INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. (in press)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/joc.1588

RMetS
ROYAL METEOROLOGICAL SOCIETY

Predicting summer rainfall in the Yangtze River basin with neural networks

Heike Hartmann,[a]* Stefan Becker[b] and Lorenz King[a]
[a] *Department of Geography, Justus Liebig University, Senckenbergstr. 1, 35390 Giessen, Germany*
[b] *Department of Geography and Urban Planning, University of Wisconsin Oshkosh, Oshkosh WI 54901, USA*

ABSTRACT: Summer rainfall in the Yangtze River basin is predicted using neural network techniques. Input variables (predictors) for the neural network are the Southern Oscillation Index (SOI), the East Atlantic/Western Russia (EA/WR) pattern, the Scandinavia (SCA) pattern, the Polar/Eurasia (POL) pattern and several indices calculated from sea surface temperatures (SST), sea level pressures (SLP) and snow data from December to April for the period from 1993 to 2002. The output variable of the neural network is rainfall from May to September for the period from 1994 to 2002, which was previously classified into six different regions by means of a principal component analysis (PCA). Rainfall is predicted from May to September 2002.

The winter SST and SLP indices are identified to be the most important predictors of summer rainfall in the Yangtze River basin. The Tibetan Plateau snow depth, the SOI and the other teleconnection indices seem to be of minor importance for an accurate prediction. This may be the result of the length of the available time series, which does not allow a deeper analysis of the impact of multi-annual oscillations.

The neural network algorithms proved to be capable of explaining most of the rainfall variability in the Yangtze River basin. For five out of six regions, our predictions explain at least 77% of the total variance of the measured rainfall.

KEY WORDS principal component analysis; correlation analysis; neural network analysis; summer rainfall; China; Yangtze

Received 13 December 2006; Revised 30 May 2007; Accepted 3 June 2007

1. Introduction

In the 1990s, frequent floods, causing horrible human as well as economic losses, drew international media attention to the Yangtze River basin. There are two main types of flooding disaster: flooding which affects only a smaller region; and flooding which hits most parts of a river basin. The first type is caused by focused rainstorms, whereas the second type is produced by widespread rainstorms which affect tributaries of the upper, middle and lower reaches, as well as the main river at the same time (Zhao, 1999). It is of great importance to predict heavy precipitation – rainstorms – as early as possible in order to increase the time until the flood situation starts, and by that to improve the preparation for floods.

Several studies have been undertaken to relate the occurrence of widespread rainstorms in the region to changes in the atmospheric or oceanic circulation. Most of the studies only focus on one influencing factor:

Several studies have examined the impact of ENSO cycles on precipitation or the flood/drought index in the study region. Gao *et al.* (2006) analyse correlations within the last 50 years whereas Jiang *et al.* (2006) and Wu *et al.* (2006) extend the time frame to the last 500 years. Guo *et al.* (2002) name positive sea surface temperatures (SST) anomalies over the Indian Ocean as a contributor to Yangtze River floods. Using the 1998 flood as an example, they show that large moisture transports from the Bay of Bengal to the Yangtze River catchment leads to heavy precipitation there. A negative temperature anomaly would support heavy precipitation in north China. Qian *et al.* (2002) find a connection between the northward progression of the warmest SST tongue in the northwestern Pacific and the onset of the Mei-yu rain. The Asian monsoon is influenced by thermal differences between ocean and continent. A large temperature gradient will generally lead to a strong summer monsoon (Chou, 2003). Bao (1987) explains the development of a thermal low over the Qinghai-Tibet-Plateau in summer and points towards the relevance of pressure conditions over this region for the monsoon circulation. Together with the subtropical high reaching from the Pacific to eastern China, the thermal low over the Qinghai-Tibet-Plateau leads to the inflow of warm and moist air masses from southeast to the inland and brings intensive rainfalls to China. Owing to a weakening of the thermal contrast between SSTs and the temperature over the Qinghai-Tibet-Plateau, an increase of snow cover and snow depth over the Qinghai-Tibet-Plateau in winter could weaken the intensity of the summer monsoon (Wu

* Correspondence to: Heike Hartmann, Department of Geography, Justus Liebig University, Senckenbergstr. 1, D-35390 Giessen, Germany.
E-mail: Heike.Hartmann@geogr.uni-giessen.de

and Qian, 2003). This leads to a decrease in precipitation in southern China and an increase in the Yangtze and Huaihe River catchments (Qian *et al*., 2003).

The Climate Prediction Center provides data and information (Climate Prediction Center, 2007) on several teleconnection indices, whose calculation is based on rotated principal component analysis (PCA) of monthly mean standardized 500 hPa geopotential height anomalies. Eurasia is affected by three prominent teleconnection patterns throughout the year, namely by the East Atlantic/Western Russia (EA/WR) pattern (referred to as Eurasia-2 pattern by Barnston and Livzey, 1987), the Scandinavia (SCA) pattern (referred to as Eurasia-1 pattern by Barnston and Livzey, 1987) and the Polar/Eurasia (POL) pattern. The EA/WR pattern consists of four main anomaly centres. The positive phase is connected with positive height anomalies located over Europe and northern China, and negative height anomalies over the central North Atlantic and north of the Caspian Sea. The positive phase of the pattern is generally associated with above average precipitation in eastern China. The SCA pattern has a primary circulation centre over Scandinavia and weaker centres of opposite sign over western Europe and eastern Russia/western Mongolia. A positive phase of the SCA pattern in summer is correlated with below average precipitation in central China and in China's southern inland region. In a positive phase, the POL pattern consists of negative height anomalies over the polar region, and positive anomalies over northern China and Mongolia. This pattern is linked to fluctuations in the strength of the circumpolar circulation, reflecting an enhanced circumpolar vortex in the positive phase. The positive phase of the POL pattern is correlated with above average precipitation in the Yangtze Delta region and with negative precipitation anomalies in China's southern inland regions.

The aim of the present study is to forecast summer rainfall in the Yangtze River basin using neural networks. It has been proven by several studies that the artificial neural networks are useful tools for predicting summer rainfall (e.g. Guhathakurta *et al*., 1999; Navone and Ceccatto, 1994; Sahai *et al*., 2003; Sahai *et al*., 2000; Venkatesan *et al*., 1997). Nevertheless, all of these studies deal with the Indian summer monsoon rainfall. The number of studies on forecasting Chinese summer rainfall by means of neural network analysis is limited. Wu *et al*. (2001) forecast summer precipitation in the Yangtze Delta on the basis of previous precipitation records and precipitation anomalies.

In the present study, the winter values of the following potentially influencing factors are used as input variables to forecast the summer rainfall in the whole Yangtze River basin: snow cover and snow depth over the Qinghai-Tibet-Plateau, Southern Oscillation Index (SOI), SSTs, sea level pressure (SLP), as well as several teleconnection indices, namely EA/WR pattern, SCA pattern and POL pattern.

The second goal of this study is to quantify the impact of various influencing factors on precipitation for different regions or, in other words, to quantify the teleconnections. This is essential for the prediction of precipitation patterns.

2. Data

All data series used in the present study consist of monthly averages or totals and cover the period from December 1993 to September 2002.

Snow data was provided by the Cold and Arid Regions Environmental and Engineering Research Institute (CAREERI), Lanzhou, PR China for the period from January 1993 to October 2002, and were based on Special Sensor Microwave Imager (SSM/I) data. We limited our analysis to the period of available snow data.

Further potentially influencing factors (predictors, input variables) which are used in this study are:

- SOI provided by the Climatic Research Unit, UEA, Norwich, UK from their Web site at http://www.cru.uea.ac.uk
- SSTs from the Bay of Bengal (BB) and the South and East China Sea (SCS and ECS respectively) from the dataset NOAA_ERSST_V2 (Smith and Reynolds, 2004) provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA from their Web site at http://www.cdc.noaa.gov
- SLP from NCEP/NCAR Reanalysis 1 (Kistler *et al*., 2001) provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA from their Web site at http://www.cdc.noaa.gov
- EA/WR pattern (referred to as Eurasia-2 pattern by Barnston and Livzey, 1987) provided by the Climate Prediction Center, Camp Springs, Maryland, USA from their Web site at http://www.nws.noaa.gov
- SCA pattern (referred to as Eurasia-1 pattern by Barnston and Livzey, 1987) provided by the Climate Prediction Center, Camp Springs, Maryland, USA from their Web site at http://www.nws.noaa.gov
- POL pattern (Barnston and Livzey, 1987) provided by the Climate Prediction Center, Camp Springs, Maryland, USA from their Web site at http://www.nws.noaa.gov

The location of the SST, SLP and snow grid cells which were selected for this study is shown in Figure 1.

The dependent (output) variables of our analyses are precipitation time series of 60 climate stations spread over the Yangtze River basin (Figure 1). The data were provided by the National Climatic Centre (NCC) of the China Meteorological Administration (CMA), Beijing, PR China. The homogeneity of the precipitation data has been confirmed in a previous study by Becker *et al*. (2007).

3. Methodology

The precipitation records were analysed by means of a PCA. PCA is a data reduction method which generates

DOI: 10.1002/joc

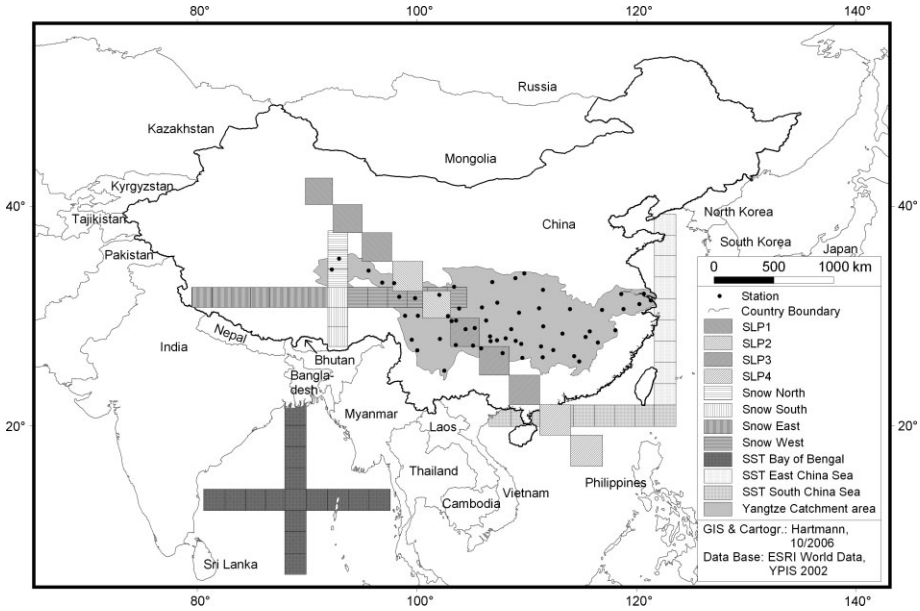

Figure 1. Location of the selected climate stations and grid cells.

a set of linearly independent spatial patterns (loadings) (Bordi *et al*., 2004). A varimax rotation of the principal components (PCs) was applied (Philipp, 2003). A maximum-loading approach was used in which each station is assigned to the PC upon which it loads most highly. Regions were classified around stations assigning to the same PC (Comrie and Glenn, 1998): Stations assigned to the first rotated PC are classified into region A, stations assigned to the second rotated PC are classified into region B etc. Instead of the rotated PC scores, we decided to use the monthly precipitation averages of different regions for further analyses. This procedure was preferred in order to keep a spatially continuous field of values with neighboring regions, including parts of common variability (Philipp and Jacobeit, 2002). Figure 2 shows the resulting rainfall regions which were generated by means of Inverse Distance Weighted (IDW) interpolation method.

The selected grid cells of SLPs and SSTs were classified into several indices according to biogeographical regions. SLP1 index covers the semi-arid and arid areas of northwestern China, SLP2 index is located over the eastern part of the Qinghai-Tibet-Plateau, SLP3 index covers the southern uplands and SLP4 index is located over the South China Sea. The SSTs are classified into three different indices: one for the Bay of Bengal, one for the South China Sea and one for the East China Sea. Snow data were grouped, North South, East West on the basis of their location. The monthly values for grid cells classified as one index were averaged and then the overall arithmetic mean was subtracted for each month. The result was divided by the overall standard deviation. Grid cells within one class are labelled by identical symbols (Figure 1).

In order to examine the relations between variables, we calculated correlation coefficients between input and output variables (Table II). However, linear models are poorly suited to model the complex and often non-linear relationships which exist between climate variables (Cannon and McKendry, 2002). Therefore, a neural network analysis was carried out which enables the modeling of even non-linear relationships. Neural networks are parallel computing structures of processing elements (neurons) which are interconnected by a network similar to the human brain (Hsieh and Tang, 1998). A conventional Multi-Layer Perceptron (MLP) neural network design was applied in the analyses. A MLP is a so-called 'feed-forward' neural network because all the information flows in one direction. The neurons of one layer are connected with the neurons of the following layer without feedback (Teschl and Randeu, 2006). The weights adjustment was performed by a back propagation algorithm: weights are modified to reduce the error occurrence between actual and desired network outputs backward from the output layer to the input layer (Backhaus *et al*., 2003).

In the present study, the input layer consisted of 15 input variables. The input variables were winter to spring (Dec−Apr) values of SST (three indices), SLP (four indices), snow depth (four indices), SOI, EA/WR pattern, SCA pattern and POL pattern for the period from 1993 to 2002. The architecture of the neural network was determined by a 'trail and error' approach. The starting point was a small network consisting of one hidden layer and two processing elements. The number of processing elements was increased until there was no improvement in the performance. This was achieved with four processing elements. Then the performance of two hidden layers was tested; however, no improvement was achieved. Thus, a neural network with one hidden layer and four processing elements was used. The output layer consisted of monthly precipitation totals from May to September for the period from 1994 to 2002. Rainfall is predicted from May to September 2002. Thirty five months of the available time series were used for training,

DOI: 10.1002/joc

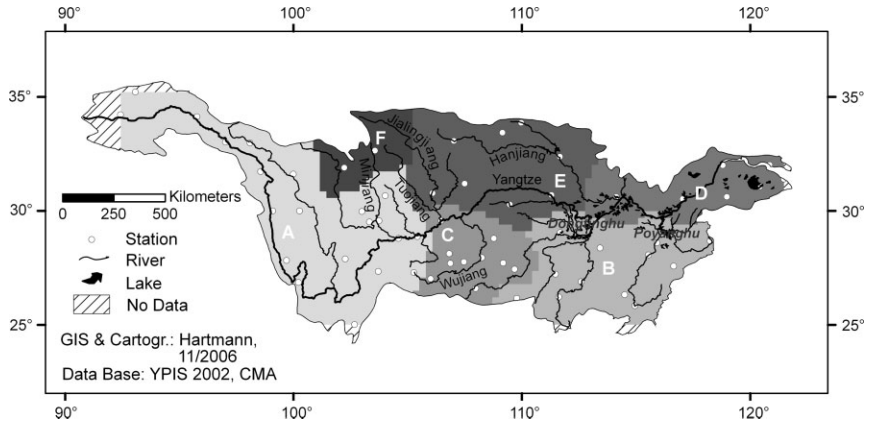

Figure 2. Location of the PCA based rainfall regions.

five months were used for cross validation and by that forecasting for five months was achieved.

A sensitivity analysis was carried out to measure the relative importance of each input variable. The testing process illustrates how the model output varies in response to variation of an input. The first input was varied between its mean and its standard deviation while all other inputs were fixed at their respective means. The network output was computed for 50 steps, above and below the mean. This procedure was repeated for each input (Principle *et al.*, 2005).

A common problem with neural network analysis is that the outcome of the sensitivity analyses can vary, depending on the random setting of the initial weights. Therefore, the neural network model run for one experiment was repeated five times. An input variable was accepted to be correlated to the output variable if a high sensitivity for it was found by all the model runs. The input variables, for which a high sensitivity was detected by all the model runs and thus the most important, were used as single input variables to test how much of the total output variance could be explained by the variance of these variables. Various combinations of input variables were tested if several input variables appeared to be highly correlated to the output variable. The results of this are shown only if they are more successful than the other tests.

4. Results: Principal component analysis

The first six rotated PCs explain altogether 80.9% of the total precipitation variance in the study region (Table I). The first rotated PC explains 54.5% of the total variance. The location of the six resulting rainfall regions can be taken from Figure 2. Rainfall region A covers large parts of the Yangtze River's upper reaches. The location of region B is around the Poyanghu and the Dongtinghu. Region C, which is mainly located south of the Yangtze, covers the Wujiang catchment. The Yangtze Delta and the Yangtze River's lower reaches are covered by region D. Region E covers large parts of the Hanjiang and the Jialingjiang catchments. The smallest region F is located in the upper reaches of the Minjiang and Jialingjiang.

Table I. Rainfall: Percentages of variance explained by the first six rotated PCs.

No. PCs	Variance (%)	Cumulative variance (%)
1	54.5	54.5
2	11.4	65.9
3	5	70.9
4	4.5	75.4
5	2.8	78.2
6	2.7	80.9

5. Results: Correlation analyses

The correlation coefficients between the input variables and the rainfall values for different regions can be taken from Table II. Also, results of the significance tests for the correlation coefficients are presented.

For region A, the highest (negative) correlation coefficient, significant at the 99% level, can be found between rainfall values and the SST index of the East China Sea. The highest (negative) correlation coefficients for the regions B, C and D, significant at the 99% level, are found between rainfall values and the SST Index of the Bay of Bengal. For region E, the highest (positive) correlation coefficient, significant at the 99% level, is detected between the rainfall values and the snow index of the southern Tibetan Plateau. Region F has the highest (positive) correlation coefficient, significant at the 95% level, between rainfall values and the SLP1 index.

6. Results: Neural network analyses

The results of the neural network analyses are summarized in Table III. Predicted rainfall of different regions (Figures 3 and 4) always refer to the most accurate results of five model runs.

DOI: 10.1002/joc

Table II. Correlation coefficients between the regional rainfall records and the input variables.

Region	SLP1	SLP2	SLP3	SLP4	SOI	SST_BB	SST_ECS	SST_SCS	Snow_E	Snow_N	Snow_S	Snow_W	EA/WR	SCA	POL
A	−0.1846	−0.3381*	−0.2172	−0.0317	−0.1507	−0.2550	−0.5566**	−0.4721**	0.1979	−0.0174	0.1232	0.1363	0.0552	−0.2479	−0.1165
B	0.6184**	0.5376**	0.5733**	0.6026**	−0.0310	−0.6288**	−0.4922**	−0.4313**	0.3990**	0.1253	0.1800	0.2369	−0.0455	0.0582	0.0098
C	0.5581**	0.4409**	0.5161**	0.5560**	−0.0267	−0.6697**	−0.5621**	−0.5249**	0.5873**	0.1846	0.2536	0.2524	0.0061	0.0994	0.1393
D	0.3438*	0.2417	0.2962*	0.3023*	−0.1364	−0.4962**	−0.4614**	−0.3949**	0.4929**	0.1114	0.1495	0.1616	−0.0654	0.1761	0.0088
E	0.1637	0.0602	0.1329	0.2316	0.0732	−0.3359*	−0.4558**	−0.3918**	0.3980**	0.3698*	0.4588**	0.3955**	0.1284	−0.0900	−0.0837
F	0.3349*	0.3249*	0.2739	0.2013	−0.0199	−0.1922	−0.0584	−0.0314	0.3313*	0.1781	0.2833	0.3056*	−0.0691	−0.0289	0.1098

* Significant at 95% confidence level
** Significant at 99% confidence level

For region A, the neural network analysis explains a large part of the total variance of rainfall on the basis of implementation of all available input variables. In addition, the absolute maximum error is small. A comparison of the measured values and the modeled output can be taken from Figure 3(a). The sensitivity analysis reveals that the SST index of the East China Sea is the most influential input variable. The results of the five model runs are in good agreement with each other. If the SST index is subsequently used as a single input variable, a large amount of the total variance can still be explained (Figure 3(b)). This clearly demonstrates the great importance of this input variable for the prediction of summer rainfall in region A.

For region B, a successful prediction of the summer precipitation is also possible if all input variables are implemented (Figure 3(c)). The result concerning the explained variance is even better than for region A; however, the maximum absolute error is higher. The sensitivity analysis shows again quite clearly that the SST index of the East China Sea is the most important input variable. The repetitions of the model runs confirm that the SST index of the East China Sea is classified as the most important input variable a further two times, whereas the SST index of the Bay of Bengal and the index SLP1 are each identified once as the most influential input variables. However, using the SST index of the East China Sea as a single input variable (Figure 3(d)) leads to a less satisfactory result in comparison with region A. The maximum absolute error is relatively high. Another neural network analysis is carried out with the SLP1 index, the SST index of the Bay of Bengal and the SST index of the East China Sea as input variables. The result is very good at predicting the rainfall (Figure 4(a)). This result is better than the outcome of the neural network analysis using all input variables. Sensitivity analysis shows that the most important input variable is the SST index of the East China Sea, followed by SLP1 index and the SST index of the Bay of Bengal. It can be concluded that although the SST index of the East China Sea is the most important input variable, the SLP1 index, as well as the SST index of the Bay of Bengal, are relevant for successful prediction of the rainfall in region B.

For region C, the prediction of summer rainfall, based on the use of all input variables, was not particularly successful (Figure 3(e)). Three of the five model runs show the SST index of the East China Sea as being the most important input variable, one indicating the SST index of the Bay of Bengal and one the SST index of the South China Sea. Limiting the analysis to only one input variable (SST index of the East China Sea) yields considerably better results (Figure 3(f)). All input variables, which have been identified to be important according to the correlation coefficients and the sensitivity analysis, are tested as single input variables, as well as in different combinations, to examine whether it is possible to model the peaks in June and August. However, all these forecasts are less accurate than the prediction based on the

DOI: 10.1002/joc

Table III. Results of the neural network analyses.

Region	Implementation of all input variables (five model runs)			Implementation of one input variable (five model runs)			Implementation of several input variables (five model runs)*			
	Maximum explained variance (%)	Maximum absolute error (mm)	Most important input variable	Maximum explained variance (%)	Maximum absolute error (mm)	Used input variable for best result	Maximum explained variance (%)	Maximum absolute error (mm)	Used input variables for best result	Most important input variable
A	77	27	5* SST_ECS	62	41	SST_ECS	–	–	–	–
B	90	67	3* SST_ECS, 1* SST_BB, 1* SLP1	64	95	SST_ECS	92	38	SST_ECS, SST_BB, SLP1	3* SST_ECS, 1* SST_BB, 1* SLP1
C	63	120	3* SST_ECS, 1* SST_BB, 1* SST_SCS	69	105	SST_ECS	–	–	–	–
D	72	61	4* SST_ECS, 1* SST_SCS	24	97	SST_ECS	79	59	SLP1, SLP4, SST_BB, SST_ECS, SST_SCS, Snow_E	2* SST_ECS, 1* SLP1, 1* SLP4, 1* SST_SCS
E	68	50	3* SST_ECS, 1* SLP1, 1* SLP4	4	70	SST_ECS	82	58	SLP1, SLP4, SST_BB, SST_ECS, SST_SCS, Snow_N	4* SST_ECS, 1* SLP4
F	78	41	3* SST_ECS, 1* SST_SCS, 1* SLP1	50	53	SLP1	–	–	–	–

* Results presented only if better than others

DOI: 10.1002/joc

SST index of the East China Sea. All in all, the prediction based on the SST index of the East China Sea only proved to be the most successful approach.

For region D, summer rainfall predictions show good agreement with the measured values (Figure 3(g)). The outcome of the sensitivity analysis points towards the SST index of the East China Sea as the most important input variable. Four out of five analyses confirm this finding. Only one analysis identifies the SST index of the South China Sea as the most influential input variable. Nevertheless, predicting rainfall by the SST index of the East China Sea as a single input variable is unreliable as can be seen from Figure 3(h). The most successful outcome is achieved by using the following input variables: SLP1, SLP4, SSTs Bay of Bengal, SSTs South China Sea, SSTs East China Sea and snow over the eastern Tibetan Plateau (Figure 4(b)). Two out of the five model runs point towards the SST index of the East China Sea as the most important input variable, whereas the SLP4 index, the SLP1 index and the SST index of the South China Sea are each identified once as the most influential input variable. It can be concluded that several indices are necessary for a successful prediction of the summer rainfall in region D and that it would be doubtful to name one input variable as the most important factor.

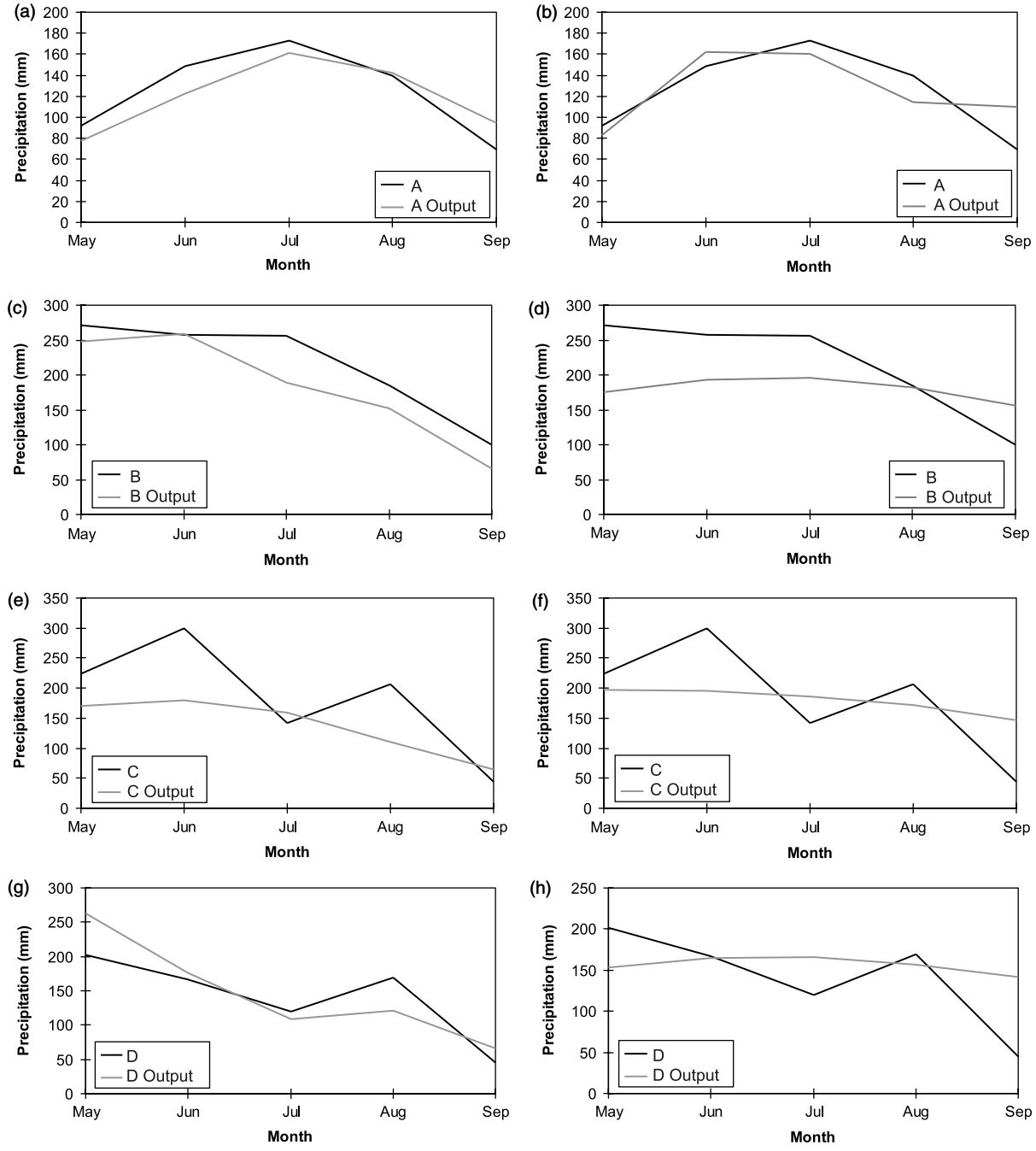

Figure 3. Measured (in black) and predicted (in gray) rainfall from May to September 2002 of the different regions modeled by 15 input variables (left) and by the most important input variable (right).

DOI: 10.1002/joc

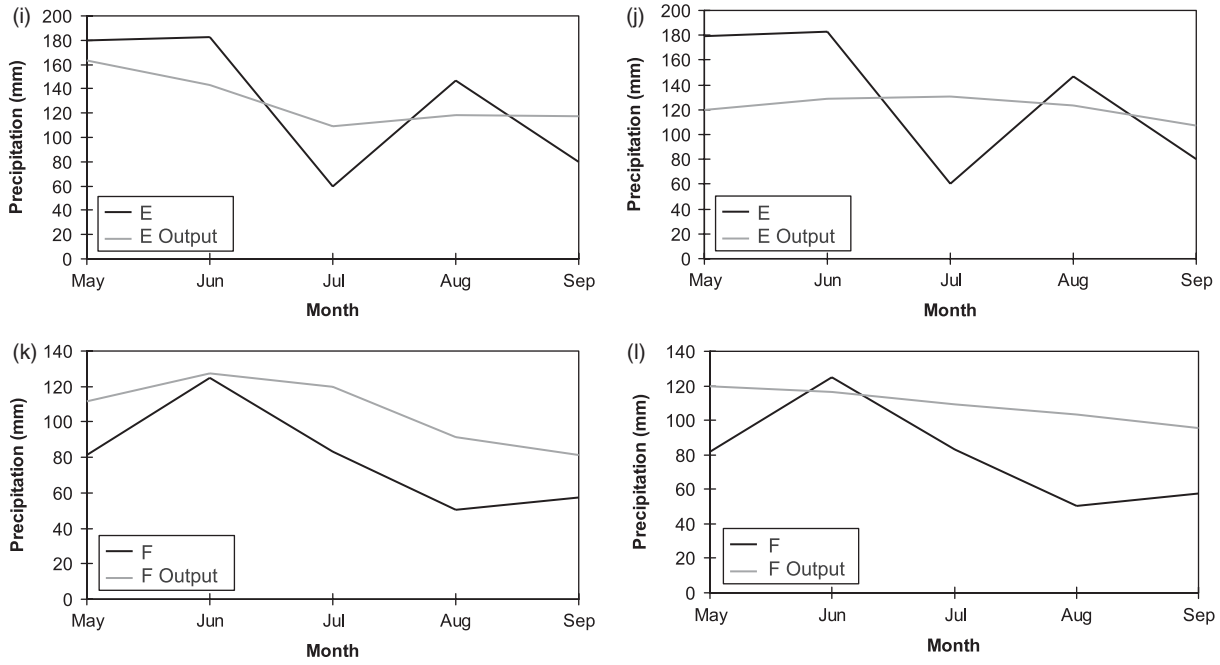

Figure 3. (*Continued*).

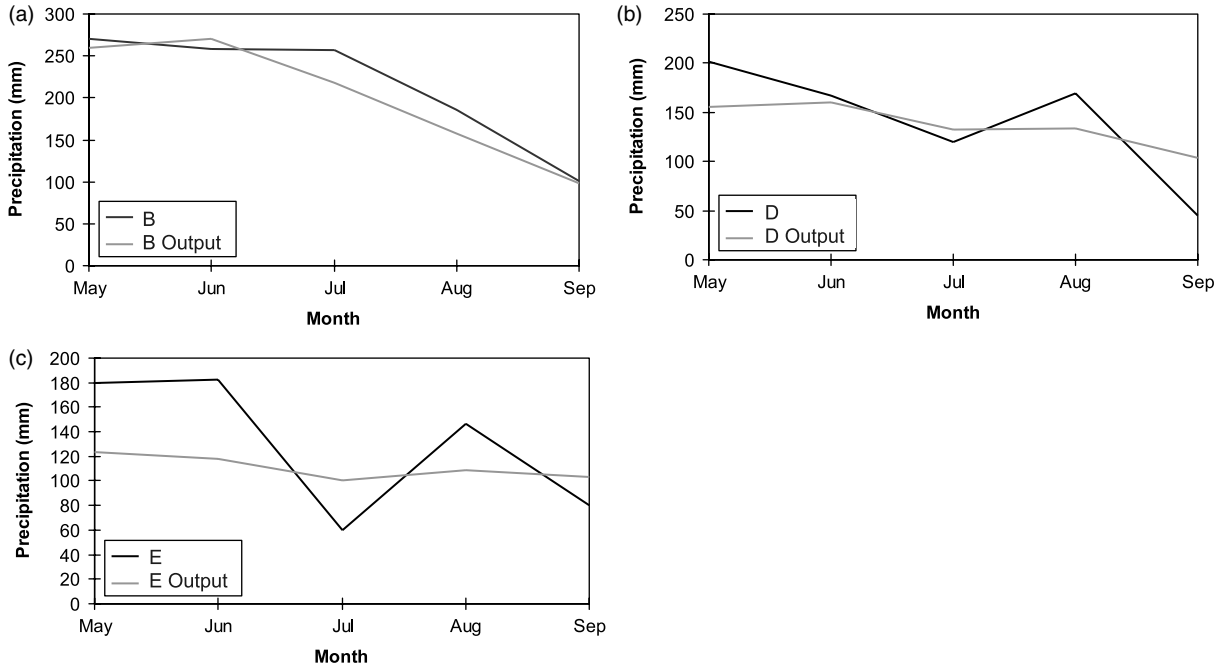

Figure 4. Measured (in black) and predicted (in gray) rainfall from May to September 2002 of different regions modeled by important input variables.

For region E, the prediction of the summer rainfall shows relatively good agreement with the measured values if all input variables are used (Figure 3(i)). However, the peaks in June and August are not predicted accurately. The sensitivity analysis identifies the SST index of the East China Sea as the most important input variable, which is confirmed by three out of the five model runs. Additional model runs point towards the relevance of the SLP1 index and the SLP4 index. The result of neural network analysis using the SST index of the East China Sea as the single input variable is not convincing (Figure 3(j)). It is neither possible to model the peaks in June and August nor the minimum in July. The best result is achieved on the basis of the predictors SLP1, SLP4, SSTs Bay of Bengal, SSTs South China Sea, SSTs East China Sea and snow over the northern Tibetan Plateau (Figure 4(c)). The experiment explains a larger part of the total variance but it also results in an absolute maximum

DOI: 10.1002/joc

error in May that is slightly larger than in the all-variables approach. The July minimum and the August peak are recognized correctly. The sensitivity analyses identify the SST index of the East China Sea as the predominately relevant input variable. We conclude that although the SST index of the East China Sea appears to be the most important input variable, several other indices are required for a successful prediction of the rainfall in region E.

For region F, the forecast of precipitation is successful in so far as the peak in June is well predicted, which would be most important for early flood warnings. Using all input variables, a large amount of the total variance can be explained (Figure 3(k)). The maximum absolute error is relatively high compared to the monthly total maximum. Three out of the five model runs identify the SST index of the East China Sea as the most important input variable; one points towards the SST index of the South China Sea and another one to the SLP1 index. To see how much of the variance can be explained by these three input variables, we use all of them individually. The most successful approach is based on the implementation of the SLP1 index (Figure 3(l)). The reliability of rainfall prediction for this region based on a combination of the three input variables is also tested. The result of this approach is less convincing than the result of the all-input-variables approach. The situation is comparable to that of region D. It can be concluded that several indices are required for successful prediction of the summer rainfall in region F and that it would not be meaningful to name one input variable as the most important factor.

7. Discussion and conclusions

We note that most previous studies rely on only one predictor for the explanation of summer rainfall variability in the Yangtze River catchment. The results of our analyses show that higher prediction accuracies can be achieved by implementing several predictors into the neural network analyses.

Nevertheless, not all the input variables have been proven to be valuable predictors. The teleconnection patterns failed to improve our predictions. This is due to the unpredictability of the indices several months ahead. In addition, winter teleconnection patterns are not further correlated with summer precipitation.

Gao *et al*. (2006) come to the conclusion that the significance of ENSO as a predictor of summer precipitation in China has weakened since the late 1970s. This finding is in agreement with our observation that the SOI is a weak predictor for summer precipitation patterns in the Yangtze River catchment. However, we have to take into account the short time period of our analysis (nine years), which makes it difficult to find connections between the factors showing dependent cycles of several years.

Wu and Qian (2003) explain that a winter of heavy snowfall over eastern Tibet or over southwestern Tibet is typically followed by a summer with weak summer monsoon circulation and consequently a high amount of rainfall in the Yangtze River valley. Our correlation analysis shows a coefficient of 0.5873 between snow over the eastern Tibetan Plateau and rainfall in region C. Only the SST index of the Bay of Bengal seems to be a more important predictor in this analysis. However, the result of neural network analysis identifies the SST index of the East China Sea as the most important input variable, followed by two other SST indices. Another relatively high correlation coefficient (compared with the other correlation coefficients) is found between snow over the southern Tibetan Plateau and rainfall in region E. However, the neural network analyses do not support this connection. Using this snow index as a single input factor allows an explanation of only 1.4% of the total rainfall variance. Therefore, we conclude that the Tibetan Plateau snow depth is a relatively weak predictor for summer rainfall in the Yangtze River basin.

The indices of the two climate factors SLP and SST – some of them being highly intercorrelated – appear to be the most valuable predictors for precipitation (Table IV).

The SLP indices are important for forecasting precipitation for regions B, D, E and F. The SLP1 index is an important predictor for forecasting the rainfall in region F, in particular. Chan and Shi (1999) use, among others, an index of the area of the subtropical high over the Pacific Ocean (110 °E–115 °W) as a predictor of summer monsoon rainfall over South China but do not give a synoptical explanation for this. Studies dealing with atmospheric pressure as the predictor of summer precipitation in China often contain hypotheses regarding the mechanisms which link SSTs and other indicators to rainfall variability in different regions. Xu (1995) focuses on the West Pacific high at 500 hPa and reasons that positive anomalies (Dec–Jan) of SSTs in the Kuroshio Current and the South China Sea lead to a northwestward extension of the West Pacific high. This would support the inflow of warm and moist air masses from southwest from the Bay of Bengal to the Yangtze and Huai River basins and would cause an earlier onset of the Mei-yu rain. Chang *et al*. (2000) come to a different conclusion. They argue that an increased moisture supply from the Bay of Bengal is at least partially offset by the 850 hPa wind anomalies, mostly directed away from the Mei-yu area in the wet season. They conclude that warm eastern Pacific SST anomalies in winter often precede a strong western Pacific subtropical ridge. The subsidence in the ridge area would be modulated by anomalous convection associated with the warm eastern Pacific SST anomaly causing a complementary cooling in the western Pacific through atmospheric Rossby wave responses and an evaporation–wind feedback. This ridge would extend further to the west from the previous winter to the consecutive fall, resulting in a 850 hPa anomalous anticyclone near China's southeast coast. This anticyclone would support heavy rainfall in the Yangtze River basin. In addition, it would induce anomalous warming of the SST of the South China Sea through increased downwelling, which would result in a higher moisture supply to the area of

DOI: 10.1002/joc

Table IV. Correlation coefficients between the input variables.

	SLP1	SLP2	SLP3	SLP4	SOI	SST_BB	SST_ECS	SST_SCS	Snow_E	Snow_N	Snow_S	Snow_W	EA/WR	SCA	POL
SLP1	1														
SLP2	0.9097**	1													
SLP3	0.9638**	0.9224**	1												
SLP4	0.8396**	0.7676**	0.9036**	1											
SOI	0.3255*	0.2534	0.3071*	0.1787	1										
SST_BB	−0.7741**	−0.5598**	−0.7340**	−0.7538**	−0.3758*	1									
SST_ECS	−0.4469**	−0.1756	−0.4050*	−0.5344**	−0.2894	0.8651**	1								
SST_SCS	−0.5192**	−0.2439	−0.4829**	−0.6114**	−0.3900**	0.8743**	0.9437**	1							
Snow_E	0.5276**	0.4424**	0.5243**	0.5682**	0.1090	−0.6405**	−0.5709**	−0.5213**	1						
Snow_N	0.3495*	0.3871**	0.3440*	0.2961*	0.1821	−0.3260**	−0.2752	−0.2367	0.5573**	1					
Snow_S	0.2093	0.2838	0.2416	0.3281*	−0.1060	−0.2141	−0.2217	−0.1234	0.6312**	0.7935**	1				
Snow_W	0.1386	0.2397	0.1716	0.2473	−0.3277*	−0.0841	−0.0893	0.0563	0.5414**	0.5449**	0.8802**	1			
EA/WR	−0.1176	−0.0683	−0.0414	−0.0121	−0.1922	0.0200	0.0332	0.0843	−0.0721	0.0712	0.2162	0.2769	1		
SCA	−0.0513	−0.0891	−0.1443	−0.2177	0.0235	0.0279	0.0132	0.0905	0.0810	−0.1416	−0.0856	0.0305	−0.0160	1	
POL	0.0725	0.1388	0.1603	0.1994	−0.0559	0.0178	0.1091	0.0558	0.0708	−0.0633	0.0262	0.0205	0.1858	0.0543	1

* Significant at 95% confidence level.
** Significant at 99% confidence level.

rainfall. However, not only the pressure constellations above the sea and the associated SST variability seem to be of importance for the prediction of rainfall, but also the pressure conditions over the Qinghai-Tibet-Plateau are important for the strength of the monsoon circulation (Bao, 1987), although their use as predictors seem to be of minor importance.

We conclude that there is sufficient evidence to support the suitability of SSTs as predictors of summer rainfall in China. On the basis of our analyses, the SSTs of the East China Sea are particularly important predictors, which explain up to 69% of the total rainfall variability in the region. The impact of the SST index of the Bay of Bengal, the SST index of the South China Sea and the SLP indices on the prediction is somewhat weaker, even though they are relevant for some regions and contribute to an increase in prediction accuracy in combination with the SSTs of the East China Sea. Other predictors such as the snow over the Qinghai-Tibet-Plateau and the SOI seem to be of less importance for the time period of our analysis.

We also conclude that neural network analysis is a suitable method for the prediction of summer rainfall in the Yangtze River basin. For all regions except region C, our forecasts explain at least 77% of the total variance of the real rainfall. We expect the implementation of additional input variables such as precipitation records of preceding years (Wu *et al*., 2001) to yield even better results. Region A covers large parts of the Yangtze River's upper reaches. A prediction of rainfall for this region is very important due to the fact that flood waves resulting from rainfall in that region can be intercepted by the Three Gorges Dam, if enough storage capacity can be provided. A good prediction of rainfall in the Yangtze River's upper reaches could contribute to an adjusted storage capacity management of the Three Gorges Reservoir, which could help to enhance the flood safety for the regions downstream of the dam. The results could allow for an earlier and better preparation for potential flood disasters and, therefore, a decrease of losses in properties and human lives.

Acknowledgements

This study has been carried out in the context of the projects KI 261/17-1 and BE 1962/6-1, funded by the German Science Foundation (DFG). Special thanks are due to our colleagues from the Cold and Arid Regions Environmental and Engineering Research Institute (CAREERI), Lanzhou for providing snow data and to Prof. Dr. Jiang Tong, Nanjing Institute for Geography and Limnology, Chinese Academy of Sciences (NIGLAS) for his help in obtaining precipitation data and for discussions. We would also like to thank Dr. Wang Run, Department of Geography, JLU Giessen and Dr. Zhang Qiang, NIGLAS, for their contribution to the data processing.

References

Backhaus K, Erichson B, Plinke W, Weiber R. 2003. *Multivariate Analysemethoden*. Springer: Berlin.

Bao C. 1987. *Synoptic Meteorology in China*. Springer: Berlin.

Barnston AG, Livzey RE. 1987. Classification, seasonality and persistence of Low-Frequency atmospheric circulation patterns. *Monthly Weather Review* **115**: 1083–1126.

Becker S, Hartmann H, Zhang Q, Wu Y, Jiang T. 2007. Cyclicity analysis of precipitation regimes in the Yangtze river basin, china. *International Journal of Climatology*, (in press).

Bordi I, Fraedrich K, Jiang JM, Sutera A. 2004. Spatio-temporal variability of dry and wet periods in eastern China. *Theoretical and Applied Climatology* **79**: 81–91, DOI: 10.1007/s00704-004-0053-8.

Cannon AJ, McKendry IG. 2002. A graphical sensitivity analysis for statistical climate models: application to Indian monsoon rainfall prediction by neural networks and multiple linear regression models. *International Journal of Climatology* **22**: 1687–1708.

Chan JCL, Shi JE. 1999. Prediction of the summer monsoon rainfall over south China. *International Journal of Climatology* **19**: 1255–1265.

Chang CP, Zhang Y, Li T. 2000. Interannual and Interdecadal variations of the east Asian summer monsoon and tropical pacific SSTs. Part I: Roles of the subtropical ridge. *Journal of Climate* **13**: 4310–4325.

Chou C. 2003. Land-sea heating contrast in an idealized Asian summer monsoon. *Climate Dynamics* **21**: 11–25, DOI: 10.1007/s00382-003-0315-7.

Climate Prediction Center. 2007. *Northern Hemisphere Teleconnection Patterns*. Available: http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml, Accessed on May 2, 2007.

Comrie AC, Glenn EC. 1998. Principal components-based regionalization of precipitation regimes across the southwest United States and northern Mexico, with an application to monsoon precipitation variability. *Climate Research* **10**: 201–215.

Gao H, Wang Y, He J. 2006. Weakening significance of ENSO as a predictor of precipitation in China. *Geophysical Research Letters* **33**: L09807, DOI: 10.1029/2005GL025511.

Guhathakurta P, Rajeevan M, Thapliyal V. 1999. Long range forecasting Indian summer monsoon rainfall by a hybrid principal component neural network model. *Meteorology and Atmospheric Physics* **71**: 255–266.

Guo Y, Zhao Y, Wang J. 2002. Numerical simulation of the Relationships between the 1998 Yangtze river valley flood and SST anomalies. *Advances in Atmospheric Sciences* **19**: 391–404.

Hsieh WW, Tang B. 1998. Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society* **79**: 18551870.

Jiang T, Zhang Q, Zhu D, Wu Y. 2006. Yangtze floods and droughts (China) and teleconnections with ENSO activities (1470–2003). *Quaternary International* **144**: 29–37, DOI: 10.1016/j.quaint.2005.05.010.

Kistler R, Kalnay E, Collins W, Saha S, White G, Woollen J, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, van den Dool H, Jenne R, Fiorino M. 2001. The NCEP_NCAR 50-Year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society* **82**: 247–267.

Navone HD, Ceccatto HA. 1994. Predicting Indian monsoon rainfall: a neural network approach. *Climate Dynamics* **10**: 305–312.

Philipp A. 2003. *Zirkulationsdynamische Telekonnektivität des Sommernbiederschlags im südhemispherischen Afrika*. PhD dissertation, University of Wuerzburg, Wuerzburg.

Philipp A, Jacobeit J. 2002. Comprehensive modes of teleconnections between southern African summer rainfall and the large-scale atmospheric circulation. In *Geowissenschaftliche Untersuchungen in Afrika IV*, *Wuerzburger Geographische Arbeiten 97*. Sponholz B (ed.). Selbstverlag des Instituts für Geographie: Wuerzburg.

Principe J, Lefebvre C, Lynn G, Fancourt C, Wooten D. 2005. *NeuroSolutions–Documentation*. NeuroDimension, Inc.: Gainesville, FL.

Qian W, Kang HS, Lee DK. 2002. Distribution of seasonal rainfall in the East Asian monsoon region. *Theoretical and Applied Climatology* **73**: 151–168, DOI: 10.1007/s00704-002-0679-3.

Qian YF, Zheng YQ, Zhang Y, Miao Q. 2003. Responses to China's summer monsoon climate to snow anomaly over the tibetan plateau. *International Journal of Climatology* **23**: 593–613, DOI: 10.1002/joc.901.

Sahai AK, Soman MK, Satyan V. 2000. All India summer monsoon rainfall prediction using artificial neural network. *Climate Dynamics* **16**: 291–302.

Sahai AK, Pattanaik DR, Satyan V, Grimm AM. 2003. Teleconnections in recent time and prediction of Indian summer monsoon rainfall. *Meteorology and Atmospheric Physics* **84**: 217–227, DOI: 10.1007/s00703-002-0595-1.

Smith TM, Reynolds RW. 2004. Improved extended reconstruction of SST (1854–1997). *Journal of Climate* **17**: 2466–2477.

Teschl R, Randeu WL. 2006. A neural network model for short term river flow prediction. *Natural Hazards and Earth System Sciences* **6**: 629–635.

Venkatesan C, Raskar SD, Tambe SS, Kulkarni BD, Keshavamurty RN. 1997. Prediction of all India summer monsoon rainfall using Error-back-propagation neural networks. *Meteorology and Atmospheric Physics* **62**: 225–240.

Wu TW, Qian ZA. 2003. The relation between the Tibetan winter snow and the asian summer monsoon and rainfall: An observational investigation. *Journal of Climate* **16**: 2038–2051.

Wu YJ, Gough WA, Jiang T, Kung HAT. 2006. The variation of floods in the middle reaches of the Yangtze River and ist teleconnections with El Niño events. *Advances in Geosciences* **6**: 201–205.

Wu X, Cao H, Flitman A, Wei F, Feng G. 2001. Forecasting monsoon precipitation using artificial neural networks. *Advances in Atmospheric Sciences* **18**: 950–958.

Xu Q. 1995. Analysis of causes and seasonal prediction of the severe floods in Yangtze/Huaihe basins during summer 1991. *Advances in Atmospheric Sciences* **12**: 215–224.

Zhao M. 1999. Analysis of the 1998 Yangtse river flood and consideration of flood control strategy. In *Proceedings of the '99 International Symposium on Flood Control*, November 11–13, Beijing: 244–253.

DOI: 10.1002/joc