peter E. Thornton and Steven W. Running, “Generating surfaces of daily meteorological variables over large regions of complex terrain,” Journal of Hydrology, vol. 190, 1997, pp. 214-251.

[10] Miquel Ninyrola, Xavier Pons and Joan M Rouge, “A methodological approach of climatological modeling of air temperature and precipitation through GIS techniques,” Int. J. Climatol., vol. 20, 2000, pp.1823-1841.

[11] VEMAP Members, “Vegetation/ecosystem modeling and analysis project (VEMAP): comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem responses to climate changes and CO2 doubling,” Global Biogeochem. Cycles, vol. 9, 1995, pp. 407-437.

[12] Liao Shunbao and Li Zehui, “A methodology of spatialization of observed data based on GIS,” Progress In Geography, vol. 22, 2003, pp.87-93

[13] Liao Shunbao and Li Zehui, “Some practical problems related to rasterization of air temperature,” Meteorological Science and Technology, vol. 32, 2004, pp.352-356

[14] Liao Shunbao and Li Zehui, “Study on methodology for rasterizing accumulated temperature data,” Geographical Research, vol. 23, 2004, pp.633-640

[15] Cai Fu, Yu Guirui, Zhu Qinglin, He honglin, Liu Xinan, et al., “Comparison of precisions between spatial methods of climatic factors: A case study on mean air temperature,” Resources Science, vol. 27, 2005, pp.173-179.

[16] Peng Dailiang, Huang Jingfeng, Cai Chengxia and Deng Rui, “Spatialization on monthly average air temperature in Xinjiang and the analysis of the result,” Journal of Zhejiang University (Science Edition), vol. 34, 2007, pp.580-584.

[17] Yu Guirui, He Honglin, Liu Xinan and Niu Dong, “Study on spatialization technology of terrestrial eco-information in China(I): The approach of spatialization in meteorology/climate information,” Journal of Natural Resources, vol. 19, 2004, pp. 537-544.

[18] Liu Xinan, Yu Guirui, Fan Liaosheng, Li Zhengquan, He Honglin, et al., “Study on spatialization technology of terrestrial eco-information in China(III): Temperature and precipitation,” Journal of Natural Resources, vol. 19, 2004, pp. 818-825

[19] He Honglin, Yu Guirui, Liu Xinan, Su Wen, Niu Dong, et al., “Study on spatialization technology of terrestrial eco-information in China(II): Solar radiation. Journal of Natural Resources,” vol. 19, 2004, pp. 679-687.

[20] Yang Cunjian and Zhang Zengxiang, “Models of accuracy loss during rasterizing landuse vector data with multi scale grid size,” Geographical Research, vol. 20, 2001, pp. 416-422.

[21] Zhang Xinchang, Huang Qiuhua and Yang Jian, “Accuracy loss model of land-use data on different scales,” Acta Scientiarum Naturalium Universitatis Sunyatseni, vol. 46, 2007,pp.103-106.