Drought prediction in the Vietnamese central highlands

TINH DANG NGUYEN[1], DAN ROSBJERG[2], CINTIA UVO[3] & KIM QUANG NGUYEN[1]

1 *Faculty of Planning & Management of Water Resources Development Systems, Water Resources University, 175 Tayson, Dongda, Hanoi, Vietnam*
dr@er.dtu.dk

2 *Institute of Environment & Resources, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark*

3 *Department of Water Resources Engineering, Lund University, SE-221 00 Lund, Sweden*

Abstract Rainfall over the Vietnamese central highlands is governed by the Asian summer monsoon. The El Niño-Southern Oscillation (ENSO) phenomenon and related large-scale circulation anomalies, however, introduce disturbances that may lead to drought in the central highlands. Droughts cannot be prevented, but the consequences for human livelihood and economic losses can be alleviated if better prediction tools become available. Sea surface temperature (SST) is an indicator for the ENSO phenomenon and has been used here in an effort to develop prediction models for precipitation in the Vietnamese central highlands with a lead-time of up to three months. SST in both the Indian and Pacific oceans has been related to precipitation by means of canonical correlation analysis, a linear statistical technique. The best results were obtained for rainfall at the outset and at the end of the rainy season. Nonlinear techniques in the form of artificial neural networks (ANN) were subsequently applied. Additionally, discharge in three river basins in the central highlands was predicted with SST and meteorological variables as predictors. Although local effects have a considerable influence in certain parts of the area, reasonable prediction results were obtained for both rainfall and discharge.

Key words artificial neural networks; canonical correlation analysis; drought prediction; ENSO; sea surface temperature

INTRODUCTION

The Vietnamese central highlands (VCH) region is located in the southeastern part of the Indochina Peninsula. It is embedded in the Asian monsoon system and flanked by the Indian and Southeast Asian monsoons (Chen *et al.*, 2000). Water is intensively used to sustain livelihood and agricultural production, the most important factor for socio-economic well-being in the region. Lack of water affects ecosystems, agricultural production, power generation, livelihood and urban areas, among others.

The forecast of monthly summer monsoon rainfall over the VCH is significant for planning and management of reservoirs, agricultural practice, and especially for drought mitigation. Medium-range planning and management of water resources require good monthly or seasonal rainfall forecasts well in advance.

Statistical forecast techniques have been rapidly improved over the last two decades. Multiple linear regression models have been commonly used for hydrometeorological analysis and drought forecasting (Lui *et al.*, 1998; Cordery &

McCall, 2000) and Indian monsoon rainfall forecasting (Hastenrath, 1988; Kumar *et al.*, 1995). Canonical correlation analysis (CCA) (Shabbar & Barnston, 1996; Uvo & Graham, 1998; Ntale *et al.*, 2003) and artificial neural networks (ANN) (Vemuri & Rogers, 1994; Uvo *et al.*, 2000) are important methods for establishing linear and nonlinear relationships between variables. They have been used successfully to forecast field variables (such as precipitation and air temperature) based on other fields (such as SST or sea level pressure).

Since significant links between ENSO and the Indian monsoon are identified by, e.g. Rasmusson & Carpenter (1983), empirical models for monsoon prediction have developed rapidly and are available in, e.g. Kumar *et al*. (1995), Rajeevan (2001). Shukla & Mooley (1987) identify Indian monsoon predictors related to ENSO and develop statistical models for forecasting summer monsoon rainfall over India. Within the framework of these predictors, statistical (Rajeevan, 2001) and ANN models (Sahai *et al.*, 2000) have been developed for operational forecasting at the India Meteorological Department Monograph Climatology. Recently, Singhrattna (2003) indicates that Thailand summer rainfall is strongly correlated with pre-monsoon land-surface temperatures (March to May), which represent the land-ocean thermal gradient. Singhrattna *et al.* (2005) investigate the possibility of predicting the SASM rainfall over Thailand from large-scale variables related to ENSO such as SST and sea level pressure.

The purpose of this study is to develop statistical linear and nonlinear models to estimate monthly rainfall and discharge, respectively, at selected stations over the VCH based on SST over the tropical Pacific and Indian oceans, and other, local variables. Description of the research region and data are presented in the next section, then the methodology applied is described, followed by the results of the forecast models and discussion. Finally, the conclusions are drawn.

RESEARCH AREA AND DATA

Research area description

The central highlands region comprises approximately five million ha of rugged mountain peaks, widespread forests, and flat plateaus of basaltic land. The highlands account for 16% of total arable and 22% of total forested land in Vietnam. The inter-annual variations of the rainfall are mainly influenced by seasonal winds. The annual rainfall ranges from 1500 to 2000 mm, and the rainy season accounts for about 80% of annual rainfall. The discharge usually peaks in October and November following massive rainfall.

The whole research area is displayed in Fig. 1, with 24 selected gauging stations, where 16 stations have rainfall records, five stations are measuring rainfall and other meteorological variables such as air temperature, humidity and wind speed, and three are hydrological stations for discharge data in the three main rivers: Sesan, Srepok and Ba. Sesan and Srepok rivers flow from mountainous areas with average elevation ranging from 500 to 1000 m a.s.l., both basins spread over areas in the VCH and Cambodia, and the rivers flow westward toward the Mekong River system in Cambodia. The Ba River catchment area is totally located in Vietnam. The river originates at the northern mountainous area of the VCH with average elevation ranging from 300 to 500 m a.s.l., and flows eastward through the lowland to the South China Sea.

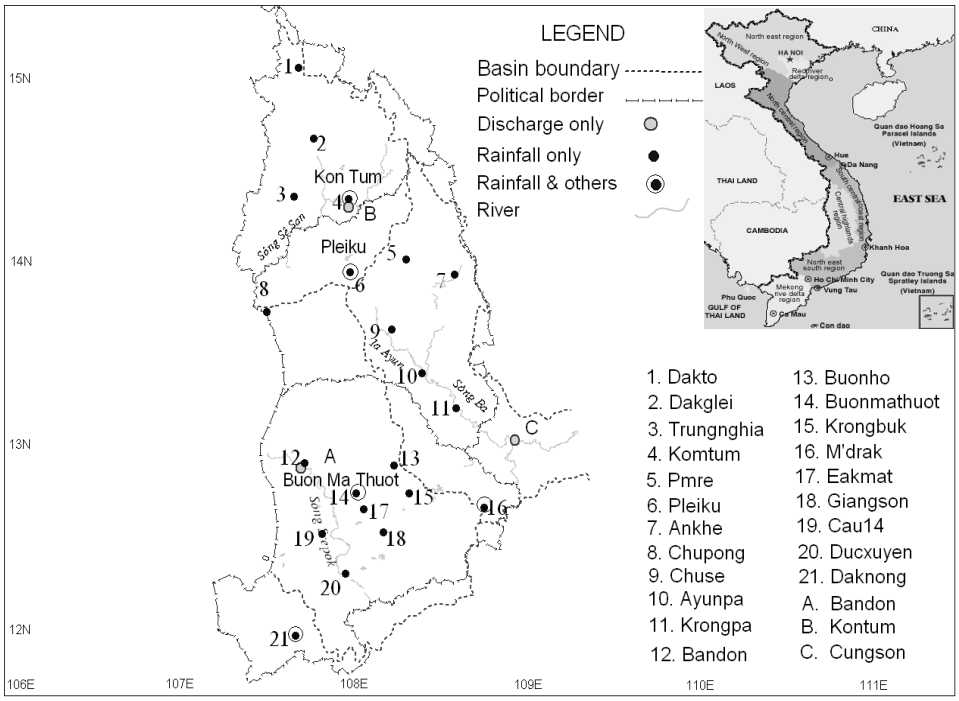

Fig. 1 The distribution of the selected raingauge stations/topography over the central highlands, Vietnam.

Data

Precipitation The data set used in Canonical Correlation analysis (CCA) includes monthly precipitation from 21 selected gauging stations over the VCH covering the period from January 1980 to December 2000. The precipitation data was primarily compiled by the National Hydrometeorology Institute of Vietnam. Most of the rainfall records are complete, while some have less than 10% missing data. A multiple linear temporal regression was used to fill missing data, when the data series contained more than 3% of missing data or a full year of missing data. In other cases, missing data were filled by inserting the appropriate long term mean.

Discharge Data sets used as input to Artificial Neural Network (ANN) models are monthly discharge at a site in each catchment, Kontum (Sesan catchment), Bandon (Srepok catchment) and Cungson (Ba catchment). The data values span the period January 1980 to December 2000 in which the data sets are divided into three subsets, 13 years for calibration, the next 3 years for checking and the last 5 years for validation.

SST For the interval of January 1980 to December 2000, the average monthly SST were obtained from the Comprehensive Ocean Atmosphere Data Set (COADS) (Reynolds *et al.*, 2002) extracted from the data library of the International Research Institute for Climate and Society—IRI. Tropical SST data were selected in the Pacific and Indian oceans, i.e. 28°N–28°S; 120°E–85°W and 28°N–28°S; 30°E–105°E, respectively. The data, available at a 2° × 2° latitude–longitude grid point resolution, were transformed to a 4° × 4° latitude–longitude resolution by simple average, and were used as input of CCA models. Moreover, meteorological data such as air temperature, humidity, wind speed, etc. were also used as ANN model input.

All time series were standardized, prior to use, by extracting the long-term mean value and dividing by the standard deviation. For ANN models, after standardization as suggested by Dawson & Wilby (2001), all data series were transformed to the interval (–0.9, 0.9).

METHODOLOGY

Canonical Correlation Analysis (CCA)

CCA is a linear multivariate technique. An early version was introduced by Hotelling (1935, 1936), and the technique has been used widely in the social sciences since the 1960s (Bretherton *et al.*, 1992). The technique isolates the linear combination of the data from one field and the linear combination of the data from another field that are maximally correlated. As simple linear regression, CCA minimizes squared errors in hindcasting the predictand from the predictors (Barnett & Preisendorfer, 1987). The CCA has been commonly used for both diagnostics (Nicholls, 1987; Díaz *et al.*, 1998) and prediction studies (Glahn, 1968; Barnett, 1981; Barnett & Preisendorfer, 1987; Barnston & Ropelewski, 1992; Gershunov & Cayan, 2003; Ntale *et al.*, 2003). CCA is able to define the highest spatial and temporal evolution of the predictor field that best predicts the predictand field. We use CCA to forecast monthly rainfall over the VCH during the rainy season using tropical Pacific and Indian Oceans SST as predictors. A thorough explanation about how a CCA model is developed can be found in Barnett & Preisendorfer (1987).

The diagnosis was made by means of CCA to show the temporal and spatial evolution in the predictor field that gives the best forecast skill to the predictand field, the so-called canonical predictor pattern (g-map). The g-map is the linear combination in predictor data that contributes most hindcast skill to the predictand data set. The analysis was made from monthly SST two months prior to the monthly rainfall, e.g. April rainfall forecast was made from February SST over the Pacific, and November precipitation forecast from September SST over the Indian Ocean.

Artificial Neural Networks (ANNs)

The ANN ability to imitate the performance of the human brain is well known. Conceptually, an ANN is composed of simple elements working in parallel, which are inspired by the biological nervous system. Mathematically, ANN is a theorized mind model in which the network function is determined by interconnecting elements in the system. Thus, a neural network can map a nonlinear complex relationship through a training process by adjusting the weights that link interconnected elements in the network.

ANNs have been used in climate and hydrological sciences during recent years. Dawson & Wilby (2001) present a thorough review of the use of ANNs in hydrological modelling. Many studies explore the forecasting potential of ANNs, e.g. hourly discharge (Deo & Thirumalaiah, 2000), daily discharge (Phien & Danh, 1997; Brikundavyi *et al.*, 2002; Cigizoglu, 2004), monthly discharge (Phien & Siang, 1993;

Salas *et al.*, 2000), seasonal discharge (Uvo *et al.*, 2000), hourly rainfall (Olsson *et al.*, 2004), and monthly rainfall (Freiwan & Cigizoglu, 2005). Successful achievements were obtained through the studies implying that ANNs can provide good forecast when appropriate choices of network types, training methods and data handling techniques are identified.

ANN design The most appropriate design of an ANN is normally found after many trial and error tests. In this study, the best accuracy was achieved when using a multi-layer feed-forward neural network. The ANN is composed of three layers, one hidden layer and one output layer containing one neuron each, and an input layer containing as many neurons as the number of input variables. A hyperbolic tangent sigmoid transfer function is used for the neurons of the hidden layer and output layer in the discharge forecast. The ANN was developed using MATLAB.

The mathematical expression of the hyperbolic tangent sigmoid transfer function is as follows:

$$a = \frac{e^{n} - e^{-n}}{e^{n} + e^{-n}}$$

where n is input and a is the output.

ANN training During the training process, the ANN weights and bias factors are estimated using the scaled conjugate algorithm. The scaled conjugate algorithm shows super linear convergence by using a step size scaling mechanism to avoid time-consuming line-search per learning iteration, which is assumed to be a sufficiently accurate and robust technique to train the network (Moller, 1993).

The critical issue in training an ANN is avoiding overfitting, as it reduces its capacity of generalization. If too many layers and neurons are used, the network has too many free parameters and may overfit the data. In contrast, if too few layers and neurons are included in the network, it might not be possible to fully detect the signal and variance of a complex data set. Thus, an optimal ANN needs to be designed by considering the number of layers and number of neurons in each layer and an appropriate transfer function and training technique. The present study adopted a stopping rule corresponding to an error between one interaction and the next as small as 0.001, or after 300 epochs. In this way overfitting can be avoided so that a satisfactory generalization can be reached.

Model verification The quality of the results obtained from the model verification was expressed by the correlation coefficient between observed and estimated values, as well as by the root mean squared error (RMSE). The correlation coefficient shows a skilful training performance, and the RMSE presents, on average, how close observed and estimated values are. The correlation coefficient is defined as the Pearson product-moment coefficient of linear correlation between two variables. It ranges in the interval (–1, 1), where higher positive values indicate better models. Conceptually, the best results are obtained by the highest correlation coefficient and the lowest value of RMSE. The correlation coefficient and RMSE are mathematically expressed as:

$$Corre.coeff = \frac{\sum_{i=1}^{n}(Y_{iObs} - \alpha)(Y_{iEst} - \beta)}{\sqrt{\sum_{i=1}^{n}(Y_{iObs} - \alpha)^2 \sum_{i=1}^{n}(Y_{iEst} - \beta)^2}} \quad \text{and} \quad RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_{iObs} - Y_{iEst})^2}{n}}$$

where Y_{iObs}, Y_{iEst} are the observed and estimated value of point i, respectively; α, β, the mean of observed and estimated time series, respectively; and n, the size of the time series (number of data points).

RESULTS AND DISCUSSION

Precipitation

The results of monthly precipitation forecast based on SST from the Pacific and Indian oceans using statistical models are presented in this section. The models were developed to estimate monthly rainfall through the VCH rainy season based on monthly SST over the Pacific or Indian oceans two months in advance. Table 1 shows correlation coefficients between standardized observed and estimated monthly rainfall series, where the latter was obtained from cross-validation of the CCA models, at all sites studied over the VCH. Correlation coefficients above 0.45 are statistically significant at levels >95%, and above 0.53 at levels >99%. For simplification, only correlation coefficients for the most skilful combinations of predictor and predictand are presented.

Table 1 Correlation coefficients between standardized observed and estimated precipitation are obtained by cross-validation in CCA models. * and ** indicate correlation significant at >95% and >99%, respectively. N, C, and S indicate station in the north, centre and south of the region, respectively.

SST over	Pacific	Pacific	Pacific	Indian	Pacific	Pacific	Pacific	Indian
Precipitation	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
SST in	Feb	Mar	May	Apr	Jun	Jul	Aug	Sep
Dakglei (N)	0.21	–0.26	0.27	0.05	0.02	0.28	0.22	0.48*
Dakto (N)	0.63**	0.08	0.46*	0.29	–0.13	0.35	0.23	0.64**
Trungnghia (N)	0.17	0.12	0.31	0.27	–0.14	–0.03	–0.01	–0.17
Kontum (N)	0.07	0.18	0.50*	0.23	–0.26	–0.09	0.41	0.61**
Pmre (C)	0.59**	0.33	0.31	0.10	0.23	0.21	0.38	0.69**
Pleiku (C)	0.60**	0.34	0.44	0.23	0.29	–0.21	0.19	0.60**
Ankhe (C)	0.12	0.16	–0.18	–0.07	0.45*	–0.21	0.40	0.26
Chupong (C)	0.52*	0.19	0.37	–0.15	0.21	–0.34	0.13	0.31
Chuse (C)	0.51*	0.29	0.23	0.00	–0.25	0.04	0.20	0.60**
Ayunpa (C)	0.31	0.00	0.08	0.26	–0.12	0.15	0.36	0.45*
Krongpa (C)	0.63**	0.12	0.16	0.55**	0.24	–0.33	0.34	0.37
Bandon (S)	0.09	0.08	0.06	0.21	0.51*	0.07	0.30	0.53*
Buonho (S)	0.60**	0.40	0.41	0.19	–0.19	0.55**	0.33	0.52*
Buonmathuot (S)	–0.46*	–0.05	0.37	0.15	–0.17	0.31	0.47*	0.10
Krongbuk (S)	0.41	–0.06	0.30	0.45*	–0.21	–0.02	0.46*	0.23
Eakmat (S)	0.58**	0.03	0.05	0.15	0.30	–0.23	0.39	0.47*
M'drak (S)	0.37	–0.01	0.18	–0.39	0.40	0.16	0.61**	0.29
Cau14 (S)	0.53*	0.50*	–0.25	0.13	0.43	0.30	0.54**	0.50*
Giangson (S)	0.39	0.12	–0.16	–0.14	0.21	–0.07	0.39	0.49*
Ducxuyen (S)	0.66**	–0.14	0.17	0.23	0.55**	0.52*	0.46*	0.43
Daknong (S)	0.67**	0.15	0.28	0.05	0.13	0.31	–0.02	0.39

In general, the quality of the results obtained from the cross-validation varies spatially and temporally during the rainy season. The best correlation coefficients are found for the first and the two last months of the rainy season (April, October and November). As can be seen from Table 1, the Pacific SST is a good predictor for April rainfall at most sites and for October rainfall at some sites two months in advance. On the other hand, Indian Ocean SST is a skilful predictor for November precipitation at most sites, also two months in advance.

Figure 2 shows the time series of standardized observed and forecasted precipitation in April at some sites, obtained from the CCA cross-validation using the Pacific Ocean SST in February as predictor. The correlation coefficient varies from 0.51 to 0.67. In general, the model reproduces well the variation in the sign of a monthly rainfall anomaly. Intensities, however, are normally underestimated. The model generates better forecasts for monthly rainfall anomalies during the 1990s for most sites of the VCH, but it remarkably underestimates the extreme rainfall anomalies in

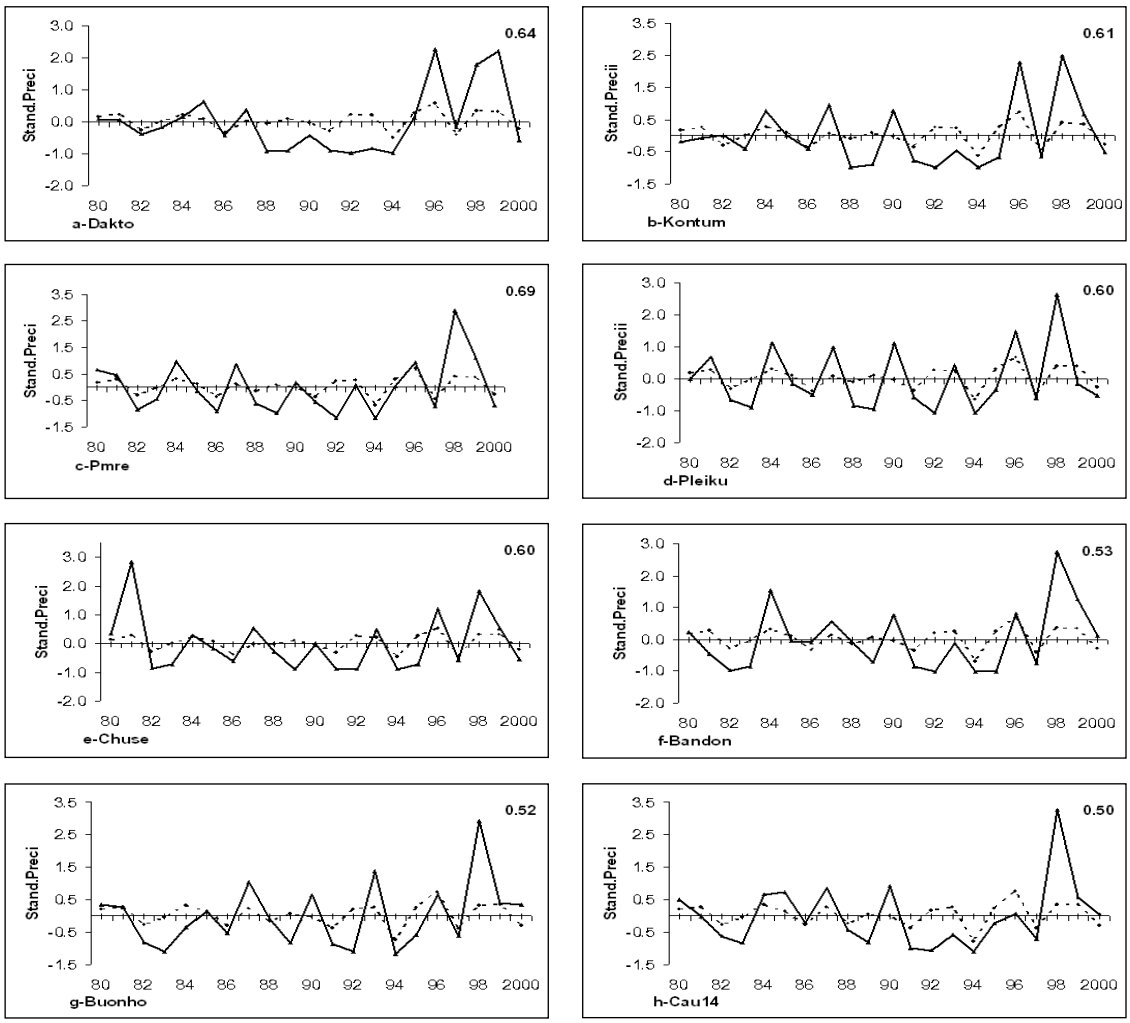

Fig. 2 The time series of standardized observed (solid line) and estimated (dashed line) precipitation in April obtained from cross-validation model using SST in the Pacific Ocean as predictor. The correlation coefficients between two time series presented on the top right of each graph are statistically significant at levels >95%.

1996 and 1997 for Pmre (Fig. 2(b)), in 1997 for Cau14 (Fig. 2(i)), Ducxuyen (Fig. 2(j)) and Daknong (Fig. 2(k)), and in 1999 for Chupong (Fig. 2(d)). The model is able to forecast well rainfall anomalies during the early to middle 1990s for the sites of Cau14, Ducxuyen, Daknong in the southern VCH, and from the middle to late 1990s for Dakto in the northern and Pleiku and Chupong in the central VCH.

For most cases the model cannot reproduce well the intensity of the extreme precipitation anomalies, despite the fact that they are mainly associated with ENSO events (e.g. 1987, 1992 and 1998). However, it is clear that the model, when using SST over the Pacific as predictor, is able to indicate dry or wet conditions at the beginning of the rainy season, two months in advance. This information is highly valuable for decision-makers in relation to agriculture management, among others.

In Fig. 3, time series of standardized observed and forecasted precipitation in November at some sites are presented, obtained from the CCA cross-validation using Indian Ocean SST in September as predictor. The eight selected stations include two sites in the northern (Dakto, Komtum shown in Fig. 3(a), (b)), three in the central (Pmre, Pleiku and Chuse in Fig. 3(c)–(e)) and three in the southern VCH (Bandon, Buonho and Cau14 shown in Fig. 3(f)–(h)). The best correlation coefficient between observed and estimated precipitation is 0.69 at Pmre (Fig. 3(c)) and the worst is 0.50 at Cau14 (Fig. 3(h)). Like in the beginning of the rainy season, the model is also able to indicate wet or dry conditions at the end of the rainy season, but not its intensity. This information, however, is valuable for water management, given that it is provided two months before the end of the rainy season.

Results obtained for May to October do not possess the same quality as those for the beginning and the end of the rainy season. However, some interesting remarks can be made. As shown in Table 1, a correlation coefficient of 0.5 between observed and forecasted precipitation was obtained at Cau14 for precipitation in May forecasted from Pacific Ocean SST in March. In June, correlation coefficients of 0.46 and 0.5 are obtained for precipitation at Dakto and Komtum, both located in the northern VCH, based on Pacific Ocean SST in May. For July, the best correlation coefficients 0.55 and 0.45 are obtained for Krongpa and Krongbuk, based on April Indian Ocean SST. In August, the best correlation coefficients 0.45, 0.51 and 0.55 are obtained for sites located in the centre (Ankhe) and the south (Bandon, Ducxuyen), based on SST over the Pacific Ocean two months in advance. In September, the best correlation coefficients 0.55 and 0.52 are obtained for two sites in the south (Buonho, Ducxuyen). In October, significant results are obtained for five selected stations with correlation coefficients varying between 0.46 and 0.61 The best correlation coefficients are obtained for the sites Buonmathuot (0.47), Krongbuk (0.46), M'drak (0.61), Cau14 (0.54) and Ducxuyen (0.46) that are all located in the south of the VCH.

Overall, the model using SST in the Pacific Ocean up to two months in advance reproduces the sign of the precipitation anomaly, but not its intensity, at most sites in the south of the VCH at the beginning of the rainy season. The best rainfall forecasts at the end of the rainy season are generated from the model using SST in the Indian Ocean up to two months in advance as the predictor, and found at sites confined to the north and central VCH. As in previous cases, the model is best in indicating the sign of the precipitation anomaly, not its intensity.

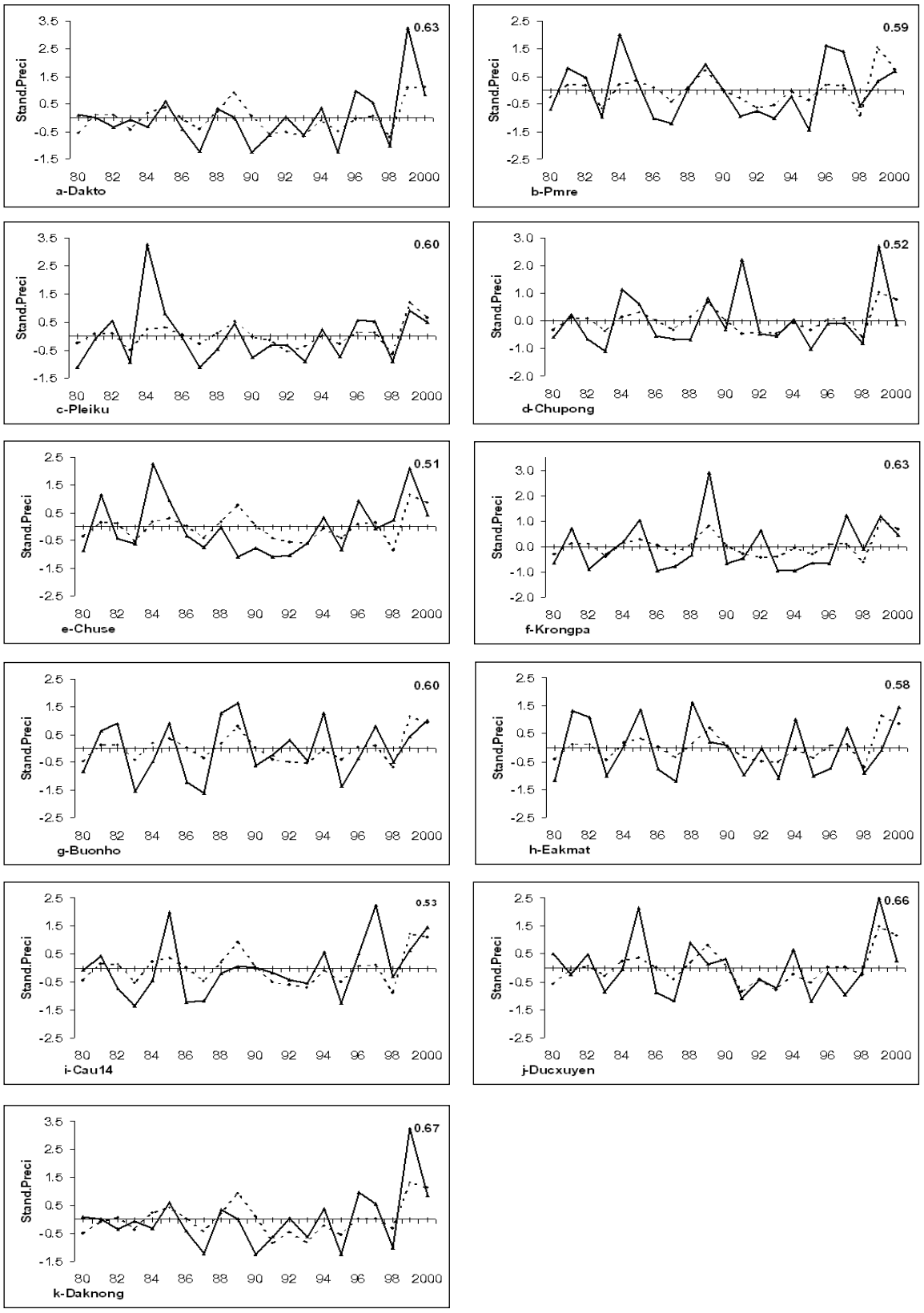

Fig. 3 The time series of standardized observed (solid line) and estimated (dashed line) precipitation in November obtained from a cross-validation model using SST in the Indian Ocean as predictor. The correlation coefficients between two time series presented on the top right of each graph are statistically significant at levels >95%.

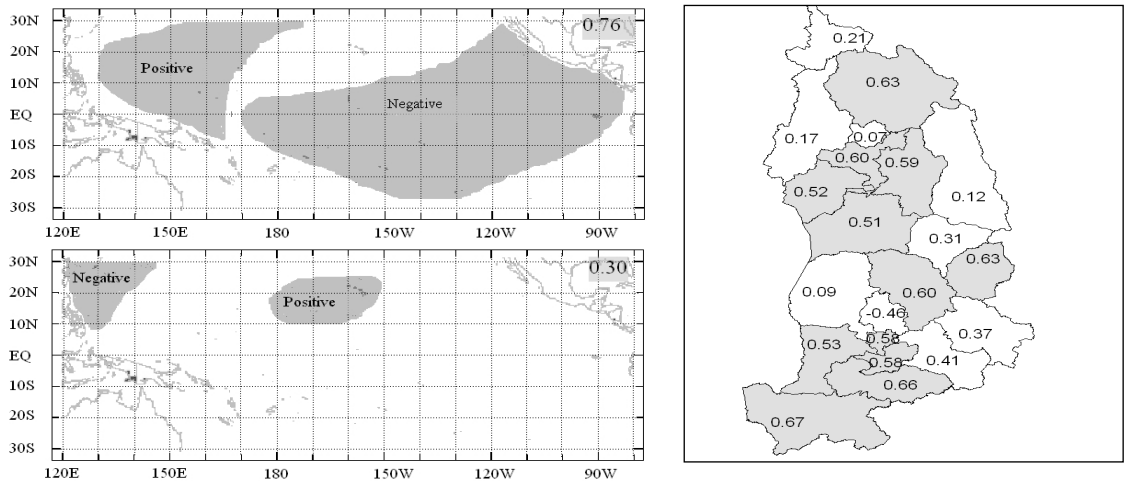

Fig. 4 The first (upper left panel) and second (lower left panel) mode of canonical predictor maps associated with April rainfall forecast skill based on February SST in the Pacific Ocean. The shading shows the SST areas that contribute most to the forecast. The correlation coefficient between canonical components of predictor and predictand is shown on the top right of each left figure. The right panel indicates the correlation coefficients between standardized observed and estimated precipitation obtained from all modes of cross-validation models. The correlation coefficients above 0.45 and 0.55 are significant at levels >95% and >99%, respectively.

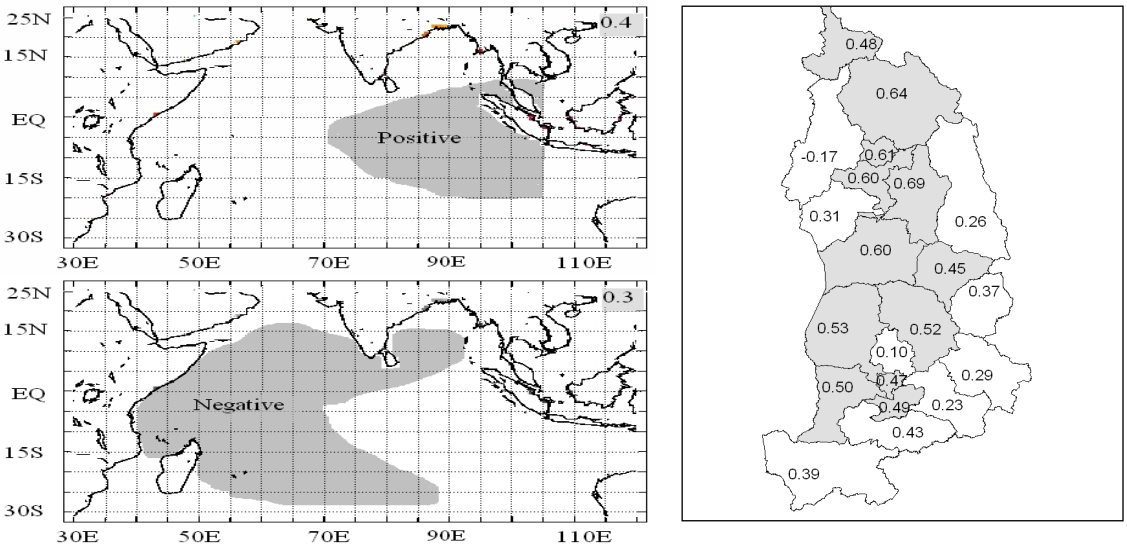

Fig. 5 The first (upper left panel) and second (lower left panel) mode of canonical predictor maps associated with November rainfall forecast skill based on September SST in the Indian Ocean. The shading shows the SST areas that contribute most to the forecast. The correlation coefficient between canonical components of predictor and predictand is shown on the top right of each left figure. The right panel indicates the correlation coefficients between standardized observed and estimated precipitation obtained from all modes of cross-validation models. The correlation coefficients above 0.45 and 0.55 are significant at levels > 95% and > 99%, respectively.

It would be interesting to identify those temporal and spatial evolutions in the SST field in both oceans that significantly contribute to the rainfall forecast skills. Figure 4

presents the first two g-map modes (left panels) of the linear combination of February SST over the Pacific Ocean that contributes most hindcast forecast skill to April rainfall. The right panel shows the correlation coefficients between standardized observed and estimated precipitation by cross-validation of the CCA model based on SST over the Pacific Ocean. Correspondingly, the g-map modes of the linear combination of September SST over the Indian Ocean that contributes most hindcast skill to November rainfall forecasting are presented in Fig. 5.

As shown in the right panel of Fig. 4, the best April rainfall forecasts obtained from the model using Pacific Ocean SST in February as predictor are at sites located in the central-west and south VCH, only one site is located in the north of the VCH (right panel). The upper left panel in Fig. 4 shows that the first g-map mode explains maximally about 50% of variance in April rainfall, whereas the second mode accounts for only about 10% of rainfall variance (lower left panel). The correlation coefficients in the first g-map mode imply that cold (warm) February SST over the equatorial central (northwestern) Pacific (upper left panel) is highly correlated with rainfall over the VCH, contributing most to the April rainfall forecast. This is also true for January and March SST (not shown). During April, the heavy rainfall over the VCH is mainly caused by convective activities. The positive SST anomalies over the northwestern and negative SST anomalies in the equatorial central to the eastern Pacific Ocean, representing ENSO, could enhance the convective rainfall over the VCH. In other words, La Niña increases and El Niño decreases precipitation over the VCH. This pattern was also found by, e.g. Lau & Yang (1997), Zhang *et al.* (2002, 2004) and Nguyen *et al.* (2007).

The best forecasts for November precipitation were obtained from the CCA cross-validation using Indian Ocean September SST. Warm (cold) SST over the equatorial eastern (western) Indian Ocean is highly correlated with the canonical components that were used for forecasting rainfall, but it does not contribute much to the November rainfall forecast. The left panels in Fig. 5 show the first two g-map modes that explain 20% of total variance of the November rainfall. The highest values of correlation between observed and estimated precipitation are at sites located in the north and the southwestern part of the VCH (right panel). In November, the meteorological conditions are such that northeasterly winds replace the southwesterly ones (summer monsoon). During this month, convective activity is the main cause of rainfall over the VCH expressed by positive September SST anomalies in the equatorial eastern Indian Ocean that feed the moisture to the convective rain band. This pattern is in agreement with the findings of Qian & Lee (2000) and Nguyen *et al.* (2007).

The variations of the rainfall over the VCH are not only associated with large-scale atmospheric and oceanic circulations. In particular, the two major mountain ranges in the northeastern and southeastern parts of the VCH (the average elevation is 1000 to 1500 m) play an important role for the rainfall on the lee-side of the mountain ranges. The northwestern Pacific and equatorial eastern Indian oceans feed the water vapour for the convection that causes precipitation over the VCH through large-scale atmospheric–oceanic circulations. The convection is triggered at the lee-side of the mountains forming the squall lines that generate the precipitation on the lee-side foot of the mountains (Satomura, 2000). This partially explains the spatial rainfall variations over the VCH, and also the spatial variability of model skills.

Discharge

The results of discharge forecasts obtained from ANN models are presented in detail in Table 2 and Fig. 6. In general, the models are potentially able to forecast discharge over the VCH, with correlation coefficients ranging from 0.78 to 0.88. The models cannot catch the discharge peaks correctly. This might be because the inputs to the models do not contain enough information about the peaks and the fact that most heavy rainfall events over the VCH are caused by storms and typhoons that in a complex way are influenced by large-scale circulation.

Table 2 Forecast results obtained from ANN models for discharge in the three studied sites over the VCH.

Catchment	Predictand	Input to network	Network structure	Cor. coeff	RMSE
Srepok	Discharge	SST_2, H_1,$T_{1\text{-max,min,mean,}}$ n, Q_1	7-10-1	0.82	162.6
Sesan	Discharge	SST_2, H_1, $T_{1\text{-max,min,mean}}$, Q_1	6-7-1	0.78	51.3
Ba	Discharge	SST_2, H_1, $T_{1\text{-max,min,mean}}$, Q_1	6-7-1	0.88	210

The inputs to the ANN models are: sea surface temperature (SST), humidity (H), air temperature (T), sunshine hours per month (n) and discharge (Q). The subscript shows the lag time prior to the forecast. The network structure a-b-c indicates the number of neurons in the input-hidden-output layer, respectively, and RMSE stands for the root mean squared error.

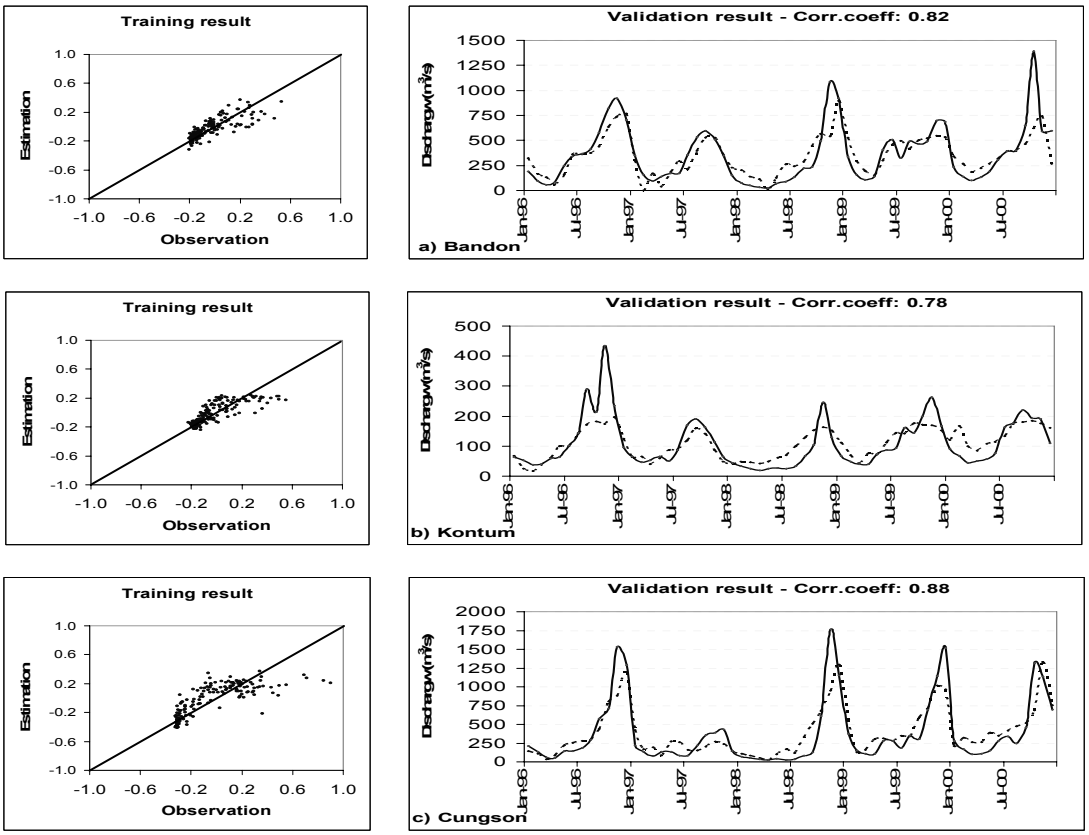

Fig. 6 Forecast results for monthly discharge obtained from ANN models. Left panels show the training results. Right panels present time series of observed (solid) and estimated (dash) discharge during the validation period for sites in three sites, Bandon (Srepok catchment), Kontum (Sesan catchment) and Cungson (Ba catchment).

Figure 6 shows the results of discharge forecast for three sites over the VCH. The left panels present results for the training period, and the right panels display the validation results. The scattering of the results for the training period indicates that the models tend to present well the discharge variability (left panels). This implies that drought conditions can be well-predicted in advance. The right panels show the time series of observed (solid line) and estimated (dash line) discharge values. Two-month lag SST in the El Niño3 area and one-month lag humidity, air temperature and discharge are used as input to the models. The time series of observed and estimated values are presented in the right panels, and high correlation coefficients are obtained for all three catchments over the VCH, 0.82 for Bandon (Srepok catchment), 0.78 for Kontum (Sesan catchment), and 0.88 for Cungson (Ba catchment).

CONCLUSIONS

Motivated by the strong relationships between precipitation over the VCH and SST over the Pacific and Indian oceans up to three months in advance, as demonstrated by Nguyen *et al*. (2007), this work has developed predictive statistical and nonlinear models using CCA and artificial neural network, respectively, to forecast the monthly rainfall and discharge for selected sites over the VCH.

The rainfall results obtained from the models for the different sites are of varying quality, because the precipitation over the VCH is influenced by local climatic patterns as well as the large-scale circulation governing the prevailing winds during the year. The local climatic patterns are mainly formed by the south Asian monsoon in combination with the topographical conditions in the region. Moreover, during the main rainy season, the heavy rainfall over the VCH is primarily caused by cyclones, typhoons and other disturbances associated with complex large-scale atmospheric circulations, so in this period few significant results could be obtained from the models based on only SST. The best discharge results obtained by ANN models are for sites where the rainfall is strongly correlated with the large-scale circulation and catchment characteristics.

The forecasting results for both rainfall and discharge show a very significant role in early drought warning over the VCH. Good performance of models for both forecasts imply that drought indices generated from rainfall or discharge, or a combination, can be presented in advance and be combined with real-time monitoring of other hydro-meteorological elements within the region, e.g. water levels in reservoirs, lakes and rivers, to provide valuable information on future drought conditions to water managers and decision-makers helping them to prepare mitigation and response actions.

Acknowledgments We thank the National Hydrometeorology Institute of Vietnam for providing the hydrological and meteorological data sets.

REFERENCES

Barnett, T. P. (1981) Statistical prediction of North American air temperature from Pacific predictors. *Mon. Weath. Rev.* **9**(5), 1021–1041.

Barnett, T. P. & Preisendorfer, R. W. (1987) Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Weath. Rev.* **115**, 1825–1850.

Barnston, A. G. & Ropelewski, C. F. (1992) Prediction of ENSO episodes using canonical correlation analysis. *J. Climate* **5**, 1316–1345.

Bishop, C. M. (1995) *Neural Networks for Pattern Recognition*. Oxford University Press, New York, USA.

Bretherton, C. S., Smith, C. & Wallace, J. M. (1992) An intercomparison of methods for finding coupled patterns on climate data. *J. Climate* **5**, 541–560.

Brikundavyi, S., Labib, R., Trung, H. T. & Rousselle, J. (2002) Performance of neural networks in daily streamflow forecasting. *J. Hydrol. Engng ASCE* **7**(5), 392–398.

Chen, T. C. & Yoon, J. H. (2000) Interannual variation in Indochina summer monsoon rainfall. Possible mechanism. *J. Climate* **13**, 1979–1986.

Cigizoglu, H. K. (2004) Discussion of performance of neural networks in daily streamflow forecasting. *J. Hydrol. Engng ASCE* **9**(6), 556–557.

Cordery, I. & McCall, M. (2000) A model for forecasting drought from teleconnections. *Water Resour. Res.* **36**(3), 763–768.

Dawson, C. W. & Wilby, R. L. (2001) Hydrological modelling using artificial neural networks. *Prog. Phys. Geog.* **25**(1), 80–108.

Deo, M. C. & Thirumalaiah, K. (2000) Real time forecasting using neural networks. In: *Artificial Neural Networks in Hydrology* (ed. by R. S. Govindaraju & A. R. Rao), 53–71. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Díaz, A. F., Studzinski, C. D. & Mechoso, C. R. (1998) Relationships between precipitation anomalies in Uruguay and southern Brazil and sea surface temperature in the Pacific and Atlantic Oceans. *J. Climate* **11**, 251–271.

Freiwan, M. & Cigizoglu, H. K. (2005) Prediction of total monthly rainfall in Jordan using feed forward backpropagation method. *Fresenius Environ. Bull.* **14**(2), 142–151.

Gershunov, A. & Cayan, D. (2003) Heavy daily precipitation frequency over the contiguous United States: Sources of climate variability and seasonal predictability. *J. Climate* **16**, 2752–2765.

Glahn, H. R. (1968) Canonical correlation analysis and its relationship to discriminant analysis and multiple regression. *J. Atmos. Sci.* **25**, 23–31.

Hastenrath, S. (1988) *Climate and Circulation in the Tropics*, 2nd edn. D. Reidel, Boston, Massachusetts, USA.

Hotelling, H. (1935) The most predictable criterion. *J. Educ. Psychol.* **26**(2), 139–143.

Hotelling, H. (1936) Relations between two sets of variates. *Biometrika* **28**, 321–377.

Kumar, K. K., Soman M. K. & Kumar, K. R. (1995) Seasonal forecasting of Indian summer monsoon rainfall: A review. *Weather* **50**, 449–467.

Lau, K. M. & Yang, S. (1997) Climatology and interannual variability of the Southeast Asian summer monsoon. *Adv. Atmos. Sci.* **14**, 141–162.

Lui, Z., Valdes, J. B. & Entekhabi, D. (1998) Merging and error analysis of regional hydrometeorological anomaly forecasts conditioned on climate precursors. *Water Resour. Res.* **34**(8), 1959–1969.

Moller, A. F. (1993) A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks* **6**, 525–533

Nguyen, T. D., Uvo, C. & Rosbjerg, D. (2007) Relationship between tropical Pacific and Indian oceans sea surface temperature (SST) and monthly precipitation over the central highlands, Vietnam. *Int. J. Climatol.* (conditionally accepted).

Nicholls, N. (1987) The use of canonical correlation to study teleconnections. *Mon. Weather Rev.* **115**, 393–399.

Ntale, H. K., Gan, T. Y. & Mwale, D. (2003) Prediction of East African seasonal rainfall using simplex canonical correlation analysis. *J. Climate* **16**, 2105–2112.

Olsson, J., Uvo, C. B., Jinno, K., Kawamura, A., Nishiyama, K., Koreeda, N., Nakashima, T. & Morita, O. (2004) Neural networks for rainfall forecasting by atmospheric downscaling. *J. Hydrol. Engng ASCE* **9**(1), 1–12.

Phien, H. N. & Danh, N. T. (1997) A hybrid model for daily flow forecasting. *Water SA* **23**(3), 201–208.

Phien, H. N. & Siang, J. J. (1993) Forecasting monthly flows of the Mekong River using back propagation. In: *Proc. IASTED Int. Conf.*, 17–20.

Qian, W. H. & Lee, D. K. (2000) Seasonal march of Asian summer monsoon. *Int. J. Climatol.* **20**(11), 1371–1386.

Rajeevan, M. (2001) Prediction of Indian summer monsoon: status, problems and prospects. *Curr. Sci. India* **81**, 1451–1457.

Rasmusson, E. M. & Carpenter, T. H. (1983) The relationship between the eastern Pacific sea surface temperature and rainfall over India and Sri Lanka. *Mon. Weath. Rev.* **111**, 354–384.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C. & Wang, W. (2002) An improved in situ and satellite SST analysis for climate. *J. Climate* **15**, 1609–1625.

Sahai, A. K., Grimm, A. M. & Satyan, V. (2000) All India summer monsoon rainfall prediction using an artificial neural network. *Clim. Dynam.* **16**, 291–302.

Salas, J. D., Markus, M. & Tokar, A. S. (2000) Streamflow forecasting based on artificial neural networks. In: *Artificial Neural Networks in Hydrology* (ed. by R. S. Govindaraju & A. Ramachandra Rao), 23–51. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Satomura, T. (2000) Diurnal variation of precipitation over the Indo-China Peninsula: Two-dimensional numerical simulation. *J. Meteor. Soc. Japan* **78**(4), 461–475.

Shabbar, A. & Barnston, A. G. (1996) Skill of seasonal climate forecasts in Canada using canonical correlation analysis. *Mon. Weath. Rev.* **124**, 2370–2385.

Shulka, J. & Mooley, D. A. (1987) Empirical prediction of the summer monsoon rainfall over India. *Mon. Weath. Rev.* **115**, 695–703.

Singhrattna, N. (2003) Interannual and interdecadal variability of Thailand summer monsoon: diagnostic and forecast. Master of Science Thesis, University of Colorado, Boulder, Colorado, USA.

Singhrattna, N. B., Rajagopalan, M. C. & Kumar, K. K. (2005) Seasonal forecasting of Thailand summer monsoon rainfall. *Int. J. Climatol.* **25**, 649–664.

Uvo, C. R. B. & Graham, N. E. (1998) Seasonal runoff forecast for northern South American: A statistical model. *Water Resour. Res.* **34**(12), 3515–3524.

Uvo, C. R. B., Tolle, U & Berndtsson, R. (2000) Forecasting discharge in Amazonia using artificial neural network. *Int. J. Climatol.* **20**, 1495–1507.

Vemuri, V. R. & Rogers, R. D. (1994) *Artificial Neural Networks: Forecasting Time Series*. IEEE Compu. Soc. Press., Los Alamitos, California, USA.

Zhang, Y., Li, T., Wang, B. & Wu, G. (2002) Onset of Asian summer monsoon over Indo-china and its interannual variability. *J. Climate* **15**, 3206–3221.

Zhang, Z. Z., Chan, C. L. & Ding, Y. (2004) Characteristics, evolution and mechanisms of the summer monsoon onset over Southeast Asia. *Int. J. Climatol.* **24**, 1461–1482.