Spatial Interpolation of meteorology monitoring data for Western China using Back-Propagation Artificial Neural Networks

Yaonan Zhang
Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China
yaonan@lzb.ac.cn

Guohui Zhao
Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China
zhgh@lzb.ac.cn

Yang Wang
Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China
wangyang@lzb.ac.cn

***Abstract*– Spatial interpolation algorithms are vital to environmental monitoring systems, especially for the real-time monitoring systems of critical variables in converting the point measurements to spatial continuous surfaces. This paper describes the spatial interpolation of meteorological observations (air temperature as an example) using a feed-forward back-propagation neural network based on the environment-affecting factors. These model independent estimators were (1) meteorological stations' longitude, latitude, altitude; (2) Normalized Difference Vegetation Index; (3) slope and aspect. This is a first to consider all the factors for are temperature spatial interpolation when interpolating using a neural network. Especially the study area covers large region of complex terrain, which includes only 241 national meteorological stations over almost half-total area of China. However, the simulated results show that the model could provide reliable spatial estimations of monthly mean air temperature. Goodness of fit of model was very high (R>0.95) and efficient.**

***Keywords*: Spatial interpolation; BP Neural Network; Western China; Air Temperature.**

I. INTRODUCTION

Spatial continuous dataset converted from local observations play a significant role in planning, risk assessment and decision making in environmental management. The high-resolution spatial dataset is also the foundation for explanation the changing trends of the whole region and modeling environmental future. Automating preparation the data set from the sparse ground-based networks sites is very useful for early warning, disaster emergency and scientific research. In case of hazards and emergencies (e.g. earthquake, fire, flash floods, earthquakes, etc.), maps need to be automatically generated in real-time with minimum human intervention. The timely information is important for an efficient management of environmental risks. Today more and more research on geoscience requires spatial continuous and to get intuitionistic information and further modeling. However, spatial distribution variables of natural phenomena are gathered from point sources. Therefore, spatial interpolation techniques are essential for estimating observed variable of the un-sampled locations.

Because there is nothing such as a universal spatio-temporal mapping algorithm, the tools currently used for spatial data analysis either provide users with very simple mapping functions that allow quick processing of the data, or, what appears to become a new trend, allow them to interact with decision trees that direct them towards more advanced functions. These wizards usually require that the users have some prior knowledge and experience in geostatistics [1]. Especially the basic statistical methodologies are the foundation for most climatological interpolations [2]. Traditional approaches have ranged from relative simple inverse distance weighting (IDW) to more complex techniques like kriging, splines [3]. However, these methods generate maps just based on the statistics information. They are data-specific or even variable-specific. Many factors including sample size, sampling design and data properties affect the spatial estimations of these methods.

The recent development of artificial neural network (ANN) as one universal estimator is very good at detecting and representing nonlinear relationships between dependent and independent variables. ANN is a mathematical model or computational model that tries to simulate the structure and functional aspects of biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases, an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Modern neural networks are non-linear statistical data modeling tools. They are usually used to model complex relationships between inputs and outputs or to find patterns in dataset. In addition, in many practical applications, ANNs have made significant achievements including atmospheric time series data [4, 5].

The aim of this paper is to explore an efficient method for spatial interpolation or disaggregation especially for the stations network Automating mapping. We focus on a spatial interpolation mode based on artificial neural network as a highly adaptive nonlinear method for performing spatial estimation of environmental observation variable and in consideration of external factor influence for sparse sites area. It provides

II. MATERIAL AND METHOD

A. Study Area and Data sources

The study area locates western China (31°N-54°N, 73°E-126°E) and composes of six provinces and autonomous regions. The dimension has an area of 4.26 million square kilometers almost equal to half of total area of China. It covers a complex terrain and elevation ranges from -153m to 7615m. The climates varying from warm temperate to cold temperate and from High Mountain to desert. There are many important geography units distributed in it including glacier, desert, oasis, inland lake and inland rivers et al.

Heihe Basin is one of the three inland rivers within this study area, which covers 13×10^4 km^2, as shown in Fig.1. Heihe River originates from the Qilian Mountains of Qinghai Province; flows through the middle basin, called the Hexi Corridor of Gansu Province; reaches the lower reaches in the Inner Mongolia Autonomous Region. There are 13 national meteorological stations within Heihe watershed. Therefor it is a challenge to interpolate the observed data into continuous surfers by usual methods for this region.

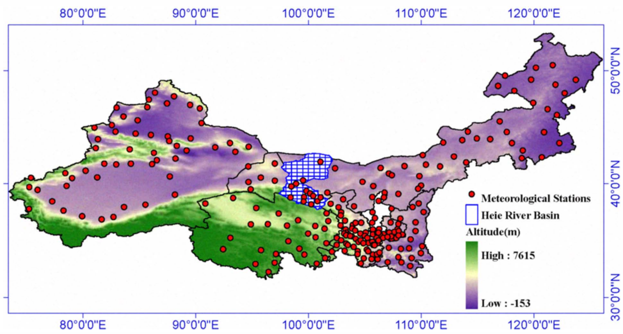


Fig1. Study area and locations of meteorological stations

The dataset used in this study comes from China Meteorological Data Sharing Service System (http://cdc.cma.gov.cn). The dataset contains 241 national meteorological observation stations and four meteorological factors which temperature, precipitation, evaporation and wind speed form 1951-2005. The sites are scarce and unevenly distributed (seen in Fig.1 & Table 1). They trend to appear in densely populated areas.

TABLE I. METEOROLOGICAL OBSERVATION SITES DISTRIBUTION

Regions	*No. of stations*	*Area(10^4km^2)*
Inner Mongolia	45	118.3
Shaanxi	19	20.6
NingXia	12	6.6
GauSu	80	45.5
QingHai	31	72.2
XinJiang	54	163.3
*Heihe River Basin**	13	13.0

**Heihe River Basin is an important landscape within the study area.*

In this paper, we choose monthly mean temperature as an example to illustrate the applications ANN on spatial interpolation. It is usually a good idea to scan dataset, which is used to spatial interpolation. The frequency histogram for the July 2000 temperature is shown in Fig.2. The corresponding frequency shows the strong skewed distribution.

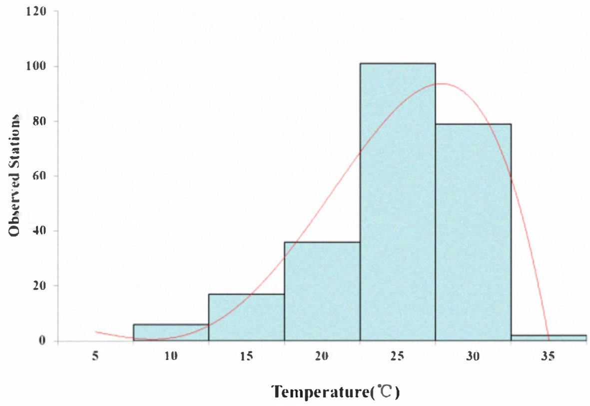


Fig 2. Frequency histogram of the July 2000 monthly mean air temperature of 241 national stations, the smooth curve is trend line of stations distribution.

B. Model Development

ANNs are a very simplified version of real neural networks. One of the most common and widely used topologies of neural networks is the so-called back propagation (BP) [6]. It is the generalization of the Widrow-Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer function. Networks with biases, a sigmoid layer, and a linear output layer are capable of approximating any function with a finite number of discontinuities. Therefore, we choose BP neural network for model development, as in previous studies of air temperature prediction [4, 5, and 7]. Because two-layer feed-forward, networks can potentially represent any input-output relationship with a finite number of discontinuities. So our BP neural network model is a two-layer feed-forward network (see Fig 3), and the hidden layer of final model has 10 neurons. The model's training algorithm adopts Levenberg-Marquardt algorithm (LMA). The LMA is an iterative technique that locates the minimum of a multivariate function that is expressed as the sum of squares of non-linear real-valued functions, which is also the fastest method up to now [8]. We use tan-sigmoid transfer functions for hidden layers and the output layer is linear respectively. The functions are given by:

$$y_{\tan sig}(x) = \frac{2}{1+e^{-2x}} - 1 \tag{1}$$

$$y_{purelin}(x) = x \tag{2}$$

Air temperature displays complex and nonlinear variation with location, topography, and complex surface characters.

Elevation has guided the spatial interpolation of temperature in numerous other studies [9, 10, and 11]. In our study, the model inputs included six impact factors: latitude, longitude, slope, aspect and altitude, normalized difference vegetation index (NDVI). The input dataset was prepared by geographical information system (GIS) methods and resample to 1000m resolution. The output of the model was temperature of stations observation.

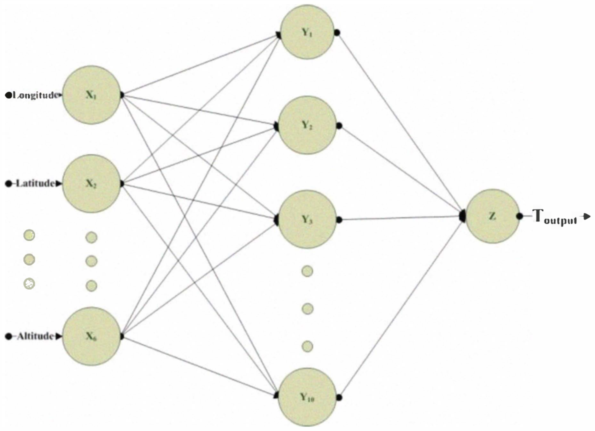


Fig 3. Architectural structure of spatial interpolation based on the BP neural work.

The spatial distribution of sites used in this work was illustrated in Fig. 4.We randomly divided the total 241 inputs and targets vectors into three sets. Sixty percent of the vectors were used to train the network and twenty percent of the vectors were used to validate how well the network generalized. The training vectors continue to train the model as long as the network's validation error reduces to the goal or timeout reached. Finally, the last Twenty percent of the vectors provide an independent test dataset that the network has never seen.

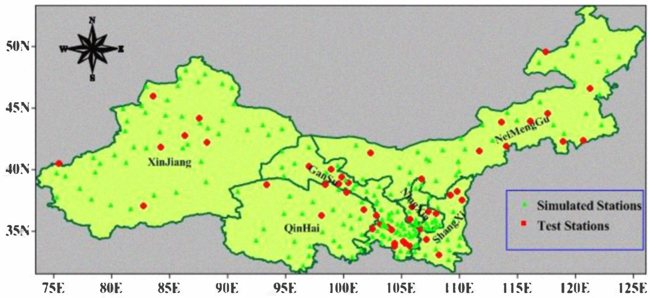


Fig4. Distribution and division of the meteorological observatories in study area

To make neural network training more efficient，the network inputs and targets were performed certain preprocessing steps. Equation (2) is often used to scale the inputs and targets values x through a linear transformation so that they always fall within a specified range [y_{min}, y_{max}]. The Equation is given by ：

$$x_{scale} = (y_{\max} - y_{\min}) * \frac{x - x_{\min}}{x_{\max} - x_{\min}} + y_{\min} \qquad (2)$$

Where y_{min} and y_{max} are the scaled range, in ours research The range was [-1, 1], x_{min} and x_{max} were the minimum and maximum values of the variable the data set.

III. Results and Discussion

The model trained for July 2000 mean air temperatures. As we can see in Fig.4, that model present a good learning curve with a final mean square error (MSE) very close to the optimum value of 0.01 after 6 epochs.

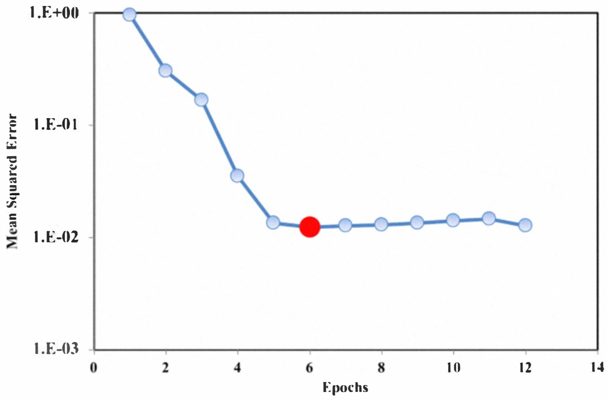
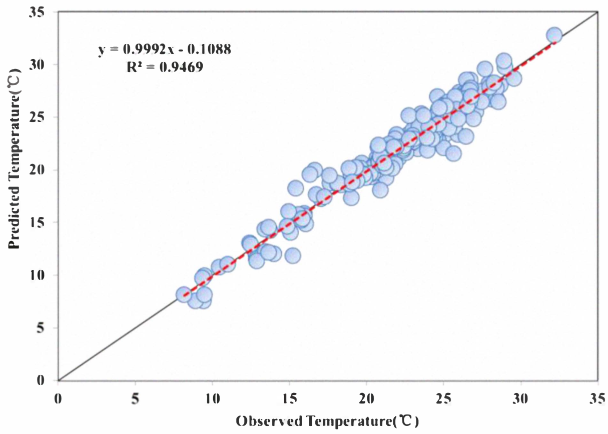
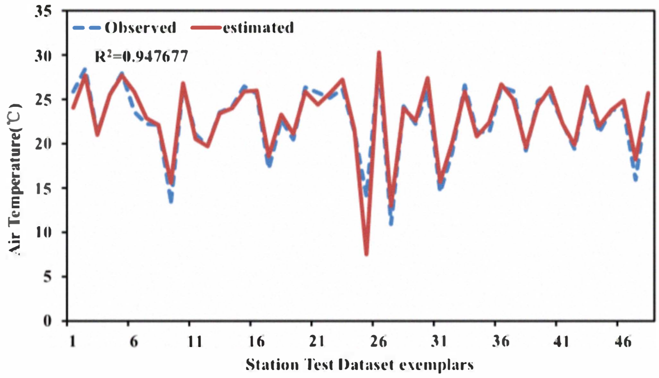


Fig 4. The learning curve BP model

In order to check the network performance with new data, we tested the model with the 48 stations of the test dataset. The network present low errors and a high correlation coefficient of 0.95 between measured and predicted values as it showed in Fig 6. The strong relation between temperature and independent variables (latitude, longitude, slope, aspect and altitude, NDVI) can explain this very good result and the fast and easy training.



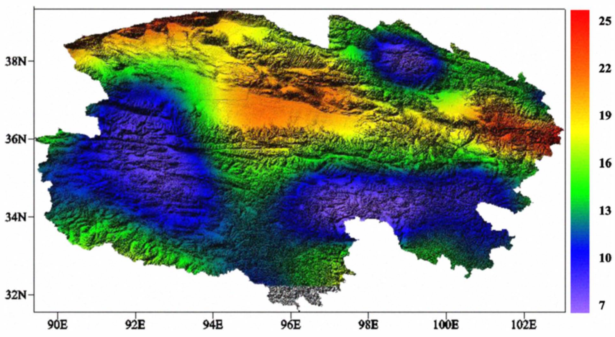
(a)



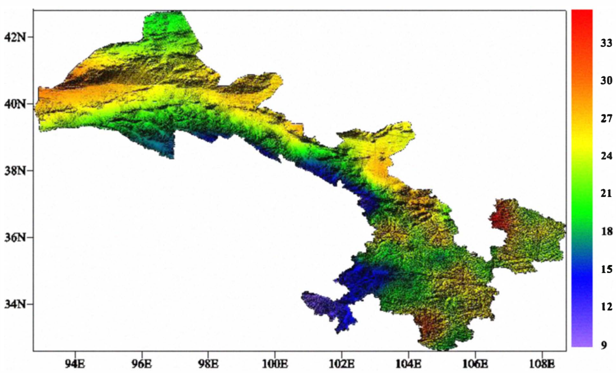
(b)

Fig 5. Observed vs. estimated July 2000 monthly mean temperature of western China whole dataset *(a)* and test dataset*(b)*

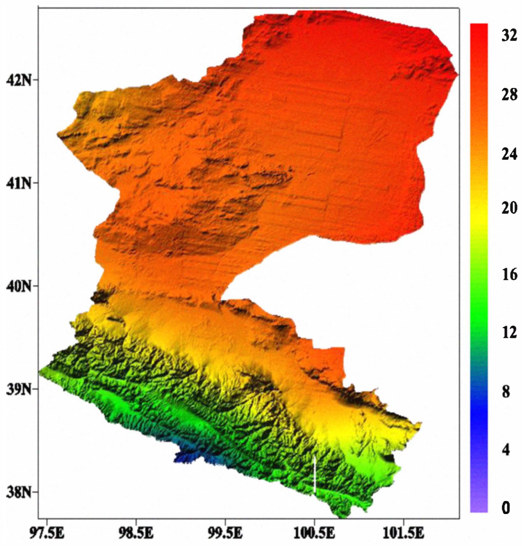
After successful training and testing the model, which be applied to all new data to obtain a grid of predicted values. Then we extracted the point dataset of different areas by GIS system and disaggregated to spatial continuous dataset as in Fig 6.



(a)



(b)



(c)

Fig 6. Distribution of monthly mean temperature in July 2000 of (a) Qinghai province (b) Gansu province and (c) HeiHe River Basin by predicted the BPNN model

The comparison between estimated and measured values for the whole training dataset and using the final model leads to an mean bias error (MBE) of 0.012. The statistical results for the Heihe river basin are listed in Table 2(the end of the paper), where covers 1.43 $\times 10^5 km^2$ and has only 13 national stations. However, the errors between observed and predicted values are still small. The relative error was ranged from 0.6 % (station ID 52446) to -7.6% (station ID 52643).

TABLE II. Statistical Results Obtained From the BP Neural Network Model for Each Station of Heihe River Basin

Station ID	Observed Temperature(℃)	Predicted Temperature(℃)	Relative Error(%)
52446	26.3	26.4	0.6
52447	26.3	25.5	-3.2
52533	23.9	24.7	3.3
52557	24.3	25.7	5.5
52546	24.4	25.7	4.8
52633	13.0	12.9	-0.7
52643	18.6	17.2	-7.6
52645	12.0	12.7	6.0
52652	24.3	24.8	2.1
52656	18.6	18.8	1.2
52657	15.6	15.5	-0.9
52661	23.0	23.9	3.8
52267	29.5	28.6	-3.1

IV. CONCLUSION

In this research, we developed a new method for the spatial interpolation the sparse observed data by taking advantage of artificial neural networks evaluation. In case of western China'1951-2005 monthly mean climate variables taking monthly mean air temperature as example, a BP neural network model has been constructed to spatially interpolate routine monthly weather observations using location (latitude and longitude), topography (slope, aspect and elevation), normalized difference vegetation index as inputs for developing spatial interpolation functions that can produce gridded weather datasets. We used model-derived temperatures to produce the spatio-temporal dataset in Western China. In addition to producing historical spatial weather fields, we are also investigating the application of BP neural network combined with wavelet method to interpolate the missing data or predict future value.

Finally, note that the methodology applied in this work calibrated the model proposed for the air temperature studied, but the same approach can be applied on other climatic factor or environmental observed variable.

ACKNOWLEDGMENT

This work is supported by Chinese Academy of Sciences Special Fund for Informatization (NO.INFO115D02) and National Fundamental Science Talents Science Foundation of China (NO.J0630966)

REFERENCES

[1] Dubois G., and Galmarini, S. "Introduction to the Spatial Interpolation Comparison (SIC) 2004 exercise and presentation of the data sets", *Applied GIS*, Vol.1,No. 2, 2005.

[2] Meyers, D.E. "Spatial interpolation: An overview", *Geoderma*,Vol.62, pp. 17-28, 1994.

[3] Arthur T. DeGaetano and Brian N. Belcher." Spatial Interpolation of Daily Maximum and Minimum Air Temperature Based on Meteorological Model Analyses and Independent Observations", Journal of Applied Meteorology and Climatology, July, 2006.

[4] Ahmet Oztopal. "Artificial neural network approach to spatial estimation of wind velocity data ", *Energy Conversion And Management* ,Vol. 47, pp.395-406, October, 2006.

[5] Brian A. Smith, Gerrit Hoogenboom, RonaldW. McClendon."Artificial neural networks for automated year-round temperature prediction", *computers and Electronics in Agricultur*e, Vol.68, pp. 52–61, 2009.

[6] Mutasem khalil Sari Alsmadi, Khairuddin Bin Omar and Shahrul Azman Noah. "Back Propagation Algorithm: The Best Algorithm Among the Multi-layer Perceptron Algorithm", *International Journal of Computer Science and Network Security*, Vol.9, No.4, April, 2009.

[7] Snell,S.E., S.Gopal and R.K.Kaufmann. "Spatial interpolation of surface air temperature using artificial neural networks: Evaluating their use for downscaling ", *Journal of Climate*, Vol. 13, pp. 886-895, 2000.

[8] Hagan, M. T. and Menhaj, M., "Training feedforward networks with the Marquardt algorithm", *IEEE Transactions on Neural Networks*, Vol.5, No. 6, pp. 989-993, 1994

[9] Price, D.T., D.W. McKenney, I.A. Nalder, M.F. Hutchinson and J.L. Kesteven. "A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data", *Agricultural and Forest Meteorology* ,Vol. 101,pp. 81-94, 2000.

[10] Johnson, G.L., C. Daly, G.H. Taylor and C.L. Hanson. "Spatial variability and interpolation of stochastic weather simulation model parameters", *Journal of Applied Meteorology*, Vol. 39,pp.778-796, 2000.

[11] Kurtzman, D. and R. Kadmon. "Mapping of temperature variables in Israel: a comparison of different interpolation methods", *Climate Research*, Vol. 13, pp. 33-43,1999.