BioOne COMPLETE

A Multivariate Regression Model for Predicting Precipitation in the Daqing Mountains

Authors: Ranhao, Sun, Baiping, Zhang, and Jing, Tan

Sun Ranhao, Zhang Baiping, and Tan Jing

A Multivariate Regression Model for Predicting Precipitation in the Daqing Mountains

Multivariate regression analysis, combined with residuals correction, was carried out to develop a precipitation prediction model for the Daqing Mountains of Inner Mongolia in northern China. Precipitation data collected at 56 stations between 1955 and 1990 were used: data from 48 stations for model development and data from 8 stations for additional tests. Five topographic factors—altitude, slope, aspect, longitude, and latitude—were taken into account for model development. These topographic variables were acquired from a 100-m resolution digital elevation model (DEM) of the study region, and the mean values of the sub-basin in which a precipitation station is located were used as the values of the respective variables of that station. The multivariate regression model can explain 72.6% of the spatial variability of precipitation over the whole year and 74.4% of variability in the wet season (June–September). Precipitation in the dry season (October–May) is hard to model owing to little rainfall (21.78% of annual rainfall) and a different synoptic system. Interpolation-based residuals correction did not significantly improve the accuracy of our model, which shows that our model is quite effective. The model, as presented in this paper, could potentially be applied to other mountains and in mountain climate research.

Introduction

Environmental research, management and planning—such as agricultural and forest management—hydro-ecology modeling, and environmental assessments usually require spatially continuous data for precipitation. However, precipitation is usually measured at a very limited number of stations, especially in mountain areas. As a result, accurate estimation and prediction of precipitation represents a great challenge due to lack or non-validity of observation data (Daly et al 1994; McGuffie and Henderson-Sellers 2001; Alijani et al 2007). The development of geographic information systems (GIS) in recent years provides increased opportunities for precipitation modeling.

Interpolation methods have been developed for rainfall modeling. Most of them are based mainly on the similarity and topological relations of nearby sample points and on the value of the variable to be measured (Beek et al 1992; Gemmer et al 2004; Chang et al 2005). Interpolation can be achieved using simple mathematical models (inverse distance weighting, trend surface analysis, splines and Thiessen polygons, etc) or more complex models (geostatistical methods, such as kriging). Geostatistical interpolation has become an important tool in climatology because it is based on the spatial variability of the variables of interest and makes it possible to quantify the estimation uncertainty (Martinez-Cob 1996; Holawe and Dutter 1999). However, interpolation methods only consider spatial relationships among sampling points, and do not take into account other important topographic variations. Consequently, the usual interpolation methods cannot provide researchers with adequate precision of precipitation estimation, especially in complicated terrains of mountainous regions (Marquínez et al 2003; Sharples et al 2005).

In recent years, geographic and topographic factors have been integrated into the modeling of precipitation (Johansson and Chen 2005). Some authors have attempted to incorporate local topographic factors, such as elevation, into geostatistical approaches (Goovaerts 2000), and others have developed models relating climate to site position and elevation (Goodale et al 1998). Relationships between topography and the spatial distribution of precipitation have been analyzed for mountainous regions (Basist et al 1994; Cheval et al 2003). The Precipitation–elevation Regressions on Independent Slopes Model (PRISM) brings a combination of climatological and statistical concepts to the analysis of orographic precipitation (Daly et al 1994), and recently PRISM has used weighting functions to incorporate gauge data of neighboring topographic facets for regressions (Daly et al 2002). More recent studies have considered more refined topographic factors by using higher resolution digital elevation models (DEM) to predict the physical influence of topographic variables on precipitation patterns. Precipitation models integrating statistical and GIS techniques have become widespread and common (Ollinger et al 1995; Goovaerts 2000). Multiple linear (Ninyerola et al 2000, 2007; Brown and Comrie 2002; Guan et al 2005) and non-linear (Marquínez et al 2003) regression models have proved to be rather effective.

The present study, focusing on the Daqing Mountains in Inner Mongolia of northern China, was conducted to develop a multivariate regression model for estimation of precipitation in mountain regions, based on five topographic factors (altitude, slope, aspect, longitude, and latitude).
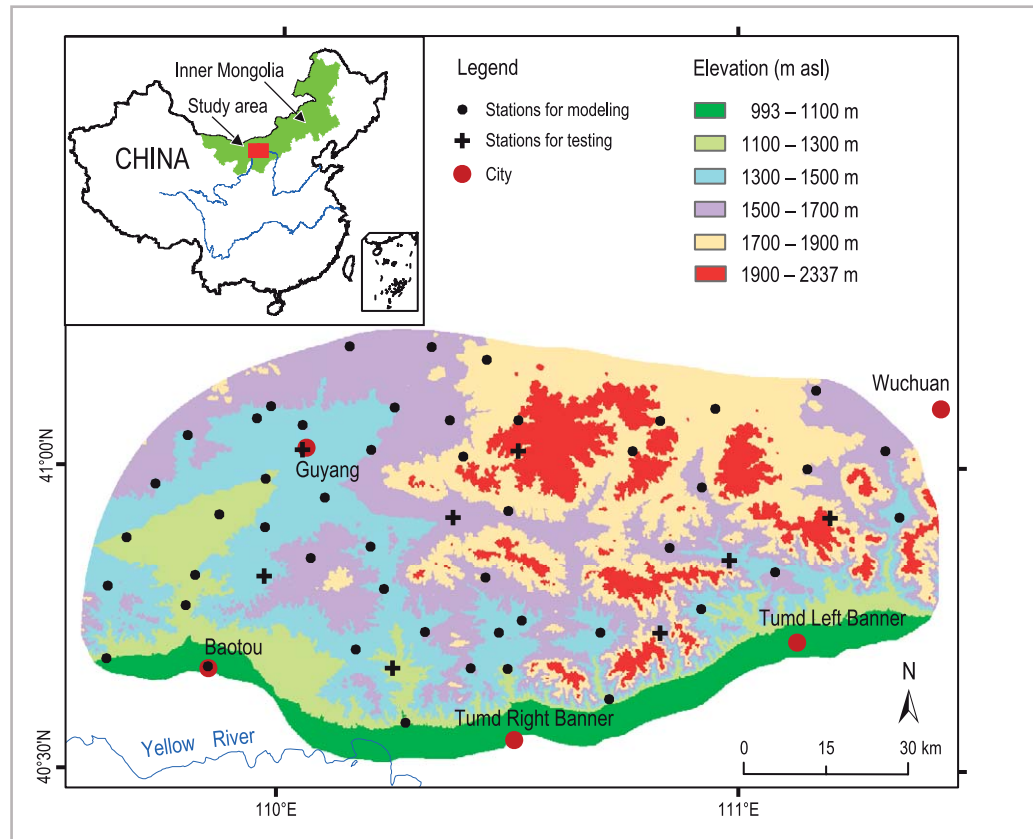
FIGURE 1 Location of the Daqing Mountains and of the 56 precipitation stations. (Map by Sun Ranhao)

Materials and methods

Study area

In northern China, many mountain ranges block the summer monsoon from the Pacific Ocean, preventing it from penetrating deep into the heartland of Asia: the Da (Great) Hinggan, Taihang, Yinshan, Helan, and Mazong mountains; as a result, they usually constitute significant moisture and temperature boundaries. The Daqing Mountains are located roughly in the center of the Inner Mongolia Autonomous Region, at 40°20′–41°20′N and 100°30′–111°30′E, covering an area of 9336 km^2 (Figure 1). Their elevation ranges from 990 m to nearly 2340 m asl within only a few kilometers of planimetric distance. The local climate is dominated by warm-moist summers and cold-dry winters. The Daqing Mountains form the dividing line between the warm-temperate and temperate zones in northern China. As a climatic and geographic transition belt, this region is particularly suited to the study of relationships between precipitation and topographic factors.

Data collecting and processing

Precipitation data from 56 stations were obtained from the observation archives of the Yellow River Conservancy Commission (1955–1990). All of these records cover over ten years, 38 of them more than twenty years. For each station, the annual average amount of precipitation was used. Density was one station for every 167 km^2, with most of these stations located at medium altitude, between 1060 m and 1954 m asl (Table 1 and Figure 1). The stations were quite evenly distributed both horizontally and vertically. The majority of annual rainfall, averaging 78.22%, occurs in the four months from June to September. This period is the wet season, and the other months make up the dry season.

There are many factors which affect precipitation and its spatial distribution. Usually, precipitation increases with growing elevation; it varies depending on slope and aspect; the amount and intensity of precipitation is larger on windward than on leeward slopes; and differences increase with growing steepness of the slope (Fu 1992). Factors closely related to precipitation in the Daqing Mountains include not only rugged orography, but also geographical location, as rainfalls in the study region are mainly due to moisture-laden air masses from the southeast (Pacific summer monsoon). Therefore, latitude and longitude, understood in terms of their relative distance to the Pacific Ocean, should be considered in the development of precipitation modeling for this area.

The introduction of GIS and the corresponding powerful software have brought new capacities to the

320

TABLE 1 Vertical distribution of precipitation stations in the Daqing Mountains.

Altitude (m asl)	Number of stations	Relative proportions (%)	Relative area (%)	Density (number of stations/km²)
<1100	2	3.6	8.9	1/414
1100–1300	5	8.9	10.5	1/195
1300–1500	19	33.9	22.7	1/111
1500–1700	18	32.1	27.2	1/141
1700–1900	10	17.9	22.1	1/206
>1900	2	3.6	8.7	1/404

TABLE 2 Pearsonian correlation coefficient matrix for independent variables and precipitation data.

Variable	Whole year	Wet season	Dry season	Altitude	Slope	Aspect	Longitude
Altitude	0.35[a]	0.32[a]	0.41[b]				
Slope	0.74[b]	0.77[b]	0.38[b]	0.16			
Aspect	0.02	0	0.09	0.16	0.13		
Longitude	0.49[b]	0.50[b]	0.28[a]	0.67[b]	0.38[b]	–0.09	
Latitude	–0.46[b]	–0.48[b]	–0.2	0.54[b]	–0.53[b]	0.02	0.15

[a] Significant at 95%; [b] Significant at 90%.

world of geography (Longley 2000). Using ArcGIS 9.0 software, we were able to obtain topographic variables from the DEM based on 1:250,000 contour maps. The DEM had a 100-m resolution, with a projected coordinate system of WGS_1984_Transverse_Mercator. Smoothing processing of predictors—eg smoothing of neighboring grids (Daly et al 1994; Wotling et al 2000; Sharples et al 2005)—or using mean values of the sub-basins (Marquínez et al 2003) is usually applied to increase the representativeness of the stations. For this purpose, we divided the study region into more than 1900 sub-basins with the help of the ArcGIS hydrology module. These sub-basins covered areas between 2.96 and 26.67 km², with an average of 4.9 km². The mean values of the topographic factors (elevation, slope, and aspect) in the sub-basin in which a station was located were used as the values of the respective variables of that station. In addition, we calculated the topographic variables for the five layers (altitude, slope, aspect, longitude, and latitude) as raster models using 100-m resolution grids. We used a 100-m grid because the study region is not large but varies greatly (the elevation ranges from 990 m to nearly 2340 m asl within only a few kilometers of planimetric distance on the southern flank of the Daqing Mountains). We also tried

smoothing values (eg with 5000 neighboring cells), and results showed that virtually no detailed information was missing in our model.

Methods

The combination of statistical (*multivariate regression*) and spatial interpolation (*ordinary kriging*) has been demonstrated to be effective in modeling precipitation (Ninyerola et al 2007). Topographic variables, if considered on their own, are poorly related to rainfall statistics, as shown in Table 2. In other words, none of the topographic variables could be used by itself to appropriately explain the precipitation pattern. Considering that linear equations are not well suited to depicting the relationships of precipitation and topographic variables, we tried to model precipitation with multivariate nonlinear regression, as follows:

$$
\begin{aligned}
P = b0 &+ b1h + b2h^2 + b3h^3 \\
&+ b4slp + b5slp^2 + b6slp^3 \\
&+ b7asp + b8asp^2 + b9asp^3 \qquad (1) \\
&+ b10X + b11X^2 + b12X^3 \\
&+ b13Y + B14Y^2 + b15Y^3
\end{aligned}
$$

TABLE 3 Regression statistics and evaluation of the multivariate regression model (including 48 stations used for modeling). R^2 is the determination coefficient that serves as a measure of the goodness-of-fit of the model; adj_R^2 is the adjusted determination coefficient, which compensates for the limitation of the determination coefficient by taking into account the size of the sample and the number of prediction variables, and it exactly represents the proportion of variation of the dependent variable (ie annual and seasonal mean precipitation) explained by the multivariate regression model; $RMSE$ is the root mean squared error, which describes the error of prediction in the modeling of precipitation; F is the value of the mathematics test; DW is the value of the Durbin-Watson statistic—a test statistic used to detect the presence of autocorrelation in residuals, based on regression analysis.

	Whole year		Wet season		Dry season	
	Unstandardized coefficients	Standardized coefficients	Unstandardized coefficients	Standardized coefficients	Unstandardized coefficients	Standardized coefficients
b0	3364.134		2955.065		407.815	
b1(h)	2.466×10^{-2}	0.1	6.892×10^{-2}	0.32	-4.466×10^{-2}	−1.01
b2(h^2)	0	0	0	0	0	0
b3(h^3)	1.75×10^{-8}	0.53	6.905×10^{-9}	0.24	1.065×10^{-8}	1.77
b4(slp)	−9.184	−0.68	−9.65	−0.82	4.877×10^{-1}	0.2
b5(slp^2)	1.574	2.06	1.762	2.65	-1.892×10^{-1}	−1.36
b6(slp^3)	-5.121×10^{-2}	−1.21	-5.985×10^{-2}	−1.62	8.656×10^{-3}	1.12
b7(asp)	-2.167×10^{-2}	−0.01	-3.439×10^{-1}	−0.27	3.225×10^{-1}	1.21
b8(asp^2)	5.171×10^{-3}	1.12	6.276×10^{-3}	1.55	-1.114×10^{-3}	−1.32
b9(asp^3)	-2.12×10^{-5}	−1.16	-2.126×10^{-5}	−1.33	8.827×10^{-8}	0.03
b10(X)	-7.086×10^{-1}	−0.51	−0.227	−1	5.201×10^{-1}	2.04
b11(X^2)	0	0	0	0	0	0
b12(X^3)	1.395×10^{-6}	0.62	2.238×10^{-6}	1.13	-8.443×10^{-7}	−2.05
b13(Y)	0	0	0	0	0	0
b14(Y^2)	0	0	0	0	0	0
b15(Y^3)	-3.198×10^{-8}	−0.64	-2.641×10^{-8}	−0.61	-5.555×10^{-9}	−0.61
R^2	0.79		0.804		0.552	
adj_R^2	0.726		0.744		0.415	
RMSE (mm)	28.11		23.72		7.49	
RMSE (%)	8.38%		8.02%		18.78%	
Measured precipitation (mm)	335.58		295.69		39.88	
F (sig<0.01)	12.34		13.449		4.037	
DW	2.341		2.255		2.289	

where *P* represents precipitation (mm); *b*0 is constant; *b*1 … *b*15 represent the coefficients obtained for each independent variable for wet season, dry season, and whole year; *h, slp, asp, X, Y* represent the variables of altitude (m), slope (°), aspect (°), longitude (km), and latitude (km), respectively.

From the 56 precipitation stations, a total of 48 were selected for modeling by random sampling using SPSS 13 software, and the remaining 8 stations were used for validation. The constants were acquired by means of the least squares method in the regression module of SPSS, and all the regression statistics were also created by the software (Table 3).

Results and discussion

About the regression model
It has been shown that when the fit set accounts for more than 80% of the whole set, the adjusted determi-

FIGURE 2 Error stability of the precipitation prediction model for different fit sets.
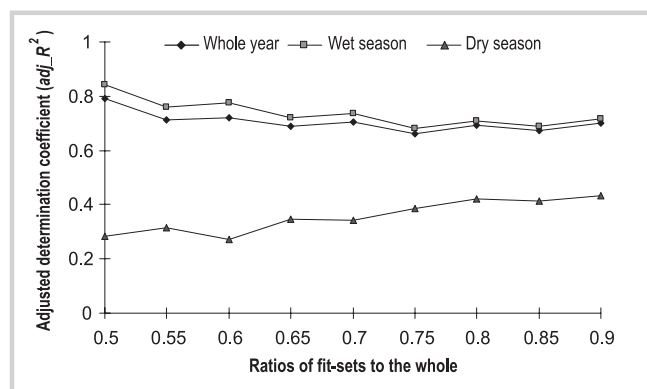
nation coefficient (adj_R^2) tends to remain stable. Hence, it is reasonable to select 85% (48 stations) of the whole set as the fit set (Figure 2).

The equations for both annual and seasonal precipitation pass the *F* tests at 0.01 significance level, and the high determination coefficient (R^2) and adj_R^2 show the goodness-of-fit of the equations (Table 3).

The capability of the model to explain the spatial variability of precipitation varies depending on the period: its accuracy is 74.4% for the wet season and 72.6% for the whole year, but only 41.5% for the dry season. The root mean squared error (RMSE)—an index for estimating relative error— is 8.02%, 8.38%, and 18.78% for the three periods, respectively. This shows that our multivariate regression model works well for both the whole year and the wet season.

Residuals estimation
A Durbin-Watson test was conducted for residuals estimation. The values in Table 3 show little autocorrelation among the residuals, meaning that our model is of good quality and able to predict most of the spatial variability of precipitation in the Daqing Mountains.

In order to improve the results of the regression model, we adopted ordinary kriging to obtain a residuals surface. An exponential model was used for predicting the residuals of the 48 precipitation stations. After

some trials, lag size was finally fixed at 8 km, with a total of 8 lags. The searching neighborhood was elliptical and divided into 4 sectors, and their maximum and minimum numbers of neighbors to be included were 5 and 2 in each case. Ultimately, the corrector layers of distribution for residuals were produced in ArcGIS. After overlaying them onto the regression surface, we obtained the spatial distribution of precipitation (Figures 3A to 3C).

Spatial pattern of precipitation
In the wet season, precipitation clearly decreases from southeast to northwest. In the dry season, precipitation is almost randomly distributed, with high values in some highlands only, such as in the center and south of the study region. The precipitation pattern for the whole year is similar to that of the wet season, with less precipitation in the northwest and on the lee sides and relatively much precipitation on the southeast slopes. We also found that precipitation in the Great Bend of the Yellow River (the Hetao Plain) is higher than in the Inner Mongolian Plateau, which is situated at much higher altitude.

Validation of the multivariate regression model with 8 test stations
We calculated the mean absolute error (MAE) and RMSE of modeled precipitation for the 8 test stations (Table 4). The RMSE for whole year, wet season, and dry season yielded precipitation values of 44.08 mm, 36.71 mm, and 12.42 mm, respectively, accounting for 12.77%, 11.94%, and 32.96% of measured precipitation. Upon considering the kriged residuals grids, predicted precipitation values increased only slightly, ie by less than 3 mm, adjusting for MAE and RMSE. This demonstrates that the residuals are random errors, and that our regression model works quite well for interpreting the spatial variability of precipitation.

It is evident that relative errors are higher for the dry season than for the wet season and for the whole year. The reason is that, in the dry season, the study

TABLE 4 Validation of the multivariate regression model with 8 test stations.

Period	MAE[a] (mm)	MAE (%)	RMSE[a] (mm)	RMSE (%)	MAE[b] (mm)	MAE (%)	RMSE[b] (mm)	RMSE[b] (%)	Measured precipitation (mm)
Whole year	33.37	9.67	44.08	12.77	32.26	9.35	42.1	12.2	345.15
Wet season	29.47	9.58	36.71	11.94	29.09	9.46	35.72	11.62	307.45
Dry season	9.66	25.61	12.42	32.96	7.33	19.44	11.31	29.99	37.7

[a]after regression only; [b]after kriging.

324
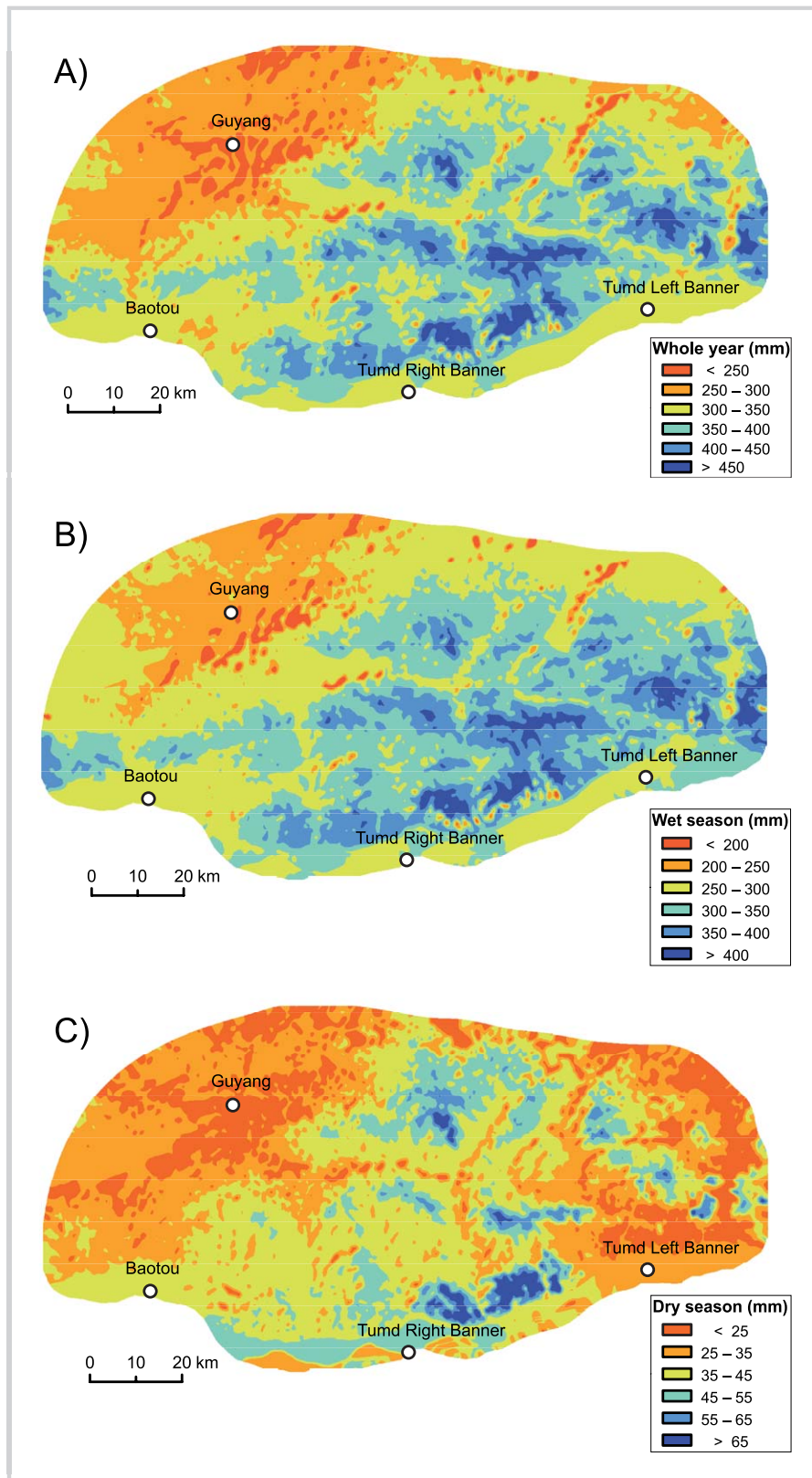
region is under the influence of completely different synoptic systems, especially the strong Siberian High Pressure system, and intensified westerly winds (Liu et al 2004). They provide little rainfall in the study region and disturb the general pattern of rainfall distribution, thus affecting the accuracy of our model.

Conclusions

The multivariate regression model developed for the Daqing Mountains is able to describe 72.6% of the spatial variability of annual precipitation in the study region, and it is much more effective for the wet season than for the dry season.

Multivariate regression models have universality in statistics, and they can depict most of the spatial variability of precipitation in mountain regions. Our model only uses DEM and precipitation data from a limited number of stations, but it could be easily applied to other mountain areas and in mountain climate research. We consider local prevailing winds, especially their direction and force, also to be an important variable for rainfall patterns. Integration of the wind variable in modeling would thus probably increase the accuracy of our model.

AUTHORS

Sun Ranhao
State Key Laboratory of Resource and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, No. A-11, Datun Road, Chaoyang District, Beijing 100101, P.R. China.
Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, No. 18 Shuangqing Road, Haidian District, Beijing 100085, P.R. China.
sunrh@lreis.ac.cn

Zhang Baiping (corresponding author), Tan Jing
State Key Laboratory of Resource and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, No. A-11, Datun Road, Chaoyang District, Beijing 100101, P.R. China.
zhangbp@lreis.ac.cn (Z.B., corresponding author); Tanj@lreis.ac.cn (T.J.)

REFERENCES

Alijani B, Ghohroudi M, Arabi N. 2007. Developing a climate model for Iran using GIS. *Theoretical and Applied Climatology,* published online, 16 May 2007. doi:10.1007/s00704-006-0292-y.
Basist A, Bell GD, Meentemeyer V. 1994. Statistical relationships between topography and precipitation patterns. *Journal of Climate* 7(9):1305–1315.
Beek EG, Stein A, Janssen LLF. 1992. Spatial variability and interpolation of daily precipitation amount. *Stochastic Hydrology and Hydraulics* 6:304–320.
Brown DP, Comrie AC. 2002. Spatial modeling of winter temperature and precipitation in Arizona and New Mexico, USA. *Climate Research* 22:115–128.
Chang CL, Lo SL, Yu SL. 2005. Applying fuzzy theory and genetic algorithm to interpolate precipitation. *Journal of Hydrology* 314:92–104.
Cheval S, Baciu M, Breza T. 2003. An investigation into the precipitation conditions in Romania using a GIS-based method. *Theoretical and Applied Climatology* 76(1):77–88. doi:10.1007/s00704-003-0004-9.
Daly C, Gibson WP, Taylor GH, Johnson GL, Pasteris P. 2002. A knowledge-based approach to the statistical mapping of climate. *Climate Research* 22:99–113.
Daly C, Neilson RP, Phillips DL. 1994. A statistical topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* 33:140–158.
Fu B. 1992. The effects of topography and elevation on precipitation [in Chinese with English abstract]. *Acta Geographica Sinica* 47(4):302–314.
Gemmer M, Becker S, Jiang T. 2004. Observed monthly precipitation trends in China 1951–2002. *Theoretical and Applied Climatology* 77:39–45.

Goodale CL, Aber JD, Ollinger SV. 1998. Mapping monthly precipitation, temperature, and solar radiation for Ireland with polynomial regression and a digital elevation model. *Climate Research* 10:35–49.
Goovaerts P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology* 228(1):113–129.
Guan H, Wilson JL, Makhnin O. 2005. Geostatistical mapping of mountain precipitation incorporating autosearched effects of terrain and climatic characteristics. *Journal of Hydrometeorology* 6:1018–1031.
Holawe F, Dutter R. 1999. Geostatistical study of precipitation series in Austria: Time and space. *Journal of Hydrology* 219(1–2):70–82. doi:10.1016/S0022-1694(99)00046-3.
Johansson B, Chen DL. 2005. Estimation of areal precipitation for runoff modeling using wind data: A case study in Sweden. *Climate Research* 29(1):53–61.
Liu C, Qian Z, Wu M, Song M, Liu J. 2004. A composite study of the synoptic differences between major and minor dust storm springs over the China–Mongolia areas. *Terrestrial, Atmospheric and Oceanic Sciences* 15(5):999–1018.
Longley PA. 2000. The academic success of GIS in geography: Problems and prospects. *Journal of Geographical Systems* 2(1):37–42. doi:10.1007/s101090050027.
Marquínez J, Lastra J, García P. 2003. Estimation models for precipitation in mountainous regions: The use of GIS and multivariate analysis. *Journal of Hydrology* 270(1):1–11.
Martinez-Cob A. 1996. Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain. *Journal of Hydrology* 174:19–35.

McGuffie K, Henderson-Sellers A. 2001. Forty years of numerical climate modeling. *International Journal of Climatology* 21(9):1067–1109. doi:10.1002/joc.632.

Ninyerola M, Pons X, Roure JM. 2000. A methodological approach of climatologically modeling air temperature and precipitation through GIS techniques. *International Journal of Climatology* 20(14):1823–1841.

Ninyerola M, Pons X, Roure JM. 2007. Monthly precipitation mapping of the Iberian Peninsula using spatial interpolation tools implemented in a Geographic Information System. *Theoretical and Applied Climatology* 89:195–209.

Ollinger SV, Aber JD, Federer AC, Lovett GM, Ellis JM. 1995. *Modeling Physical and Chemical Climate of the Northeastern United States for a Geographic Information System.* United States Forest Service General Technical Report NE-191. Radnor, PA: United States Department of Agriculture.

Sharples J, Hutchinson MF, Jellett DR. 2005. On the horizontal scale of elevation dependence of Australian monthly precipitation. *Journal of Applied Meteorology* 44(12):1850–1865.

Wotling G, Bouvier C, Danloux J, Fritsch JM. 2000. Regionalization of extreme precipitation distribution using the principal components of the topographical environment. *Journal of Hydrology* 233(1):86–101.