Hydrological Sciences Journal

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/thsj20

Hydroclimate variability and long-lead forecasting of rainfall over Thailand by large-scale atmospheric variables

Nkrintra Singhrattna \(^a\), Mukand S. Babel \(^a\) & Sylvain R. Perret \(^a\)\(^b\)

\(^a\) Water Engineering and Management, Asian Institute of Technology, Pathumthani, Thailand

\(^b\) Centre de Coopération Internationale en Recherche Agronomique pour le Développement, UMR G-Eau, Montpellier, France

Available online: 20 Jan 2012

To cite this article: Nkrintra Singhrattna, Mukand S. Babel & Sylvain R. Perret (2012): Hydroclimate variability and long-lead forecasting of rainfall over Thailand by large-scale atmospheric variables, Hydrological Sciences Journal, 57:1, 26-41

To link to this article: http://dx.doi.org/10.1080/02626667.2011.633916

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Hydroclimate variability and long-lead forecasting of rainfall over Thailand by large-scale atmospheric variables

Nkrintra Singhrattna1, Mukand S. Babel1 and Sylvain R. Perret1,2

1Water Engineering and Management, Asian Institute of Technology, Pathumthani, Thailand
msbabel@ait.ac.th, nkrintras@yahoo.com
2Centre de Coopération Internationale en Recherche Agronomique pour le Développement, UMR G-Eau, Montpellier, France

Received 27 May 2010; accepted 25 March 2011; open for discussion until 1 July 2012

Editor Z.W. Kundzewicz


Abstract The development of statistical relationships between local hydroclimates and large-scale atmospheric variables enhances the understanding of hydroclimate variability. The rainfall in the study basin (the Upper Chao Phraya River Basin, Thailand) is influenced by the Indian Ocean and tropical Pacific Ocean atmospheric circulation. Using correlation analysis and cross-validated multiple regression, the large-scale atmospheric variables, such as temperature, pressure and wind, over given regions are identified. The forecasting models using atmospheric predictors show the capability of long-lead forecasting. The modified k-nearest neighbour (k-nn) model, which is developed using the identified predictors to forecast rainfall, and evaluated by likelihood function, shows a long-term forecast of monsoon rainfall at 7–9 months. The decreasing performance in forecasting dry-season rainfall is found for both short and long lead times. The developed model also presents better performance in forecasting pre-monsoon season rainfall in dry years compared to wet years, and vice versa for monsoon season rainfall.

Key words rainfall; hydroclimate variability; ENSO; large-scale atmospheric variables; long-lead forecasting; statistical approach; modified k-nn model; cross-validated multiple regression; Chao Phraya River Basin; Ping River Basin; Thailand

Variabilité hydroclimatique et prévision à long terme des précipitations en Thaïlande à l’aide de variables atmosphériques de grande échelle

Résumé L’établissement de relations statistiques entre variables pluviométriques locales et variables atmosphériques de grande échelle permet une meilleure compréhension de la variabilité hydroclimatique. Les précipitations dans le bassin d’étude (le bassin supérieur de la Rivière Chao Phraya, Thaïlande) sont influencées par la circulation atmosphérique dans l’Océan Indien et dans l’Océan Pacifique tropical. Par l’analyse de corrélations et régressions multiples en validation croisée, les prédicteurs de variables atmosphériques de grande échelle, à savoir température, pression et vent, sont identifiés sur des régions cibles, et montrent une bonne capacité de prévision à long terme. Le modèle modifié des k-plus proches voisins (k-nn), développé par les prédicteurs identifiés pour prévoir les pluies, et évalué par fonction de probabilité, montre une prévisibilité à long terme (7–9 mois) des pluies de mousson. La moindre performance dans la prévision des pluies de saison sèche est prouvée pour la prévision à court et à long terme. Le modèle développé présente également une meilleure performance pour les pluies de pré-mousson en année sèche, par rapport à une année humide, et inversement pour les pluies de mousson.

Mots clefs précipitations; variabilité hydroclimatique; ENSO; variables atmosphériques de grande échelle; prévision à long terme; approche statistique; modèle k-nn modifié; régression multiple à validation croisée; bassin du Fleuve Chao Phraya; bassin du Fleuve Ping; Thaïlande

1 INTRODUCTION

The links between large-scale atmospheric variables and local hydroclimates are widely studied and reported to diagnose the variability of local hydroclimates (Smith and O’Brien 2001). The anomalies of the climate variables known as atmospheric indices, e.g. the anomalous sea-surface temperature (SST) index or El Niño-Southern Oscillation...
(ENSO), can strengthen or weaken a monsoon (Mason and Goddard 2001, Smith and O’Brien 2001). Thus, the atmospheric anomalies are responsible for the variability of local hydroclimates, with respect to temperature (Gershunov 1998, Pavia et al. 2006), precipitation (Masiokas et al. 2006) and streamflow (Gutiérrez and Dracup 2001, Meko and Woodhouse 2005). Due to an anomalous phase of the ocean and atmosphere, the responses of local hydroclimates may not be experienced with the same level of variability. In some regions, storms develop more strongly and frequently (Cañón et al. 2007, Kim et al. 2006), whereas in other regions, dry conditions are experienced with droughts (Mendoza et al. 2005) during longer periods (Schöngart and Junk 2007). The effects of anomalous oceanic–atmospheric circulation have changed climate patterns not only along the coast, but also in regions located far from it (Saravanan and Chang 2000). Moreover, their effects are difficult to identify because the local climate may be influenced by nearby and distant oceans through the coupled atmospheric–oceanic circulation (Wu and Kirtman 2004, Nagura and Konda 2007). The teleconnection influences could be determined by the lag time of anomalous events up to several months after a phenomenon occurs (Sourouille and Lall 2003, Grantz et al. 2005), which makes it difficult to understand.

The annual variability of the Southeast Asian climate is influenced by the interactions between the annual reversal of surface monsoonal winds from the Indian Ocean to the equatorial western Pacific Ocean, including the South China Sea, and the complex distribution of land, sea and terrain. For the seasonal development of surface winds, the Asian summer monsoon that develops due to the land–ocean temperature gradient is influenced by the sea-surface temperature (SST) anomalies over the equatorial eastern Indian Ocean (Goswami et al. 2006) and the tropical Pacific Ocean (Fasullo and Webster 2002). The non-convective anomalies over the equatorial Indian Ocean tend to weaken the convection over the Bay of Bengal, the eastern Indian Ocean, Southeast Asia and the equatorial western Pacific Ocean (Krishnan et al. 2000). Thus, the anomalous oceanic–atmospheric circulation, e.g. ENSO, affects the Asian summer monsoon and subsequently the monsoon rainfall. A warm phase of ENSO (El Niño) is associated with a weak monsoon and below-normal rainfall, whereas it is the opposite for a cold phase of ENSO, i.e. La Niña (Shrestha 2000, Shrestha and Kostaschuk 2005). However, the influences over a particular region have to be carefully investigated due to the different responses which make it hard to define a certain pattern of local hydroclimates (An et al. 2007).

The lagging relationships of local hydroclimates with anomalous atmospheric events are also of interest when developing hydroclimatic forecasting models. Several models, such as artificial neural networks (ANNs) (Silverman and Dracup 2000), multiple regression model (Schöngart and Junk 2007) and principal component analysis (Eldaw et al. 2003), have been developed, and show better performance using the predictors of large-scale atmospheric information (Singhrattna et al. 2005a, Zehe et al. 2006). The basic variables, e.g. SST, surface air temperature, atmospheric pressure, wind and ENSO standard indices, are applied in the forecasting models (Gershunov 1998, Clark et al. 2001, Hamlet et al. 2002, McCabe and Dettinger 2002, Grantz et al. 2005) based on the significant relationships between large-scale atmospheric variables and hydroclimates. The forecasting models enable successful short-lead projections to be made (Gutiérrez and Dracup 2001, Schöngart and Junk 2007); however, the model performance of long-lead forecasts still needs to be improved (Krishna Kumar et al. 1995, DelSole and Shukla 2002). The decadal variability of large-scale atmospheric variables used as the predictors of a model could weaken the relationships with hydroclimates and subsequently decrease the model performance of long-lead forecasts.

The objectives of this research are to study the climate characteristics and their variability in part of the Upper Chao Phraya River Basin located in Thailand (Southeast Asia), and to identify the predictors of large-scale atmospheric variables at long-lead times for rainfall forecasting. Using the identified variables as predictors, the rainfall forecasting model, based on a statistical-stochastic approach, is ultimately developed to forecast the seasonal rainfall at long lead times.

2 THE STUDY BASIN

Thailand, located between the Indian Ocean and the Gulf of Thailand which is connected to the Pacific Ocean (Fig. 1(a)), covers an area of 513 115 km² with a population of 62.4 million (2005). Its climate is influenced by the Indian Ocean and the Pacific Ocean due to the land–ocean interrelated temperature gradients. The summer season, from mid-February to mid-May, is responsible for developing the land–ocean temperature gradient that strengthens the southwest
monsoon from the Indian Ocean in the following rainy season. The rainy season is caused by the Inter-Tropical Convergence Zone (ITCZ) and by the southwest monsoon, and lasts from mid-May to mid-October. Due to the low intensity and uncertainty of rainfall occurrence, the period from May to July is called the pre-monsoon season, or transition period from summer to monsoon season. In contrast, heavy rainfall occurs from August to October. The average annual rainfall in the country ranges from 1200 to 1600 mm. The winter season is characterized by dry and cool winds brought by the northeast monsoon from the mid-latitudes, and lasting from mid-October to mid-February. In terms of streamflow, Thailand generates about $289 \times 10^9$ m$^3$ per year of total average runoff in 25 major river basins. However, only $38 \times 10^9$ m$^3$ per year or 13.1% of the annual runoff can be stored in reservoirs and supplied to various sectors for use within the country (RID 2007).

The Chao Phraya River Basin (Fig. 1(b)) is the largest among the 25 major basins, covering an area of 178 000 km$^2$ or 35% of the country’s land area. Four major tributaries, the Ping, Wang, Yom and Nan rivers, merge at Nakhon Sawan and form the Chao Phraya River. The portion above the confluence at Nakhon Sawan is called the Upper Chao Phraya River Basin, whereas that below the confluence is the Lower Chao Phraya River Basin. The upper basin covers an area of 102 635 km$^2$ or 58% of the Chao Phraya River Basin.

The study basin, the Ping River Basin, is located in northern Thailand, between 15°–19°N latitude and 98°–100°E longitude, and covers an area of 33 899 km$^2$ (Fig. 1(c)). The Ping River is 740 km long and many storage dams are located along the river and its tributaries. The most important is the Bhumipol Dam, constructed for the purposes of hydropower, agriculture, flood mitigation, fishery and transportation. It is 154 m high and 486 m long with a maximum storage capacity of $13.462 \times 10^9$ m$^3$ and installed capacity of 779.2 MW of hydropower. In terms of land use, 71.46% of basin area is covered with forests, which are mostly located in the upstream or the origin of the Ping River. The remainder, along both sides of the river and the plain downstream, is covered by agricultural and residential areas and water bodies.

3 DATA

3.1 Hydroclimatic data

Two hydroclimate data sets—rainfall and temperature—from the climatic stations located in and around the Ping River Basin, are used in this study. Both data are recorded on a daily basis. Out of 208 rainfall stations operated by the Royal Irrigation
Hydroclimate variability and long-lead forecasting of rainfall over Thailand by large-scale atmospheric variables

Department (RID), Thailand Meteorological Department (TMD) and Department of Water Resources (DWR) of the Royal Thai Government, 50 stations are selected based on the length of time series and the occurrence of incomplete data. Daily rainfall data observed at the selected stations are obtained for the period 1950–2007, with incomplete data constituting less than 5% of the recent 30 years’ data. Fisher’s Transformation ($z'$) is used to test the hypothesis on significance of a sample correlation coefficient by (Haan 2002):

$$z' = 0.5 \ln \left( \frac{1 + r}{1 - r} \right)$$  \hspace{1cm} (1)

where $r$ is the correlation coefficient between two independent samples, and $z'$ is approximately normally distributed, and is used to compare with $z$ of the standard normal distribution to obtain the significance level ($p$) of correlation.

Coefficients of correlation obtained for monthly rainfall of all selected stations varied from 0.46 to 0.96 and are well correlated at the 95% significance level by Fisher’s Transformation. Therefore, the average values are calculated over 50 selected stations.

For daily air temperature, 11 stations of the TMD are selected with the length of data varying from 16 to 56 years. The average is estimated over 11 selected stations which are statistically correlated at 95% confidence level. The locations of selected rainfall and temperature stations are shown in Fig. 1.

3.2 Standard sea-surface temperature (SST)

This study used monthly data of SST for 1950–2007 over four regions: NINO 3, NINO 4, ION and SCS, as follows. The standard SSTs over the tropical Pacific Ocean used in this study are:

- NINO 3: average SST over the area of 5°N–5°S latitude and 90°–150°W longitude; and
- NINO 4: average SST over the area of 5°N–5°S latitude and 150°W–160°E longitude.

These SSTs are provided by the Climate Prediction Center (CPC 2007) of the US National Oceanic and Atmospheric Administration (NOAA).

The standard SST over the Indian Ocean (ION) is the SST obtained from two data sets: the Comprehensive Ocean–Atmosphere Data Set (COADS) (Woodruff et al. 1993) and real-time surface marine data from the National Centers for Environmental Prediction (NCEP), and averaged over the region of 2°N–2°S latitude and 70°–90°E longitude. The ION data set is provided by the Physical Sciences Division of NOAA (PSD 2007b).

The SST over the South China Sea (SCS) is the SST averaged over the region of 5°N–5°S latitude and 100°–115°E longitude. The SCS data set is provided by the PSD (PSD 2007a).

3.3 Large-scale atmospheric data

The large-scale atmospheric variables used in this study are obtained from re-analysis-derived data, which are provided by the NCEP/NOAA (Kalnay et al. 1996). Several variables, on both a daily and a monthly basis, are available from 1948 to the present, and cover a global grid of 2.5° latitude $\times$ 2.5° longitude (PSD 2007a). This study uses monthly data of four principal atmospheric variables from 1948–2007: the surface air temperature (SAT); sea-level pressure (SLP); surface zonal wind (SXW); and surface meridian wind (SYW). These variables play an important role in strengthening or weakening a monsoon, and influence the convection over the study basin. The significant ability of forecasting models can be obtained using combinations of these variables (Sahai et al. 2003).

4 CLIMATE DIAGNOSTIC

4.1 Inter-annual climate

The annual cycles of air temperature and rainfall in the Ping River Basin are shown in Fig. 2. The summer
season occurs during March-April-May (MAM). The maximum temperature of 30.2°C is observed in April, and the minimum of 22.9°C in December. Figure 2 shows that rainfall is bi-modal with two peaks: one in May and another in August. Considering the primary peak, the rainy or monsoon season in the basin occurs during August-September-October (ASO). Moreover, the secondary peak is observed during May-June-July (MJJ), which is the pre-monsoon season. The secondary peak of the pre-monsoon is related to the southwest monsoon and the ITCZ passing from the Indian Ocean to Thailand in May and to the South China Sea and central China in mid-June. However, the primary peak is associated with the ITCZ as it moves back over Thailand during ASO. From 58 years of rainfall data, the maximum MJJ rainfall of 651.8 mm and the maximum ASO rainfall of 948.3 mm are found in 1950. The minimum MJJ rainfall of 254.0 mm is in 1997 and the minimum ASO rainfall of 387.7 mm in 2004. The total annual rainfall varies from 843.0 to 1605.6 mm. The MJJ and ASO rainfall is about 88% of the total annual rainfall. The remaining rainfall is the dry-season rainfall which falls in November–April and ranges from 49.4 to 295.2 mm over 58 years.

4.2 Inter-seasonal climate

The air temperature is related to rainfall in terms of developing the temperature gradient between land and sea that subsequently strengthens the monsoon. In this study, the summer season temperature is averaged for MAM. Figure 3 shows the inverse relationship between the dry-season rainfall and MAM air temperature. As expected, if more rainfall is received during November–April, it would tend to cool the land and atmosphere, thus decreasing the air temperature during MAM and vice versa.

![Fig. 3 Scatter plots between dry-season (November–April) rainfall and March-April-May (MAM) air temperature (T: temperature; R: rainfall).](image)

![Fig. 4 Scatter plots between March-April-May (MAM) air temperature and (a) May-June-July (MJJ) rainfall, and (b) August-September-October (ASO) rainfall (T: temperature; R: rainfall).](image)

The MAM temperature subsequently plays a role in driving pre-monsoon and monsoon season rainfall. Figure 4 shows the relationships between the MAM air temperature and both MJJ rainfall and ASO rainfall. During the pre-monsoon season or the transition period, the land–ocean temperature gradient is not fully developed. Hence, inverse relationships between MAM temperature and MJJ rainfall are found. During the monsoon season, as the positive relationships indicate, the higher MAM air temperature increases the ASO rainfall, whereas the lower MAM air temperature decreases the ASO rainfall.

To further enhance the understanding of the land–sea temperature gradient that influences the seasonal rainfall, the relationships between the MAM surface temperature of nearby oceans and MJJ and ASO rainfall are investigated (Fig. 5). The positive relationships between MAM air temperature over the study basin and ASO rainfall are stronger than those of MJJ rainfall, which is consistent with the relationships in Fig. 4.

For the ASO rainfall (Fig. 5(b)), the positive correlations at 95% significance level (i.e. upper and lower bounds of +0.26 and –0.26, respectively) are associated with the MAM surface temperature especially over the study basin; negative correlations at 99% significance level (between +0.34 and –0.34) are associated with MAM surface temperature over the South China Sea. This indicates the developing land–sea temperature gradient.

To confirm the influence of land–sea temperature gradient on the monsoon rainfall, the differences in MAM temperature over the study basin and over the
regions of nearby oceans such as the Pacific Ocean (NINO 3 and NINO 4), the Indian Ocean (ION) and the South China Sea (SCS) are estimated and correlated to the MJJ and ASO rainfall (Table 1). With the 90% confidence level, the significant

relationships between MJJ rainfall and MAM temperature differences are found to correspond to the NINO 3 region in the Pacific Ocean. The ASO rainfall develops stronger relationships than MJJ rainfall with the MAM temperature differences corresponding to the regions of NINO 4, ION and SCS. Thus, the land–sea temperature gradients developed from different sources have an influence on the seasonal rainfall. However, more investigation and detailed study are required to distinguish the influences on local hydroclimates.

### Table 1 Cross-correlations between seasonal rainfall during May-June-July (MJJ) and August-September-October (ASO) and temperature differences during March-April-May (MAM) over the study basin and four different regions over the oceans.

<table>
<thead>
<tr>
<th>Season</th>
<th>Correlation coefficient:</th>
<th>NINO 3</th>
<th>NINO 4</th>
<th>ION</th>
<th>SCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MJJ rainfall</td>
<td></td>
<td>0.23*</td>
<td>0.04</td>
<td>−0.07</td>
<td>−0.12</td>
</tr>
<tr>
<td>ASO rainfall</td>
<td></td>
<td>0.18</td>
<td>0.25*</td>
<td>0.25*</td>
<td>0.35**</td>
</tr>
</tbody>
</table>

NINO 3: a region over the central equatorial Pacific Ocean between 5°N–5°S latitude and 90–150°W longitude; NINO 4: a region over the western Pacific Ocean between 5°N–5°S latitude and 150°W–160°E longitude; ION: a region over the Indian Ocean between 2°N–2°S latitude and 70–90°E longitude; SCS: a region over the South China Sea between 5°N–5°S latitude and 100–115°E longitude. Figures with * and ** represent the correlations being significant at 90% and 99% confidence level, respectively.

### 4.3 Trend of seasonal hydroclimates

To support the influences of temperature on the seasonal rainfall and to investigate the variability of seasonal hydroclimates over the decades, trend analysis of seasonal rainfall and temperature is carried out. The time series of seasonal rainfall, i.e. dry (November–April), pre-monsoon (MJJ) and monsoon (ASO) season, are presented in Fig. 6, and the time series of MAM temperature are shown in Fig. 7. Based on the linear trend, the dry-season rainfall...
shows a slightly increasing trend of 16.0 mm over 57 years. Based on the inverse correlations between dry-season rainfall and MAM temperature (Fig. 3), the increasing trends of dry-season rainfall tend to decrease the MAM temperature, as shown in Fig. 7. There is a decrease of 0.6°C in the MAM temperature over 56 years of data. This trend is statistically significant at 95% confidence level by the standard \( t \)-test (Haan 2002), which is applied to test a null hypothesis of the slope of a regression line or linear trend being different from a specific value. The standard \( t \)-test equation is as follows:

\[
 t = \frac{\hat{\beta} - \beta_0}{SE} \quad (2)
\]

where \( \hat{\beta} \) is the slope of a fitting regression \( (y_i = \alpha + \hat{\beta}_i + \epsilon_i) \), \( \beta_0 \) is the specific value for testing, i.e. 0 in this case, and SE is the standard error \( (\epsilon_i) \) of least-squares of the estimates.

The decreasing trend of MAM temperature subsequently causes the decreasing trends in the MJJ and ASO rainfall, as shown in Fig. 6, due to the weakening of development of land–sea temperature gradient. For the period 1950–2007, the MJJ rainfall shows a decreasing trend of 34.2 mm, and the ASO rainfall shows a decreasing trend of 102.1 mm. It is also important to note that the years reported as the warm phase of ENSO or El Niño (CPC 2006, COAPS 2006) coincided with below-normal MJJ and ASO rainfall, and vice versa during the La Niña years. This is corroborated by Singhrattna et al. (2005b) and Chen and Yoon (2000). Moreover, the decreasing trend of rainfall is associated with that reported by the Intergovernmental Panel on Climate Change (Trenberth et al. 2007). Based on historical observations (1900–2005), they found that precipitation over the areas between 10°N and 30°N tended to decline since 1970 due to global warming and climate change. However, the precipitation variability is dependent upon region and season.

5 PREDICTOR IDENTIFICATION

5.1 Correlations with large-scale atmospheric variables

The climate diagnostic shows that the increasing land–sea temperature gradient strengthens the monsoons and influences the rainfall. Since an ultimate objective of this study is to develop a model to forecast rainfall at long lead times using the large-scale atmospheric variables, this section aims to develop the relationships between rainfall in the study basin and other large-scale atmospheric variables such as SAT, SLP, SXW and SYW at various lead times to identify the predictors of rainfall. The relationships are developed using correlation maps which are the interactive plots and analysis provided by the Earth System Research Laboratory (ESRL 2008). Although the physical long lead–lag relationship is not plausible, Sahai et al. (2003) found that the correlation patterns between Indian summer monsoon rainfall (ISMR) and SST show a slow and consistent temporal evolution which addresses the significant lead time relationships of SST even four years prior to the monsoon season. Clark et al. (2000) also used various Indian Ocean indices at long lead
times to develop the relationships for rainfall predictions. Nicholls (1983) found relationships between ISMR and 16-month leading SST near Indonesia that can show the long lead time relationships and the influences of large-scale atmospheric circulation at distance from the source (Saravanan and Chang 2000, Harshburger et al. 2002, Tereshchenko et al. 2002). Therefore, in this study, the correlation maps are adopted using the large-scale atmospheric variables, e.g. SAT, SLP and surface wind with varying lead times from 4 to 15 months prior to season starting. The correlation maps also cover the oceans, e.g. the Indian Ocean, the Pacific Ocean and the South China Sea that indicate influences on rainfall in the study basin. The predictors are selected based on long-leading relationships and over regions giving high correlations at 95% significance level. Then, the optimal combination cases of predictors will be identified by cross-validated multiple regression and used in the forecasting model.

5.2 Combination cases of the predictors

To avoid redundancy in the case of several identified predictors, the optimal sub-set of predictors that is composed of the minimum number of mutually exclusive variables is selected by model selection methods. In this study, a cross-validated multiple regression, namely the generalized cross-validation (GCV) with the leave-one-out technique is adopted to select an optimal combination of predictors. For \( k \) multiple independent variables, there are \( 2^k - 1 \) cross-combination cases of variables. Each combination case is evaluated using the leave-one-out technique, which is applied to all points of data \( n \) by dropping one point out of the dependent \( y_i \) and independent \( x_i \) data set. Then, the regression is developed to fit the remaining points of data \( (n - 1) \). At the dropped point, the estimation of dependent variable \( y_i' \) is done with the developed regression using the dropped independent variables \( x_i \). The residual \( (y_i - y_i') \) at this point is also calculated. The procedure is repeatedly applied to all points of data \( n \), and the GCV is calculated by:

\[
\text{GCV} = \frac{\sum_{i=1}^{n} (y_i - y_i')^2}{n (1 - m/n)^2}
\]  

(3)

where \( y_i \) is observed data at the dropped point \( x_i \) based on the leave-one-out technique, \( y_i' \) is the estimation at the dropped point obtained from the fitting regression, \( n \) is the total number of observed data, and \( m \) is the number of independent variables or predictors applied to the fitting multiple regression.

From the \( 2^k - 1 \) combination cases, the best combination of predictors is selected due to the minimum GCV which is associated with the smallest error from the fitting regression using the minimum number of mutually exclusive variables.

6 FORECASTING MODEL

6.1 Description of modified \( k \)-nearest neighbour (\( k \)-nn) model

The nonparametric regression is an alternative of the fitting data model. The regression equation is as follows:

\[
y = f(x_1, x_2, x_3, \ldots, x_k) + e
\]

(4)

where \( f \) is the fitting function of the independent variables \( x_1, x_2, x_3, \ldots, x_k \) (in this case, the identified large-scale atmospheric variables); \( y \) is the dependent variable (in this case, seasonal rainfall); and \( e \) is the error or residual assumed to be normally distributed with mean \( = 0 \) and variance \( = \sigma \).

Unlike the parametric approach, the prior assumption of relationships between dependent and independent variables such as linear relationships, which is one of the drawbacks of parametric models, is not required for the nonparametric approach. The fitting functions \( f \) capture relationships locally or by a small set of neighbours \( k \) at a given point \( x_i \). Because the functions are flexible and able to fit any arbitrary data, for example, bivariate data and multivariate data, the nonparametric approach can determine relationships better than the parametric regression. Moreover, a drawback of global fitting in the parametric approach, where all points of data may have a large influence on an individual point of a curve, can be solved by the local fitting functions.

There are several approaches of nonparametric regression. One that grants the discontinuities of the derivative curve is the spline approach. Another that applies a regression locally to a given point \( x_i \) and the neighbours around \( x_i \) is referred to as the local polynomials approach. This approach includes locally-weighted polynomials (Loader 1999) and \( k \)-nearest neighbour \( (k \)-nn) local polynomials (Owosina...
independent variables (identified large-scale atmospheric predictors as the
is adopted to forecast seasonal rainfall using the
functions such as GCV and likelihood.

In terms of forecast, the modified k-nn model
is adopted to forecast seasonal rainfall using the
identified large-scale atmospheric predictors as the
independent variables (\(x_i\)). The fitting of a regression,
rainfall forecast and resample data to achieve rainfall
ensembles are described in the following steps:

1. For the fitting process, the size of neighbours (\(k\))
and the order of polynomial (\(p\)), normally 1 or
2, will be selected as the combination of \(k\) and
\(p\) obtaining the minimum GCV score by:

\[
GCV(k,p) = \frac{\sum_{i=1}^{n} e_i^2}{n \cdot (1 - m/n)^2}
\]

where \(e_i\) is the error, \(n\) is the number of data points
\((x_i)\), and \(m\) is the number of parameters.

2. The regression is fitted locally with \(k\) and \(p\)
obtained from Step 1.

3. The rainfall according to the developed fitting
regressions is then estimated and called the mean
estimations \((\bar{y}_1, \bar{y}_2, \bar{y}_3, \ldots, \bar{y}_n)\). Then, the residu-
als \((e_1, e_2, e_3, \ldots, e_n)\) are computed.

4. At a new point of independent variable \((x_{new})\),
the forecast of rainfall is required. The mean estima-
tion \((\bar{y}_{new})\) is obtained from the developed fitting
regression.

5. An ensemble forecast is then obtained by adding
the residual \((e_i)\) to \(\bar{y}_{new}\). The residual \(e_i\) corre-
sponds to one of the \(k\)-nearest neighbours \((k-nn)\)
of \(x_{new}\) which is randomly selected by using a
weight function as follows:

\[
W(j) = \frac{1/j}{\sum_{i=1}^{k} (1/i)}
\]

where \(W(j)\) is the weight value of a neighbour of
\(x_{new}\) whose distance from \(x_{new}\) falls in the \(j\)th rank,
and \(k\) is the size of the neighbours, which does
not have to be the same as for fitting polynomial.
The term \(\sqrt{n-1}\) is used, in practice, to estimate \(k\)
where \(n\) is the total number of \(x_i\). The weight func-
tion gives more weight to the nearest neighbour
and less weight to the farthest neighbour. From

6. Repeat Step 5 as many times as required to
achieve the forecasting ensembles of rainfall, or a
number of simulations \((N)\); in this case \(N = 300\).

7. Repeat steps 4–6 for each forecasting point.

The \(N\) ensembles of forecasting rainfall can
develop a box-plot or a probability density function
(pdf). It can provide an exceedence and a non-
exceedence probability of the extreme events, e.g. wet
or dry. The exceedence and non-exceedence prob-
babilities can support water resources managers to
implement a strategic planning of water allocation
and management and agricultural practices.

6.2 Model evaluation

The likelihood function (lhf) is applied to evaluate
the modified k-nn model on capturing the pdf
which is an essential tool of decision making for
water resources management. The initial step of lhf
calculation is to divide the climatological data of
1950–2007 into three categories: below-normal (dry),
normal and above-normal (wet). In this case, three
categories are determined separately for each season
of rainfall by dividing the seasonal rainfall data at the
33rd percentile and the 67th percentile: the rainfall
data below the 33rd percentile fall into the dry cat-
egory, whereas those above the 67th percentile are
in the wet category; the remaining data are in the
normal category. Next, the categorical probabilities
of climatology that are the proportion of historical
data in each category are calculated. Subsequently,
for a given year, rainfall ensembles in each category
Hydroclimate variability and longlead forecasting of rainfall over Thailand by large-scale atmospheric variables

7 RESULTS AND DISCUSSION

7.1 Identifying predictors and combination cases

To develop a long-lead forecasting model, the statistical relationships between rainfall during MJJ, ASO, NDJ and FMA and the large-scale atmospheric variables are developed by correlation maps with varying lead times of the atmospheric variables. Figure 8 shows examples of developed relationships between premonsoon season (MJJ) rainfall and SAT for 12 different lead times varying from 4 to 15 months before the beginning of the premonsoon season in May. The solid box represents the selected region of predictor which has high correlation at the 95% significance level (upper and lower bounds of +0.26 and –0.26, respectively). Statistical relationships from correlation maps were also developed for the variables SLP, SXW and SYW and ASO, NDJ and FMA rainfall data (not shown). The significant relationships of monsoon rainfall (i.e. MJJ and ASO rainfall) show development with the large-scale atmospheric variables over the study basin and the nearby seas and oceans, e.g. South China Sea, the Indian Ocean and the Pacific Ocean. The developed relationships are found to be associated with long lead times of atmospheric variables of 8–15 months prior to the start of the season. This is corroborated by Sahai et al. (2003) and Nicholls (1983), who presented long-lead relationships between Indian summer monsoon rainfall and atmospheric variables, e.g. SST. The significant relationships between monsoon (i.e. MJJ and ASO) rainfall and large-scale atmospheric variables suggest the ability of the rainfall forecasting models to make long-lead forecasts using the identified atmospheric predictors, which is hardly found in any study of Thailand rainfall forecasting (Hung et al. 2009, Singhrattna et al. 2005a, Weesakul and Lowanichchaisri, 2005). However, the dry-season (i.e. NDJ and FMA) rainfall gave shorter lead-time relationships with large-scale atmospheric variables of 7–10 months. Furthermore, significant links are found over distant regions, e.g. Sumatra and Java (Indonesia), and northeast India, which shows the remote influence of atmospheric circulation (Harshburger et al. 2002, Tereshchenko et al. 2002). Table 2 summarizes the identified predictors of large-scale atmospheric variables. It is also noted that the significant relationship between SLP and NDJ rainfall is hardly found over the region, e.g. the study basin, the Pacific Ocean and the South China Sea. Moreover, the observed rainfall data (RAIN) are included in the identified predictors because, from the lag autocorrelation analysis (Fig. 9), the negative correlations at 95% significance level are associated with a 6-month lag (and the positive correlations – with a 12-month lag), which indicates the ability of forecasting models to give long-lead forecasts using RAIN as the predictor. Hence, there are five identified predictors each of the MJJ, ASO and FMA rainfall, and four predictors for the NDJ rainfall.

Subsequently, an optimal combination of predictors is selected by cross-validated multiple regression. Based on k independent predictors, the $2^k - 1$
cross-combination cases can be set. In this study, the optimal sub-sets are identified separately for each season of rainfall and each lead period of predictors that can provide forecasts varying from 1 to 12 months prior to the season starting. For example, for MJJ rainfall, the 15-month lead-time predictors, which are averaged during February-March-April (FMA) in the current year, are used to forecast the MJJ rainfall of the following year. The forecast can be issued on 1 May, i.e. 12 months in advance. From the GCV, an optimal sub-set of predictors is selected from the $2^k - 1$ cross-combination cases corresponding to the minimum GCV. Table 3 summarizes the best combination cases of predictors for rainfall during MJJ, ASO, NDJ and FMA. The selected combinations consist of 1–2 mutually exclusive predictors, except for

Fig. 8 Correlation maps between MJJ rainfall and surface air temperature (SAT) at 15- to 4-month lead times, respectively: (a) FMA, (b) MAM, (c) AMJ, (d) MJJ, (e) JJA, (f) JAS, (g) ASO, (h) SON, (i) OND, (j) NDJ, (k) DJF and (l) JFM. The 95% significance levels correspond to $\pm 0.26$. 

Downloaded by [Nkrintra Singhrattna] at 16:41 23 January 2012
two combination cases for ASO rainfall corresponding to the 6- and 10-month forecasting lead times that address three predictors of SAT, SXW and SYW, and SAT, SYW and RAIN, respectively. The identified sub-sets indicate that the univariate regression of large-scale atmospheric variables (e.g. SXW or SYW) and the multivariate regression (e.g. SYW and RAIN or SLP and SXW) are more efficient than the univariate regression of observed rainfall (i.e. RAIN). This suggests the better performance of a fitting regression and forecasting model with the incorporation of large-scale atmospheric predictors, e.g. SAT and SLP (Basson and Rooyen, 2001, Trenberth et al. 2006, Sadhuram and Ramana Murthy, 2008) compared to developing a regression from only lagging relationships of hydroclimatic variables, e.g. RAIN.

### 7.2 Forecasting evaluation

The lhf scores used to evaluate the modified k-nn model with the leave-one-out technique are computed separately for each season of rainfall, for each year of 1950–2007 and each forecasting lead time. Figure 10 shows plots of median lhf for seasonal rainfall during MJJ, ASO, NDJ and FMA. A score of 0.0 indicates lack of model performance in capturing the pdf of climatology, whereas a score greater than 1.0 indicates better model performance than climatology. The darker shading represents higher lhf scores or better model performance. From the median lhf of 1950–2007 rainfall ensembles (Fig. 10(a)), high scores can be found corresponding to the forecast of MJJ and ASO rainfall at long lead times (e.g. 7–9 months and 11 months). The successful long-lead forecasting of the modified k-nn model indicating higher skills than climatology (i.e. lhf scores greater than 1.0) was hardly found in previous studies (Kulkarni, 2000, DelSole and Shukla, 2002). The modified k-nn model also performed better than the climatology of MJJ and ASO rainfall at short forecasting lead times, e.g. 1–2 months (Singhrattna et al. 2005a, Grantz et al. 2005). However, associated with lhf scores of less than 1.0, the forecasting performance of the model with dry-season (i.e. NDJ and FMA) rainfall decreases considerably even at short lead times because, in terms of physical mechanism, the dry-season rainfall over the study basin is influenced by unstable local conditions, e.g. higher surface temperature and humidity, at finer time scales, e.g. hourly and daily (TMD 2007). Hence, the modified k-nn model, which is developed based on the relationships between rainfall and large-scale atmospheric variables at seasonal time scales, performs better in monsoon rainfall forecasting than dry-season rainfall forecasting.

To compare the prediction ability of the modified k-nn model in extreme years, Fig. 10(b) and (c) present plots of median lhf for dry rainfall years (rainfall below the 33rd percentile) and wet rainfall years

### Table 2 Summary of the large-scale atmospheric predictors identified from the correlation maps for seasonal rainfall during May-June-July (MJJ), August-September-October (ASO), November-December-January (NDJ) and February-March-April (FMA).

<table>
<thead>
<tr>
<th>Season of rainfall</th>
<th>Atmospheric variables</th>
<th>Spatial coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>MJJ</td>
<td>SAT 20.0°N, SLP 7.5–10.0°N, SXW 0°, SYW 0–2.5°N, RAIN 15.30–19.36°N</td>
<td>97.5–101.75°E</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Longitude</td>
</tr>
<tr>
<td>ASO</td>
<td>SAT 2.5–5.0°N, SLP 17.5–20.0°N, SXW 10.0°N, SYW 10.0°N, RAIN 15.30–19.36°N</td>
<td>107.5–110.0°E</td>
</tr>
<tr>
<td>NDJ</td>
<td>SAT 2.5–7.5°S, SLP 2.5–5.0°S, SXW 20.0–22.5°N, SYW 20.0–22.5°E, RAIN 15.30–19.36°N</td>
<td>97.5–100.70°E</td>
</tr>
<tr>
<td>FMA</td>
<td>SAT 7.5°S, SLP 15.0°N, SXW 17.5°N, SYW 2.5°S–2.5°N, RAIN 15.30–19.36°N</td>
<td>110.0–112.5°E</td>
</tr>
</tbody>
</table>

SAT: surface air temperature; SLP: sea-level pressure; SXW: surface zonal wind; SYW: surface meridian wind; RAIN: observed rainfall over study basin.

**Fig. 9** Correlogram or lag autocorrelation (ACF) of monthly rainfall for 1950–2007. The dashed lines represent the correlations being significant at the 95% confidence level.
Table 3 Summary of the optimal sub-sets of predictors (●) identified from the cross-validated multiple regression for seasonal rainfall during (a) May-June-July (MJJ), (b) August-September-October (ASO), (c) November-December-January (NDJ) and (d) February-March-April (FMA). (SAT: surface air temperature; SLP: sea-level pressure; SXW: surface zonal wind; SYW: surface meridian wind; RAIN: observed rainfall over study basin.)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>SAT</th>
<th>SLP</th>
<th>SXW</th>
<th>SYW</th>
<th>RAIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Season of predictors (forecasting lead time in month)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) MJJ Rainfall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FMA (12)</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>MAM (11)</td>
<td>α</td>
<td>●</td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>AMJ (10)</td>
<td></td>
<td>●</td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>MJJ (9)</td>
<td></td>
<td></td>
<td>●</td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>JJA (8)</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>JAS (7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>ASO (6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>SON (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>OND (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>NDJ (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>DJF (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>JFM (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>(b) ASO Rainfall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MJJ (12)</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>JJA (11)</td>
<td>α</td>
<td>●</td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>JAS (10)</td>
<td></td>
<td>●</td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>ASO (9)</td>
<td></td>
<td></td>
<td>●</td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>SON (8)</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>OND (7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>NDJ (6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>DJF (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>JFM (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>FMA (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>MAM (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>AMJ (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>(c) NDJ Rainfall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASO (12)</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>SON (11)</td>
<td></td>
<td>●</td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>OND (10)</td>
<td></td>
<td></td>
<td>●</td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>NDJ (9)</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>DJF (8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>JFM (7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>FMA (6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>MAM (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>AMJ (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>MJJ (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>JJA (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>JAS (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>(d) FMA Rainfall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDJ (12)</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>DJF (11)</td>
<td>α</td>
<td>●</td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>JFM (10)</td>
<td></td>
<td>●</td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>FMA (9)</td>
<td></td>
<td></td>
<td>●</td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>MAM (8)</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>AMJ (7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>MJJ (6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>JJA (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>JAS (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>ASO (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>SON (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>OND (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
</tbody>
</table>
Hydroclimate variability and long-lead forecasting of rainfall over Thailand by large-scale atmospheric variables

Fig. 10 Median likelihood (lhf) scores for seasonal rainfall during May-June-July (MJJ), August-September-October (ASO), November-December-January (NDJ) and February-March-April (FMA) for: (a) 1950–2007, (b) dry years, and (c) wet years.

For MJJ rainfall, the lhf scores of dry years are higher than those of wet years, especially at long lead times, e.g. 9–12 months. In contrast, the modified k-nn model shows the greater predictability of ASO rainfall in wet years than in dry years. However, the highest lhf scores (i.e. 2.0–2.5) for ASO rainfall forecast in both dry and wet years can be found associated with the 6-month forecasting lead time, which shows an impressive ability to forecast extreme events with a long lead period. It is also noted that the better performance than climatology (i.e. lhf greater than 1.0) of wet ASO rainfall can be observed for all 12 forecasting lead times. In addition, the good performance of long-lead rainfall forecasts is obtained for dry NDJ (and short-lead forecasts for wet NDJ). For FMA rainfall, the forecast of dry years shows better performance than wet years, in particular, at the forecasting lead time of 5–6 months.

Therefore, the large-scale atmospheric variables are identified and used as predictors of the modified k-nn model to forecast seasonal rainfall in the Ping River Basin for 12 forecasting lead periods. The optimal combination sets of predictors that are determined by GCV consist of 1–2 predictors. This can confirm the comparative performance of a fitting regression and forecasting model using the large-scale atmospheric predictors. Since the irrigated area of the study basin covers 2332 km², and the agriculture depends on both the rainfall during the monsoon season and the irrigated water (also based on rainfall), the ability of the modified k-nn model to make long-lead forecasts for monsoon rainfall, e.g. 7–9 months prior to the season starting, is very useful for water resources planning and management, agricultural practices and cropping strategies. Moreover, the operation planning of several storage dams along the Ping River including the Bhumipol Dam can be done in advance to deal with the extreme events, i.e. dry and wet episodes, and to efficiently serve the water demands of the Ping and Chao Phraya river basins, consisting of 6127 and 11 000 × 10⁶ m³ per year, respectively. The non-exceedence and exceedence probabilities of extreme events obtained from the pdf of rainfall forecasting ensembles serve as a tool for decision making in reservoir operation, and the modified k-nn model shows better performance in capturing the pdf than the climatology.

8 CONCLUSION

This study enhances the understanding of local hydroclimate variability and develops statistical relationships between rainfall and large-scale atmospheric variables. The land–sea temperature gradient plays a key role in strengthening or weakening the monsoon over the Upper Chao Phraya River Basin in Thailand. The amount of dry-season (November–April) rainfall determines the MAM temperature which is responsible for setting up the land–sea temperature gradient. Higher MAM temperature tends to increase monsoon (i.e. ASO) rainfall. Due to the anomalous atmospheric variables, the below-normal rainfall corresponds to the El Niño phase of ENSO and, conversely, the above-normal rainfall corresponds to the La Niña phase.

Using the correlation maps, the large-scale atmospheric variables are identified as the predictors
of seasonal rainfall. The significant relationships between rainfall and large-scale atmospheric variables indicate the long-lead forecasting ability of forecasting models. There are five identified predictors for each of MJJ, ASO and FMA rainfall, and four identified predictors for NDJ rainfall. Based on the cross-validated multiple regression, optimal combinations of predictors are selected separately for each season of rainfall and forecasting lead period. The selected combinations composed of one or two mutually exclusive predictors present the comparative performance of a regression and forecasting model by incorporation of the large-scale atmospheric variables.

The modified k-nn model is ultimately developed to forecast rainfall at long lead times using the optimal sets of predictors. Based on lhf scores, the modified k-nn model shows the ability to give long-lead forecasts of MJJ and ASO rainfall at 7–9 months lead time. However, decreasing performance is found in dry-season rainfall forecasts at both short and long lead times. The ability to forecast rainfall in dry MJJ is more significant than in wet MJJ, especially at lead times of 9–12 months. The best performance of ASO rainfall forecast in extreme years, i.e. dry and wet, is observed at a 6-month lead time. The long-lead forecasting ability of the modified k-nn model for rainfall in the Ping River Basin is an impressive result that can be applied to manage water resources and develop the long-term plans for reservoirs.

Acknowledgements The authors would like to thank Dr Sutat Weesakul and Dr Kiyoshi Honda for their valuable comments. The authors also thank the Royal Irrigation Department of Thailand, the Thailand Meteorological Department and the Department of Water Resources for providing the data for the study. The first author is grateful to the Asian Institute of Technology, Thailand for the graduate fellowships, to the Department of Public Works and Town & Country Planning for its kind support and to Piyawat Wuttichaikitcharoen for providing some data used in this research. The research funding from CIRAD is gratefully acknowledged. Finally, the authors would like to thank the anonymous reviewers whose comments improved the manuscript.

REFERENCES
Hydroclimate variability and long-lead forecasting of rainfall over Thailand by large-scale atmospheric variables


Owosina, A., 1992. Methods for assessing the space and time variability of ground water data. Thesis (MSc), Utah State University, Utah, USA.


