Seasonal Forecasting of Thailand Summer Monsoon Rainfall

Nkrintra Singhrattna[1, 2], Balaji Rajagopalan[+1, 3], Martyn Clark[3] and K. Krishna Kumar[3, 4]

[1] Department of Civil, Environmental and Architectural Engineering, University of Colorado at Boulder, Boulder, USA

[2] Thailand Public Works Department, Bangkok, Thailand

[3] Co-operative Institute for Research in Environmental Sciences, University of Colorado, Boulder, USA

[4] Indian Institute of Tropical Meteorology, Pune, India

Abstract

This paper describes the development of a statistical forecasting method for summer monsoon rainfall over Thailand. Predictors of Thailand summer (August-October) monsoon rainfall are identified from the large-scale ocean-atmospheric circulation variables (i.e., sea surface temperature and sea level pressure) in the Indo-Pacific region. The identified predictors are part of the broader El Niño Southern Oscillation (ENSO) phenomenon. The predictors exhibit significant relationship to the summer rainfall only during the post-1980 period when the Thailand summer rainfall also shows a relationship with ENSO. Two methods for generating ensemble forecasts are adapted. The first is the traditional linear regression, and the second is a local polynomial based non-parametric method. The associated predictive standard errors are used for generating ensembles. Both the methods exhibit significant comparable skills in a cross-validated mode. However, the nonparametric method shows improved skill during extreme years (i.e. wet and dry years). Furthermore, the models provide useful skill at 1~3 month lead time that can have strong impact on resources planning and management.

Key Words: Thailand summer rainfall; Monsoon; ENSO; Ensemble Forecast; Nonparametric methods; Local Polynomials; Seasonal Forecasting

International Journal of Climatology
November, 2004

[+] Corresponding Author

1. Introduction

Seasonal forecasts of Thailand summer monsoon rainfall can have significant value for resources planning and management - e.g., reservoir operations, agricultural practices, and flood emergency responses. In particular, increased population stress on the Chao Phraya River basin, one of the key regions for Thailand's socio-economic well being, is resulting in water quantity and quality problems. To mitigate this, effective planning and management of water resources is necessary. In the short term, this requires a good idea of the upcoming monsoon season rainfall – i.e. good seasonal forecast. In the long term, it needs realistic projections of scenarios of future variability and change. There is no known long-lead forecast of Thailand summer monsoon rainfall or stream flows. As a result, much of the water a resource planning in Chao Phraya basin and in Thailand in general is near term – i.e. responding to near term weather forecast.

There is a rich literature of studying the variability of Indian summer monsoon both from observational data (e.g., Walker, 1924; Rasmussen and Carpenter, 1983; Pant and Parthasarathy, 1981; Fein and Stephens, 1987; Webster et al., 1998) and from modeling studies (e.g., Ju and Slingo, 1995; Meehl and Arblaster, 1998). These studies have identified a strong link between El Nino Southern Oscillation (ENSO) and the Indian summer monsoon. Statistical methods for forecasting the Indian monsoon rainfall use this ENSO - monsoon relationship. For example, Krishna Kumar et al. (1995) and Shukla and Mooley (1987) identify several predictors of the Indian monsoon and developed statistical models for forecasting – almost all of the predictors are various facets of ENSO. With this framework of predictors, statistical models using time series (Thapliyal, 1981, Rajeevan 2001.) and artificial neural network techniques (Sahai et al.,

2000) have been developed by the Indian Meteorological Department and other researchers for use in operational forecast.

Krishna Kumar et al. (1999a) showed that the ENSO-Indian monsoon relationship has substantially weakened in the post-1980 period. They argue for changed ENSO characteristics and global warming as potential causes for this weakening. This is having a strong impact on the forecasting efforts of Indian monsoon as most of its predictors (mentioned above) are related to ENSO. Furthermore, Krishna Kumar et al. (1995, 1999b) show that the Indian monsoon predictors are strongly related to the Indian monsoon only when the monsoon itself is strongly related with ENSO. Interestingly, results from our research (Singhrattna et al., 2004) indicated that the Thailand monsoon is more closely related to ENSO in the post-1980 period, just when the Indian monsoon relationship with ENSO is weakening. This enhances the prospects of forecasting Thailand monsoon rainfall.

There is little in the literature studying the variability and predictability of Thailand summer monsoon. Admittedly, it is much smaller in comparison to the Indian summer monsoon but has a significant socio-economic impact in Thailand. There have been some studies of late on the variability of Thailand monsoon and rainfall over Singapore and Indonesia by Kripalani and Kulkarni (1997, 1998 and 2001) and more recently by Singhrattna et al., 2004 and Singhrattna 2003. Distributed hydrologic models for Chao Phraya and Nakon Sawan river basins have been developed (Jha et al., 1997, 1998), but these are mainly for real time or event based simulation of stream flow and not for seasonal forecasting.

It is not clear if there is a seasonal forecast mechanism in place in Thailand. Unlike in the case of the Indian summer monsoon, the Indian Meteorological Department is required to issue a seasonal forecast of the upcoming monsoon season by the end of April. The great need and utility of the Thailand monsoon forecast and the enhanced prospects of its predictability in recent decades, serves as a strong motivation for the present research.

We adapt two approaches for ensemble forecast of Thailand summer monsoon rainfall in this paper. The first is a traditional linear regression approach and the second a nonparametric technique based on local regressions.

The paper is organized as follows. Data description and predictor identification are first presented. The two forecasting methods are next described followed by cross-validated model skills in forecasting the Thailand summer rainfall. Discussions of the results conclude the paper.

2. Data

The data used in this study are:

1. Rainfall data for the Thailand summer monsoon (August – October), and surface air temperature (SAT) data during pre-monsoon months (March – June), averaged over three stations, Nakhon Sawan (15°48'N, 100°10'E), Suphan Buri (14°28'N, 100°08'E) and Don Muang (13°55'N, 100°36'E). All of these stations are in the West Central region and in the Chao Phraya River basin. These data were obtained from the GEWEX Asian

Monsoon Experiment (GAME) project website[1]. The GAME program, part of the global energy and water cycle experiment (GEWEX) has done a good job of collecting and archiving data from South East Asian countries. In general, it has been difficult obtaining long hydroclimate data from South East Asia and Thailand in particular. Please see Singhrattna (2003) for further details on these data sets.

2. Large-scale ocean and atmospheric circulation variables such as sea surface temperature (SST); sea level pressure (SLP), winds, velocity potential were obtained from NCEP/NCAR Re-analysis (Kalnay et al., 1996). These data sets span the period of 1948 – current, covering the globe on a 2.5°x2.5° grid and available at http://www.cdc.noaa.gov

3. Standard ENSO indices: NINO3, NINO1+2, Southern Oscillation Index (SOI) available at http://www.cpc.noaa.gov

4. Indian Ocean Dipole (IOD) index (Saji et al. 1999). This is an index based on SST anomaly difference between the Eastern and Western tropical Indian Ocean. The index, its impact on the adjoining continental rainfall, interactions with ENSO and teleconnections can all be obtained from the IOD home page http://www.jamstec.go.jp/frsgc/research/d1/iod/.

3. Identification of Predictors

The aim in this section is to identify predictors for the Thailand summer rainfall, which can then be used in statistical forecast models. The two main requirements for any

[1] http://hydro.iis.u-tokyo.ac.jp/GAME-T/GAIN-T/routine/rid-river/longterm.html

useful predictors are (i) good relationship with the seasonal rainfall and (ii) reasonable lead-time (i.e. months to season). Our earlier work (Singhrattna 2003, Singhrattna et al. 2004) indicated that Thailand summer rainfall is strongly correlated with ENSO in the post-1980 period and also with pre-monsoon (especially, Mar-May) land surface temperatures representing the land-ocean thermal gradient. So, the first step is to look for relationship with standard ENSO indices during the pre-monsoon seasons and follow up with correlations between the rainfall and large-scale ocean-atmospheric variables (SSTs, SLPs). This approach of correlation with large-scale ocean-atmospheric circulation variables has been used to identify predictors for stream flows in Northern Brazil (DeSouza and Lall 2003) and in the Truckee-Carson river basins in NV, USA (Grantz 2003).

3.1 Correlation with ENSO indices

Thailand summer monsoon rainfall was correlated with the standard ENSO indices and IOD index from pre-monsoon seasons and also with the spring (March-May, MAM) Thailand air temperatures (SAT). The latter is believed to be an indicator of the land-ocean thermal gradient that is important for the strength of the monsoon (Singhrattna et al. 2004). The correlations are computed for the post-1980 period and shown in Table 1. Correlation values that are statistically significant at the 95% confidence level using a t-test (Helsel and Hirsch 1995) are shown in bold in the table. It can be seen that the SLP-based ENSO index, SOI, shows a strong correlation with monsoon rainfall during the concurrent season and also 1-2 seasons prior. The spring land temperatures also exhibit a significant correlation as expected. The IOD shows a strong correlation with the monsoon rainfall at 2-season lead-time. All these brighten the prospects for a long-lead forecast.

To confirm that the correlations are strong only during the post-1980 period (as in Table 1), selected predictors from pre-monsoon seasons (JFM NINO3, MJJ SOI, MAM IOD and MAM SAT) were correlated with monsoon rainfall on a 21-year moving window (Figure 1). It can be seen that the predictors show correlations with summer rainfall only in recent decades, much as the correlations between the rainfall and ENSO (shown in solid line between summer rainfall and ASO SOI) and seen by Singhrattna et al. (2004). Similar shifts have been seen (Miyakoda et al., 2003) in pre-monsoon signals of South Asian monsoon. This suggests that the ENSO-based predictors are related to the monsoon rainfall only when the monsoon rainfall itself is related to ENSO. Interestingly, this is similar to the finding by Krishna Kumar et al. (1995) where they show that the predictors of Indian monsoon rainfall are related to the rainfall only during the period when the Indian monsoon is strongly related with ENSO. In the case of the Indian monsoon this is pre-1980 period. This is consistent with the ENSO related circulation changes during pre and post-1980 periods (Krishna Kumar et al., 1999a; Singhrattna et al., 2004). Land cover changes (Kanae et al., 2001) and decadal changes in ENSO-monsoon relationship (Krishna Kumar et al., 1999b; Torrence and Webster, 1999) could lead to trends in monsoon precipitation and consequently, add to the non-stationarity of the relationship as seen in Figure 1.

3.2 Correlation with Large-Scale Variables

While the indices show significant correlations as seen above, we would like to check their large-scale aspects and also to see if other stronger predictors could be identified. To this end, the summer monsoon rainfall was correlated with SSTs and SLPs during pre-monsoon seasons and the correlation maps are shown in Figure 2. The shaded regions indicate correlations that are significant at 95% confidence level. With SLPs (Figure 2(a)) the correlations are strong in the Pacific subtropical region indicating that a higher than normal subtropical pressure tends to enhance the easterlies and thereby increasing the moisture transport to the Thailand and consequently, the rainfall. Wang et al. (2003, see their Figures 1 and 2) found similar pressure patterns in the Pacific subtropical region to be linked with variations in the Australian and Asian monsoons. Strong positive correlations with SSTs (Figure 2(b)) are seen in the Eastern Indian Ocean and Western Pacific Ocean regions around the equator. This region is also one of the poles of the Indian Ocean Dipole index (Saji et al. 1999) and hence, the strong correlation with IOD seen in Table 1 and Figure 1. These correlation maps indicate persistence from spring leading up to the monsoon season, thus providing the potential for long-lead forecast. The solid box in the figures shows the regions of high correlation from where the predictors will be developed in the following sections.

3.3 Predictor Selection

Based on the correlations with indices and the correlation maps with large-scale variables, predictors with high correlations to the summer rainfall were identified. With this criterion, the selected predictors are (i) SSTs averaged over 10.5^{o}S-14.5^{o} S latitudes and 108^{o}-120^{o} E longitudes and (ii) SLPs averaged over 20^{o}N-30^{o}N latitudes and 165^{o}-180^{o} E longitudes. Thailand surface air temperature (SAT) is also selected as one of the predictors. This essentially captures the land-ocean gradient that gets set up by the land temperatures, especially during the Spring season before the monsoon (Singhrattna, 2003).

To check the temporal variability of the strength of the predictors to monsoon rainfall, moving window correlations are shown in Figure 3. As expected, the predictors are correlated mainly in the post-1980 period as in Figure 1. Furthermore, the predictors show significant correlations with the summer rainfall at 1~2-season lead-time.

4. Forecast Models

Typically, a regression (often linear) is fit between the identified predictors and a single dependent variable (i.e. the summer rainfall). The fitted regression is then used to forecast the mean value of the variable. There is a rich literature for fitting and testing linear regression models, and software is extensively available (e.g., Helsel and Hirsch 1995). Such models have been widely used for hydroclimate forecasting in the US (e.g., Lui et al. 1998, Piechota et al. 2001, Cordery and McCall 2000; Mccabe and Dettinger, 2002) for the Indian monsoon forecasting (Hastenrath, 1987, 1988; Krishna Kumar et al. 1995). For forecasting a field of dependent variable such as precipitation at several locations from fields of independent variables (e.g., tropical SST, SLP etc.), Canonical Correlation Analysis is typically used (e.g., Shabbar and Barnston, 1996; Ntale et al., 2003). Below, the linear regression model is briefly described.

4.1 Linear Regression

Traditional linear regression involves fitting a linear function between the response variable (i.e. summer rainfall) and the independent variables (i.e. predictors). They are of the form:

$$Y_t = a_1 * x_{1t} + a_2 * x_{2t} + a_3 * x_{3t} + \ldots + a_p * x_{pt} + e_t \qquad [1]$$

$$t = 1,2,\ldots N$$

Where the coefficients a_1, a_2,…, a_p are estimated from the data, typically, minimizing the sum of squares of the errors; e_t is the error which is assumed to be Normally (or Gaussian) distributed with mean 0 and variance σ_e^2 (also estimated from the data) and N is the number of observations. The equations for the coefficients, the error variance and methods for testing the goodness of the fitted model can be found in any standard book on statistics (e.g., Helsel and Hirsch, 1995).

Implicitly, the variables are also assumed to be normally distributed. If not, they are generally transformed to a normal distribution (e.g., log or power transform) before the model is fit. Once the model is fit (i.e. the coefficients estimated) then for any new value of the predictors, the model with the fitted coefficients (Equation [1]) is used to predict the mean value of the dependent variable, say, Y_{new}. Predictive standard error, σ_{pe}, (or the standard deviation of the error of the predicted mean) is obtained from the theory (Helsel and Hirsch 1995). Normal random deviates with a mean of 0 and standard deviation σ_{pe} provide the ensembles of errors, when added to the mean estimate, Y_{new}, results in an ensemble forecasts. This approach of using Normal distribution with the predictive standard error was applied by Clark and Hay (2004) for generating ensemble forecasts of stream flows in the Western US.

In the above model, if the independent variables happen to be past values of the response variable itself, then it forms a time series model of Auto Regressive framework. Hydrologists have developed and used such models for stream flow simulation and forecast (Salas 1985, Yevjevich 1972, Bras and Iturbe 1985).

The main drawbacks of traditional linear regression models are (i) assumption of Gaussian distribution of data and errors, (ii) assumption of linear relationship between the variables and, (iii) not portable across data sets (i.e. sites). Furthermore, if the fitted model is found to be inadequate then the alternative choices are limited, more so when the number of observations are small.

4.2 Nonparametric regression – Locally Weighted Polynomials

Nonparametric methods provide an attractive alternative in alleviating some of the drawbacks of the traditional linear regression. In this approach, the model is:

$$Y_t = f(\mathbf{x_t}) + e_t \qquad [2]$$

Where, $\mathbf{x_t} = (x_{1t}, x_{2t}, x_{3t}, \dots x_{pt})$, t = 1,2,…$N$

This is similar to the linear regression model (Equation [1]) but the function f could be linear or nonlinear, and the errors, e_t, are assumed to be normally distributed with mean 0 and variance σ_{le}^2. The key difference from linear regression is that the function f is fit "locally" to estimate Y. In that, the value of the function at any point '$\boldsymbol{x_i}$' is obtained by (i) identifying a small number K (= $\boldsymbol{a}$ *N, where $\boldsymbol{a} \in (0,1]$) of neighbors to '$\boldsymbol{x_i}$' and (ii) fitting a polynomial of order (p) to the neighbors. Neighbors are identified from the observations that are closest to ‘$\boldsymbol{x_i}$’ in terms of the Euclidian distance or other such metric (e.g., Mahalanobis distance, Yates et al., 2003). The fitted polynomial is then used to estimate the mean value of the dependent variable. The coefficients of the polynomial are estimated using weighted least squares approach. The theoretical background of the local polynomial method is described in detail in Loader (1999) and the author refers to it as LOCFIT – henceforth, we will use the same terminology in this paper.

LOCFIT also provides the local standard errors of the estimate σ_{le}, and local predictive standard errors σ_{lpe} (Loader 1999), corresponding to σ_e and σ_{pe}, respectively, in the case of linear regression described in the previous section. The steps for generating the ensembles are same as that for the linear regression: (i) for a new value of the predictor set, the mean value, Y_{new}, is first estimated using the LOCFIT approach described above, (ii) the local predictive standard error σ_{lpe} are estimated (Loader 1999) and (iii) Normal random deviates with a mean of 0 and standard deviation of σ_{lpe} when added to the mean estimate Y_{new}, result in ensemble forecasts.

The key parameters to be estimated are the size of the neighborhood (K or α) and the order of the polynomial (p). These parameters are obtained using objective criteria such as Generalized Cross Validation (GCV) function, Likelihood function:

$$GCV(\alpha , p) = \frac{\sum_{i=1}^{N} \frac{e_i^2}{N}}{\left(1 - \frac{m}{N}\right)^2} \qquad [3]$$

where e_i is the error (i.e. difference between the model estimate and observed), N is the number of data points and m is the number of parameters. For a suite of a and p values the GCV function is computed from the above equation and the combination that gives the least GCV value is selected. For stability purposes, the minimum neighborhood size should be twice the number of parameters to be estimated in the model.

Note that if a first order (i.e. linear) polynomial is selected, and if the neighborhood includes all the observations (i.e., K= N or a =1) this then results in the traditional linear regression. Thus, LOCFIT can be viewed as a superset. We used the software LOCFIT developed by Loader and is available on-line[2].

There are several nonparametric approaches to estimating the function f locally, such as, kernel-based (Bowman and Azzalini 1997), Splines, local polynomials (Rajagopalan and Lall 1998; Owosina 1992; Loader, 1999). Owosina (1992) performed an extensive comparison of a number of regression methods both parametric and nonparametric on a variety of synthetic and real data sets. He found that the nonparametric methods handily outperform parametric alternatives. All of the nonparametric methods perform similarly but LOCFIT is easy to implement, hence we adapted it in this paper.

LOCFIT has been used for several hydroclimate applications (Lall, 1995) – for spatial interpolation of precipitation (Rajagopalan and Lall, 1998); salinity modeling (Prairie et al., 2003a; Prairie, 2002); stream flow modeling (Prairie et al., 2003b, 2002); stream flow forecasting (Grantz, 2003) and flood frequency estimation (Apipattanavis et al., 2004).

[2] http://cm.bell-labs.com/cm/ms/departments/sia/project/locfit/index.html

Variants of LOCFIT also provide attractive alternative to ensemble generation. For example, the (*K*) neighbors of an estimation point '$\boldsymbol{x_i}$' identified can be re-sampled (i.e. bootstrapped) with a weight function that gives more weights to the nearest neighbor and less to the farthest, thus, generating ensembles. Lall and Sharma (1996) developed this approach and used it for stream flow simulation. Later Rajagopalan and Lall (1999) and Yates et al. (2003) extended it for stochastic daily weather generation and DeSouza and Lall (2003) applied it for stream flow forecasting.

4.3 LOCFIT with Resampled Residuals (Modified K-NN)

Often times the errors e_t are not Normally distributed. To address this issue a modification to LOCFIT was developed by Praire (2002). Prairie et al. (2003a,b) applied this for stream flow and salinity modeling. Later, Grantz (2003) demonstrated the use of this approach for stream flow forecasting on the Truckee-Carson basin in Nevada, USA.

Prairie (2002) referred this as the "Modified *K*-NN", and we do the same in this paper, henceforth. The modification is described below.

Suppose an ensemble is required for a new value of the predictor $\boldsymbol{x}_{new}$, and suppose that the polynomial order (*p*) and the size of the neighborhood (*K*) have been obtained using GCV or other objective criteria. The steps in the modification are as follows:

(i) Identify *K* nearest neighbors to $\boldsymbol{x}_{new}$ and fit a polynomial of order *p*. The fitted polynomial provides the estimate of the dependent variable at all the neighbors and consequently, the residuals.

(ii) The fitted polynomial from step (i) is used to estimate the mean value Y_{new}. (This step is just the LOCFIT process described in the previous section).

(iii) Now select one of the (*K*) neighbors of $\boldsymbol{x_{new}}$, say $\boldsymbol{x_i}$ and select the corresponding residual e_i (already obtained from step (i)), this is now added to the mean estimate $Y_{new} + e_i$; thus, obtaining one of the ensemble members. The selection of one of the neighbors is done using a weight function

$$W(j) = \frac{1}{j\sum_{i=1}^{K}\frac{1}{i}} \qquad [4]$$

As can be seen, this weight function gives more weight to the nearest neighbor and less to the farthest neighbors.

Repeat step (iii) several times, resulting in an ensemble.

The number of neighbors for fitting the local polynomial can be different from the neighbors used to resample the residuals (e.g., Prairie, 2002). In this work we have kept both to be the same. In the modification described above, if the number of observations (N) is small then the re-sampled residuals (step (iii) above) provide very limited variety in the ensembles and this is the main disadvantage.

5. Model Evaluation

The models are verified in a cross-validated mode. In that, the data (rainfall and the predictors) for a given year is dropped out and the model(s) based on the rest of the data is applied to generate ensemble forecast for the dropped year. This is repeated for all the years for the 1980 – 2000 period. Apart from visual inspection, the ensembles are evaluated on three criteria:

(i) Correlation between the observed value and the median of the ensemble forecast. This is much like evaluating the mean forecast that would come from a standard linear regression model.

(ii) Likelihood function (LLH) (Rajagopalan et al. 2002). This evaluates the skill of the model in capturing the Probability Density Function (PDF).

(iii) Rank Probability Skill Score (RPSS) (Wilks 1995). This evaluates the skill of the model in capturing the categorical probabilities (i.e. the probability distribution function).

The likelihood function (LLH) is applied to measure the skills of forecast models. Its process is to categorize forecasted values to three divisions: below, normal and above normal. The ensemble forecasts falling into these three categories are compared to historical data and then develop a skill score. The likelihood skill score for any given year of forecast is defined as:

$$LLH = \frac{\prod_{t=1}^{N} \hat{P}_{j,t}}{\prod_{t=1}^{N} P_{cj,t}} \qquad [5]$$

Where N is the number of years to be forecasted, j is the category of the observed value in year t, $\hat{P}_{j,t}$ is the forecast probability for category j in year t, and $P_{cj,t}$ is the climatological probability for category j in year t. Here we divided the rainfall into three categories at the 33rd and 66th percentile, so the probabilities of each of the category are

1/3 and N is length of data. The LLH values vary from 0 to number of categories (i.e. 3 in this study). The score of zero indicates lack of skill, a score of greater than 1 indicates that the forecasts have skill in excess of the climatological forecast and a score of 3 indicates a perfect forecast.

The ranked probability skill score (RPSS) is also applied to quantify the skills of forecast models. This method evaluates the probability of ensemble forecasts falling into many categories (i.e. in this study: below average, average and above average) and compared to historical data. The RPSS score for any given year is defined as:

$$RPS(p,d) = \frac{1}{R-1}\left[\left(\sum_{i=1}^{R} P_i - \sum_{i=1}^{R} d_i\right)^2\right] \qquad [6]$$

for *R* mutually exclusive and collectively exhaustive categories (in this case we have three categories, so *R* = 3). The vector *d* (d_1, d_2, ... d_R) represents the observation vector such that d_R equals 1 if the observation fell in category *'R'* or 0 otherwise. The RPSS is then calculated as (e.g., Toth, 2002; Wilks, 1995)

$$RPSS = 1 - \frac{RPS(\text{forecast})}{RPS(\text{climatolog y})} \qquad [7]$$

RPSS scores vary from +1 to -∝ (i.e. perfect skill to bad skill). Scores above 0 indicate improvement over climatological forecast.

For the LOCFIT and Modified *K*-NN methods, due to small sample size, we used polynomial of order (*p*=1, i.e. local linear fit). However, the neighborhood size was objectively obtained using the GCV criteria.

6. Results

From the set of predictors (based on SST, SAT, SLP fields and ENSO indices) identified in the previous section, the optimal subset was found by the combination that gave the best-forecast skill. Several formal methods are available for subset selection – such as stepwise regression or cross-validation metrics etc. Since the number of significant predictors is small, in our case, almost all combinations were tried out to find the optimal predictor set. For summer monsoon rainfall, the best set of predictors were found to be the ones based on SLP and SST that are described in section 3.3. The land temperatures (SAT) did not seem to improve the skill much. Forecasts were issued at the beginning of each month starting April 1st for each year, using all the three methods and, the predictors are the average values from the preceding season (i.e. preceding three months). Except for forecasts issued on July 1st and August 1st, SST predictor of Mar-May and the SLP predictor of preceding season are used – as this combination gave the best skill. Thus, July 1st forecast is based on SST predictor of Mar-May and SLP predictor of Apr-Jun and, August 1st forecast is based on SST predictor of Mar-May and SLP predictor of May-Jul.

The skills of the forecasts are evaluated using the three skill measures described in the previous section. Skills are also compared during high (wet) and low (dry) years.

Threshold exceedance probabilities during the extreme years are estimated and the PDFs of ensembles of a few representative years are also presented.

The skill scores are shown in Figure 4. It can be seen that the skill increases significantly as the forecast lead-time decreases for all the methods. This is intuitive and consistent with expectations. The linear regression and LOCFIT show similar skills on all the three measures. This indicates that for the most part the relationship between the predictors and the rainfall is linear, and that linear regression seems appropriate. The Modified K-NN is comparable in performance as the lead-time decreases, but early on, its performance is weak. This we believe is due to the small sample size of residuals used in resampling. Given that we only have 21 data points (the data used for forecast is for the period 1980 – 2000) *K* tends to be of the order of 7~8. With a small *K*, coupled with the fact that at long leads, the relationship between predictors and rainfall is not as strong, consequently, there is less variety in the ensembles and a bias if the predictors are not very useful - which leads to poor skill scores.

Notice the significant skill from May 1st onwards, providing a 2-month lead-time that can be very useful for resources planning and management. This useful long-lead skill, regardless of the method, is quite impressive.

In order to compare the performance of these models in extreme years, Figures 5 (a) and (b) show the skill scores for the High (wet) and Low (dry) rainfall years defined in Singhrattna (2003). Interestingly, the nonparametric models (LOCFIT and Modified K-NN) seem to show slight improvement over linear regression for forecasts starting May 1st in both the wet and dry years and generally for all the skill measures. This could be explained by the fact that subtle nonlinear relationship exists between the predictors and the rainfall at the extremes and hence there is some advantage to use nonparametric methods. Furthermore, notice that the skill in wet years is much more than that in the dry years.

PDFs of the ensemble forecasts (solid line) made on August 1st during selected wet and dry years from the three models are presented along with the *climatological PDF* (dashed line), which is estimated from the entire historical record and, the observed values (dotted line) in Figure 6 and 7, respectively. For the wet years the Modified *K*-NN (Figure 6a) shows the ensembles to be shifted to the right of the climatological PDF. For the low years (especially, 1984 and 1994 Figure 7b) it can be seen that the LOCFIT method does a good job of shifting the forecast ensemble PDF to the left of the climatological PDF relative to the linear regression.

Even though the observed values are not in the middle of the ensemble PDFs (as we would like it to be) it still can provide useful information and skill in terms of threshold exceedance probabilities - one of the key variables for decision. We chose 700 mm (the 90th percentile of the data) as a surrogate for wet (or flood) conditions and 400 mm (the 10th percentile of the data) for dry conditions. From the PDFs of the ensembles forecast on May 1st, the exceedance probabilities are computed for the selected wet and dry years and shown in Table 2. For the wet years the climatological exceedance probability is 0.10, while the ensembles in all the years except 1995 indicate a very high probability of exceedance of this threshold, thus indicating a wet condition. This information, provided on May 1st, three months ahead of the summer monsoon season could be very helpful in flood emergency response planning and management. For the dry years the models show a small non-exceedance probability of the lower threshold (400mm) when a higher probability of non-exceedance is expected. This is consistent with the fact that the models have low skill in dry years especially with the May 1st forecasts (Figure 5b). However, we found the skill in these exceedance probabilities to be higher for June, July and Aug 1st forecasts. It can be seen that the nonparametric models in general show a slight improvement upon the linear regression model. Similar estimates were obtained from forecasts issued in other months.

The threshold exceedance probabilities can be used to effectively plan annual and seasonal reservoir, emergency response preparedness, flood plain management, cropping strategies, conservation measures etc. Furthermore, they can also be used as a surrogate for wetness or dryness and provide probabilistic information on flooding potential, land slides, etc. and develop optimal response strategies. Lastly, the ensembles of rainfall can be used to drive a water balance model and generate ensembles of stream flows. The forecasts will provide a useful and powerful tool to water managers in long-term planning that is currently lacking.

7. Summary

Predictors from large-scale ocean, atmosphere and land variables that have strong correlation with Thailand summer monsoon have been identified. The predictors are consistent in terms of their physical mechanistic links to the monsoon. The predictors indicate a 1-2 seasons worth of lead-time of predictability. Interestingly, the predictors are related to the monsoon rainfall, only during post-1980 period when the monsoon rainfall is correlated with ENSO, as seen in Singhrattna (2003). This suggests the tantalizing possibility that ENSO relationship could be modulating the predictability – similar to what is seen in the case of the Indian monsoon (Krishna Kumar et al. 1995; 1999b). The nonstationarity aspect of the relationship between the predictors and Thailand summer rainfall urges caution in that, the relationships have to be tested periodically and new predictors identified if necessary.

Two modeling approaches for ensemble forecasts of Thailand summer monsoon are offered – (i) traditional linear regression (parametric) and (ii) a nonparametric method based on local polynomials are adapted. Both the models exhibit significant skill at 2-5 months lead-time. The nonparametric method seems to show improved skills in the extreme years, especially in the wet years.

The proposed models for forecasting Thailand summer rainfall make a significant contribution as no official forecast models exist to our knowledge. This has tremendous implications to water management, early warning and preparedness and also for resources planning in general. Further testing and improvements of these models are required.

Acknowledgments

The first author (NS) is grateful to Thailand Government for the graduate student fellowship and to Thailand Public Works Department for their continued support. Travel support from IUGG to NS to present preliminary results from this work at the IUGG General Assembly 2003 in Sapporo, Japan is very much appreciated. Funding for the fourth author via CIRES Visiting Fellowship is also thankfully acknowledged. We thank the GEWEX Asian Monsoon Experiment (GAME) for the precipitation data and also providing a wealth of information on the Monsoon efforts in Asia. The authors are grateful to two anonymous reviewers whose comments greatly improved the manuscript. Finally, the authors would like to thank Somkiat Apipattanavis and Tom Chase for useful discussions.

Table 1: Correlations (post-1980 period) between Thailand summer rainfall (Aug – Oct) and large-scale climate indices

(The 95% significant level is ±0.41. Values in bold are statistically significant at 95%)

	JFM	**FMA**	**MAM**	**AMJ**	**MJJ**	**JJA**	**JAS**	**ASO**
Nino 1+2	**0.41**	0.31	0.29	0.28	0.25	0.17	0.08	-0.06
Nino 3	**0.42**	0.33	0.15	-0.01	-0.13	-0.19	-0.24	-0.31
Tahiti-Darwin (SOI)	0.40	0.27	-0.07	-0.27	**0.44**	**0.45**	**0.57**	**0.59**
Indian Ocean Dipole (IOD)	-0.37	**-0.44**	**-0.70**	**-0.55**	-0.32	-0.17	-0.22	-0.34
Air Surface Temperature (SAT)	0.30	**0.51**	**0.48**	0.34	0.20	0.10	-0.01	-0.11

Table 2: (a) Exceedance probabilities for selected wet years and (b) Non-exceedance probabilities for selected dry years

(a)

Wet Years				
Year	**Climatology**	**Modified *K*-NN**	**LOCFIT**	**Linear Regression**
1983	10.0%	81.0%	73.5%	71.3%
1988	10.0%	39.9%	54.7%	33.6%
1995	10.0%	3.1%	4.6%	1.1%

(b)

Dry Years				
Year	**Climatology**	**Modified *K*-NN**	**LOCFIT**	**Linear Regression**
1984	10.0%	1.0%	1%	1%
1987	10.0%	2.3%	3.7%	9%
1994	10.0%	1.0%	1.5%	1%

References

Apipattanavis, S., B. Rajagopalan and U. Lall, 2004: Local Polynomial Technique for Flood Frequency Analysis, (in review) *Journal of Hydrologic Engineering*

Bowman, A. W. and A. Azzalini, 1997: *Applied Smoothing Techniques for Data Analysis*. Oxford: Clarendon press.

Bras, R. L. and I. R. Iturbe,1985: *Random Functions and Hydrology*. Massachusetts: Addison-Wesley Publishing.

Clark, M., P. and L. E. Hay, 2004: Use of medium-range numerical weather prediction model output to produce forecasts of streamflow, *Journal of Hydrometeorology*, **5(1)**, 15-32.

Cordery, I. and M. A. McCall, 2000: A Model for Forecasting Drought from Teleconnections, *Water Res. Res.*, **36(3)**, 763-768.

DeSouza, Filho, F. A. and U. Lall, 2003: Seasonal to Interannual Ensemble Streamflow Forecasts for Ceara, Brazil: Applications of a Multivariate, Semi-Parametric Algorithm, *Water Res. Res.*, in press.

Fein, J. S. and P. Stephens, (Eds) 1987: *Monsoons*, Wiley-Interscience Publication, 632ppp, John Wiley and Sons, New York

Grantz, K, 2003: Usinig large-scale climate information to forecast seasonal streamflows in the Truckee and Carson rivers, Master of Science thesis, University of Colorado, Boulder, CO.

Hastenrath,S., 1987: On the prediction of Indian summer rainfall anomalies, J. Clim. Appl. Meteorol., 26, 847-857.

Hastenrath, S., 1988: Prediciton of Indian monsoon rainfall : further exploration, Bull. Amer. Met. Soc., 69, 819-825.

Helsel, D. R. and R. M. Hirsch, 1995: *Statistical Methods in Water Resources*. Amsterdam: Elsevier Science B. V.

Jha, R., Herath, S. and Musiake, K., 1997: Development of IIS Distributed Hydrological Model and its Application in Chao Phraya River Basin Thailand., *Ann. J. Hydraul. Eng. JSCE*, **41**, 227-232.

Jha, R., Herath, S. and Musiake, K., 1998: Application of IIS Distributed Hydrological Model in Nakon Sawan Catchment Thailand., *Ann. J. Hydraul. Eng. JSCE*, **42**, 145-150.

Ju, J. and J. M. Slingo, 1995: The Asian summer monsoon and ENSO, *Q. J. Royal Meteorological Societ*, **122**, 1133-1168.

Kalnay, E. and Coauthors, 1996: The NCEP/NCAR 40-year Reanalysis Project, *Bull. Amer. Meteor. Soc.*, **77**, 437-471.

Kanae, S., T. Oki and K. Musiake, 2001: Impact of deforestation on regional precipitation over the IndoChina peninsula, *Journal of Hydrometeorology*, **2**, 51-70.

Kripalani, R. H. and Ashwini Kulkarni, 1997: Rainfall Variability over South-East Asia Connections with Indain Monsoon and ENSO Extremes: New Perspectives. *Journal of Climatology*, **17**, 1155-1168.

Kripalani, R.H. and A. Kulkarni, 1998: The Relationship between Some Large-Scale Atmospheric Parameters and Rainfall over Southeast Asia: A Comparison with Features over India. *Theo. And Appl. Cli.*, **59**, 1-11.

Kripalani, R. H. and Ashwini Kulkarni, 2001: Monsoon Rainfall Variations and Teleconnections over South and East Asia. *Inter. J. of Cli.*, **21**, 603-616.

Krishna Kumar, K., M. K. Soman and K. Rupa Kumar, 1995: Seasonal Forecasting of Indian Summer Monsoon Rainfall: A Review, *Weather,* **50(12)**, 449-467.

Krishna Kumar, K., Balaji Rajagopalan and Mark A. Cane, 1999a: On the Weakening Relationship between Indian Monsoon and ENSO, *Science*, **284**, 2156-2159.

Krishna Kumar, K., R. Kleeman, M. A. Cane, and B. Rajagopalan, 1999b: Epochal changes in the Indian Monsoon – ENSO precursors, *Geophysical Research Letters,* **26(1)**, 75-78.

Lall, U, 1995: Recent advances in nonparametric function estimation, *U. S. Natl. Rep. Int. Union Geod. Geophys. 1991-1994, Reviews of Geophysics,* **33**, 1093-1102.

Lall, U. and A. Shama, 1996: A Nearest Neighbor Bootstrap for Resampling Hydrologic Time Series, *Water Res. Res.*, **32(3)**, 679-693.

Loader, C, 1999: *Local Regression Likelihood*. New York: Springer.

Lui, Z., J. B. Valdes and D. Entekhabi, 1998: Merging and Error Analysis of Regional Hydrometeorological Anomaly Forecasts Conditioned on Climate Precursors, *Water Res. Res.*, **34(8)**, 1959-1969.

Mccabe, G., and M. D. Dettinger, 2002: Primary modes and predictability of year-to-year snowpack variations in the Western United States from teleconnections with Pacific Ocean climate, *Journal of Hydrometeorology,* **(3)**, 13-25.

Meehl, G. A., and J. M. Arblaster, 1998: The Asian-Australian monsoon and El Niño-Southern Oscillation in the NCAR Climate System Model. *J. Climate*, **11**, 1356-1385.

Miyakoda, K., J. L. Kinter III and, S. Yang, 2003: The role of ENSO in the south Asian monsoon and pre-monsoon signals over the Tibetan plateau, *Journal of Meteorological Society of Japan*, **81(5)**, 1015-1039.

Ntale, H. K., T. Y. Gan and D. Mwale, 2003: Prediction of East African seasonal rainfall using canonical correlation analysis, *Journal of Climate,* **16(12)**, 2105-2112.

Owosina, A., 1992: Methods for Assessing the Space and Time Variability of Ground Water Data. *Thesis (M.S.),* Utah: Utah State University.

Pant, G.B. and Parthasarathy, B., 1981: Some aspects of an association between the southern oscillation and Indian summer monsoon. *Arch. Meteorol. Geophys. Biokl., Sr. B.*, 29, 245-251.

Piechota, T. C., F. H. S. Chiew, J. A. Dracup and T. A. McMahon, 2001: Development of an Exceedence Probability Streamflow Forecast, *ASCE J. of Hydrologic Eng.*, **6(1)**, 20-28.

Prairie, J. R. 2002: *Long-term Salinity Prediction with Uncertainty Analysis: Application for Colorado River above Glenwood Springs,* M.S. Thesis. Colorado: University of Colorado at Boulder

Prairie, J., B. Rajagopalan, T. Fulp and E. Zagona, 2003a: Statistical nonparametric model for natural salt estimation (in press) *ASCE Journal of Environmental Engineering*.

Prairie, J., B. Rajagopalan, T. Fulp and E. Zagona, 2003b: A modified K-NN Model for Generating Stochastic Natural Streamflows, (submitted to) *Journal of Hydrologic Engineerng.*.

Rajagopalan, B. and U. Lall, 1998: Locally Weighted Polynimial Estimation of Spatial Precipitation, *J. of Geographic In formation and Decision Analysis*, **2(3)**, 48-57.

Rajagopalan, B. and U. Lall, 1999: A Nearest Neighbor Bootstrap Resampling Scheme for Resampling Daily Precipitation and other Weather Variables, *Water Resources Research,* **35(10)**, 3089-3101.

Rajagopalan, B., U. Lall and S. Zebiak, 2002: Optimal Categorical Climate Forecasts through Multiple GCM Ensemble Combination and Regularization, *Monthly Weather Review,* **130**, 1792-1811.

Rajeevan, M., 2001: Prediction of Indian summer monsoon: Status, problems and prospects. Current Science, 81, 1451-1457

Rasmusson, E. M., and T. H. Carpenter, 1983: The relationship between the eastern Pacific sea surface temperature and rainfall over India and Sri Lanka, *Monthly Weather Review,* **111**, 354-384.

Sahai, A. K., A. M. Grimm, V. Satyan 2000: All India summer monsoon rainfall prediction using an artificial neural network, *Climate dynamics,* **16**, 291-302.

Saji, N. H., B. N. Goswami, P. N.Vinayachandran and T. Yamagata, 1999: A Dipole Mode in the Tropical Indian Ocean, *Nature*, **401**, 360-363.

Salas, J. D., 1985: Analysis and Modeling of Hydrologic Time Series, *Handbook of Hydrology,* New York: McGraw-Hill, 19.1-19.72.

Shabbar A. and G. A. Barnston, 1996: Skill of seasonal climate forecasts in Canada using cnonical correlation analyss, *Monthly Weather Review,* **124**, 2370-2385.

Shukla, J., and D. A. Mooley, 1987: Empirical prediction of the summer monsoon rainfall over India, *Monthly Weather Review,* **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.

Singhrattna, N., Balaji Rajagopalan, Krishna Kumar, K. and Clark, M., 2004: Interannual and Interdecadal Variability of Thailand Summer Monsoon, *Journal of Climate* (under revision)

Thapaliyal, V, 1981: ARIMA model for long-range predictio n of monsoon rainfall in Peninsula India, *India Meteorological Department Monograph Climatology 12/81*

Torrence, C., and P. Webster, 1999: Interdecadal changes in ENSO-Monsoon System, *Journal of Climate,* **12(8)**, 2679-2690.

Toth, Z., 2002: Assessing the Value of Probabilistic Forecasts from a Scientific Perspective, Validation of Probabilistic Forecasts, *Predictability Seminar, ECMWF*, Sept 9-13

Wang, B., R. Wu and T. Li, 2003: Atmosphere-Warm ocean interaction and its impacts on Asian-Australian Monsoon variation, *Journal of Climate,* **16**, 1195-1211

Webster, P., V. O. Magana, T. N. Palmer, J. Shukla, T. A. Tomas, M. Yanai and T. Yasnari, 1998: Monsoons: Processes, predictability, and the prospects for prediction, *Journal of Geophysical Research,* **103(C7),** 14451-14510.

Walker, G. T., 1924: Correlation in seasonal variations of weather, IV, A further study of world weather, *Mem. Indian Meteorological Department,* **24**, 275-332.

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Science: An Introduction. SanDiego: Academic Press.

Yates, D., S. Gangopadhyay, B. Rajagopalan and K. Strzepek, 2003: A Nearest Neighbor Bootstrap Technique for Generating Regional Climate Scenarios for Integrated Assessments, *Water Resources Research*, **39(7)**, 1199.

Yevjevich, V. M., 1972: Stochastic Processes in Hydrology, *Water Res. Publi.* Colorado: Fort Collins.

Figure Captions

Figure 1: 21-year moving window correlation between Thailand summer (Aug – Oct.) rainfall and selected predictors from pre-monsoon seasons (Jan-Mar NINO3; Mar – May IOD; Mar – May SAT; May – Jul SOI). The dashed horizontal lines are 95% significant levels.

Figure 2: Correlation maps of Thailand summer rainfall and pre-monsoon season (a) Sea Level Pressures and, (b) Sea Surface Temperatures. Shaded regions are significant at 95% confidence level.

Figure 3: Same as Figure 1 but with the identified predictors for the pre-monsoon seasons, (a) Mar – May, (b) Apr – Jun and (c) May – Jul.

Figure 4: Cross-validated skill scores for Thailand summer rainfall forecasts issued on the 1st of each month from April through August using the three ensemble forecast methods. The skill measures - correlation, LLH and RPSS are shown in the three plots.

Figure 5: Same as Figure 4 but for (a) wet years and (b) dry years. Correlation measure is not shown due to small sample size.

Figure 6: PDF of ensemble forecasts (solid line) and the climatological PDF (dotted line) for three selected wet years, 1983, 1988 and 1995 from the three methods (a) Modified K-NN, (b) LOCFIT and (c) Linear Regression

Figure 7: Same as Figure 6 but for selected dry years, 1984, 1987 and 1994.

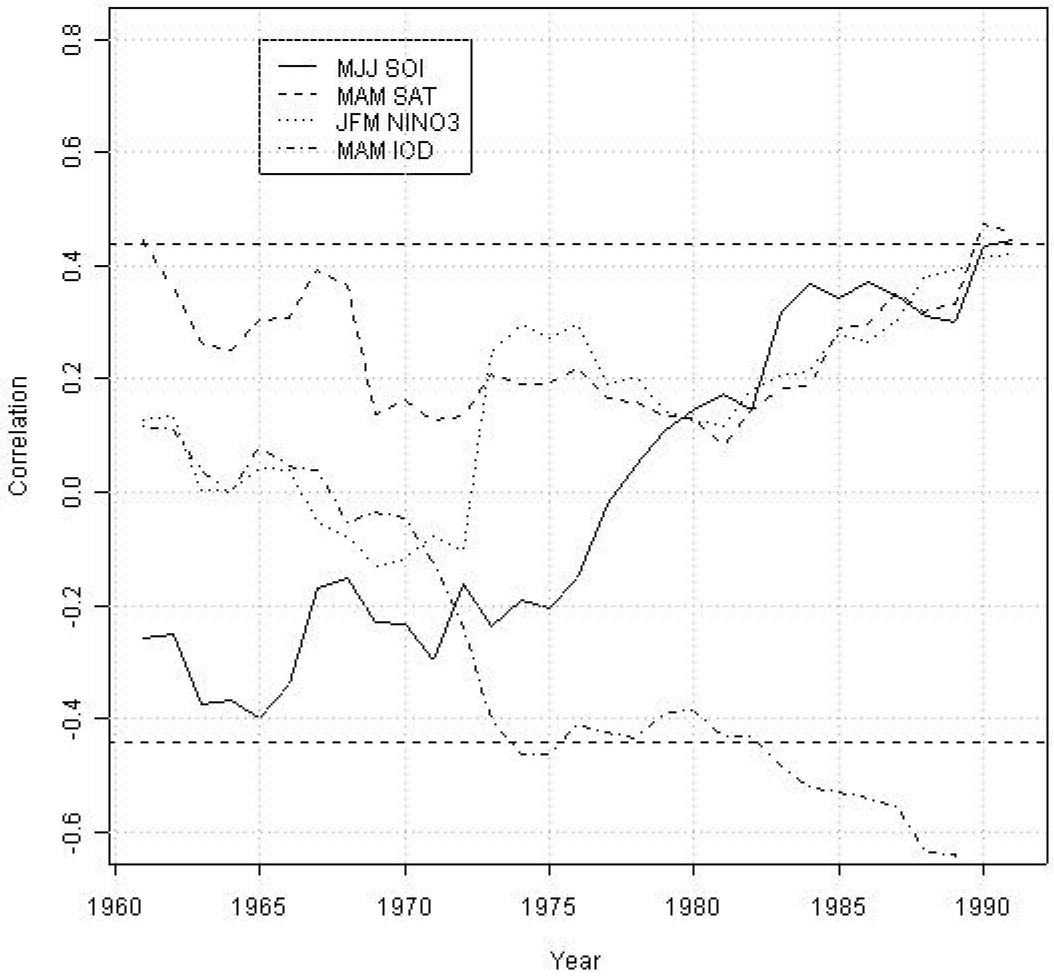

MJJ SOI
MAM SAT
JFM NINO3
MAM IOD
Correlation
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1960
1965
1970
1975
1980
1985
1990
Year

Figure 1

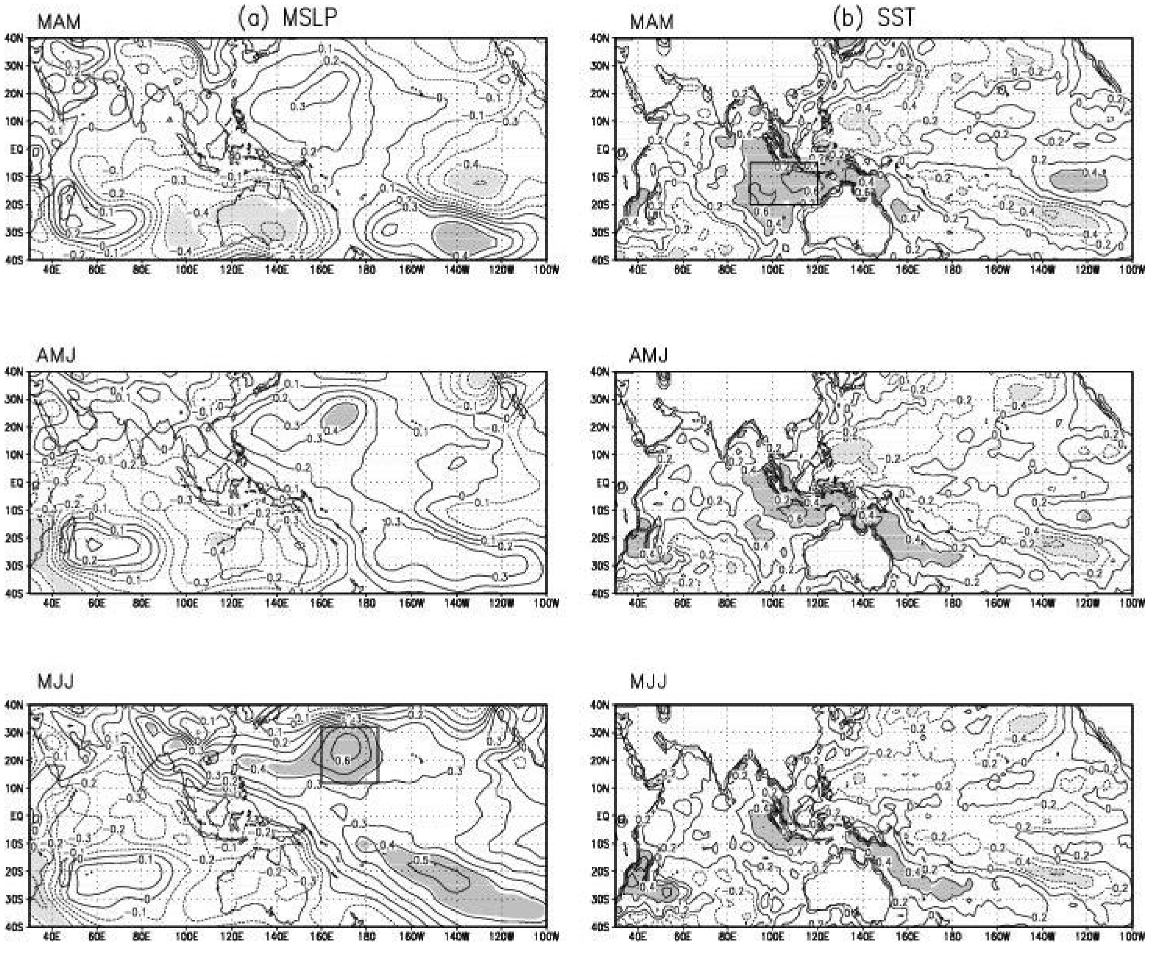

(a) MSLP
(b) SST
MAM
AMJ
MJJ
40N
30N
20N
10N
EQ
10S
20S
30S
40S
40E
60E
80E
100E
120E
140E
160E
180
160W
140W
120W
100W

Figure 2

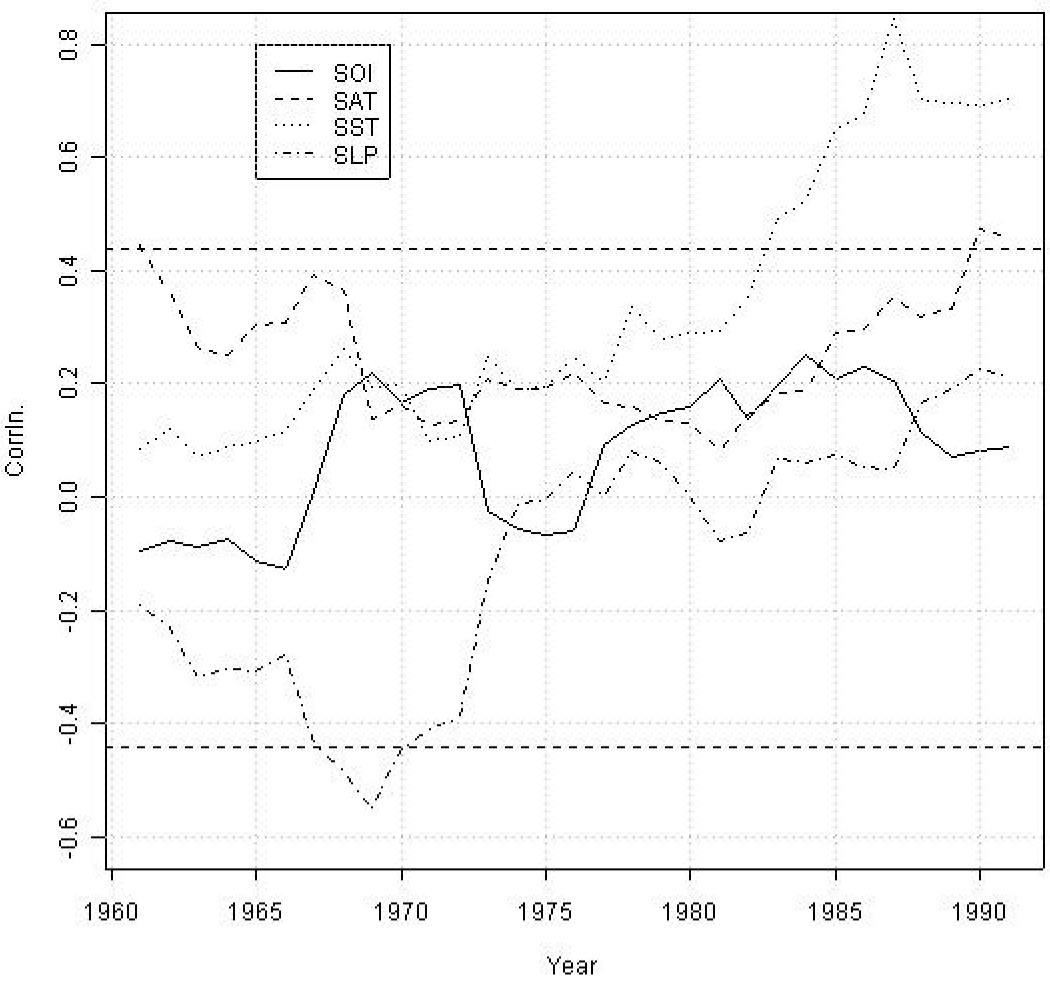

SOI
SAT
SST
SLP
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
Corrln.
1960
1965
1970
1975
1980
1985
1990
Year

Figure 3(a)

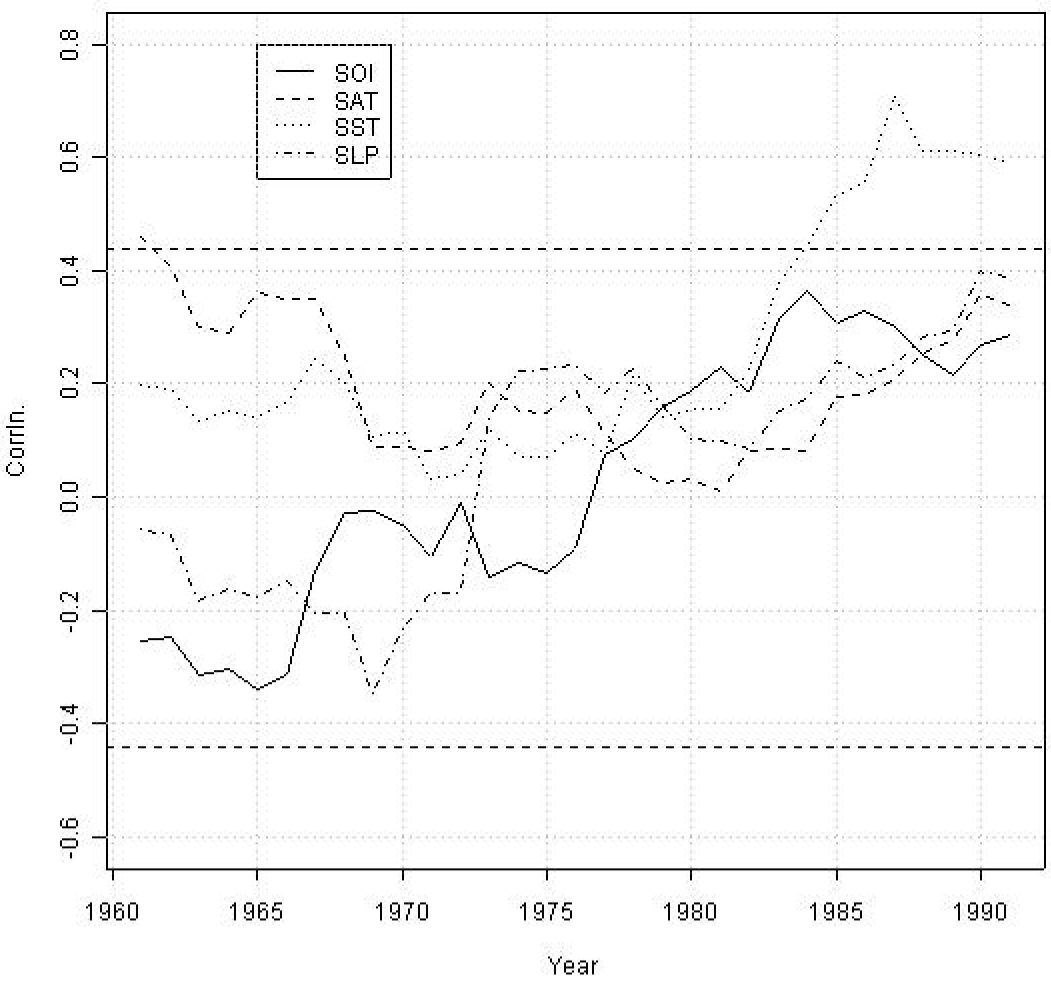
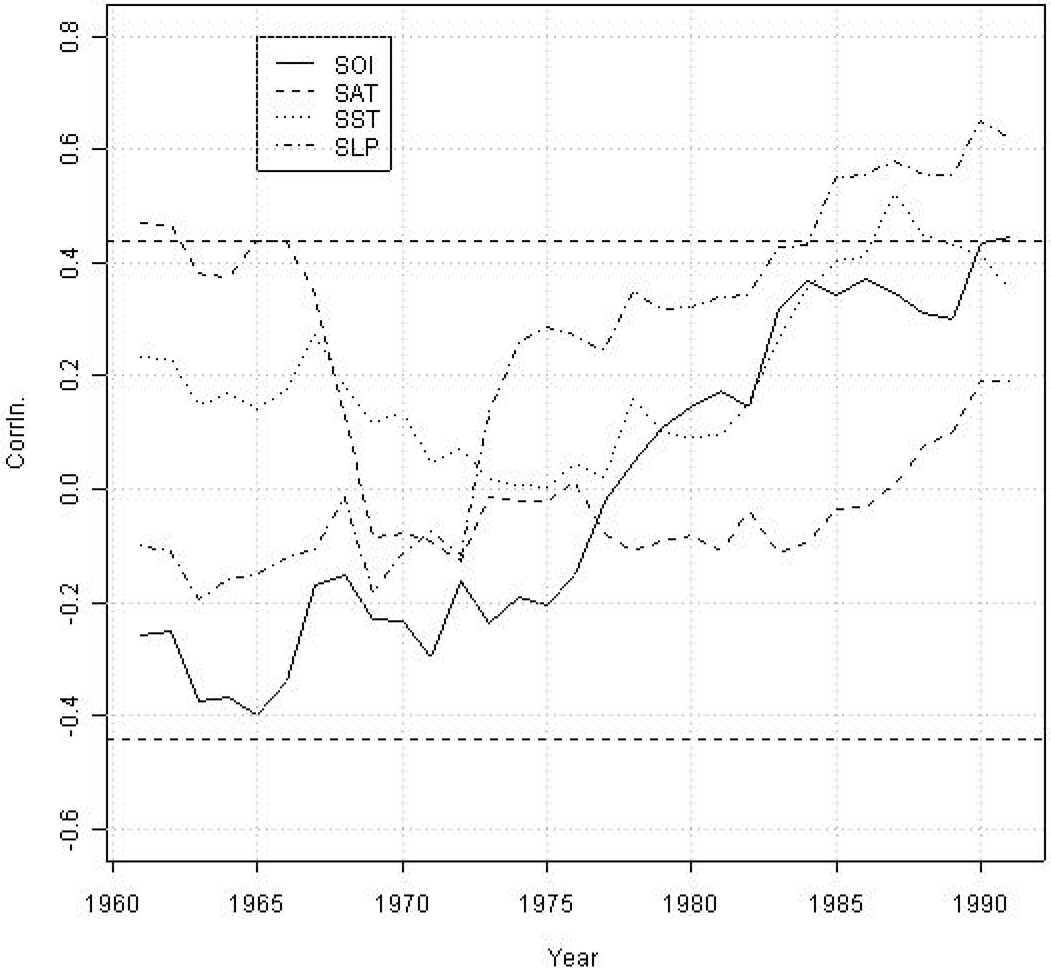

SOI
SAT
SST
SLP
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
Corrln.
1960
1965
1970
1975
1980
1985
1990
Year

Figure 3(b)



SOI
SAT
SST
SLP
Corrln.
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
1960
1965
1970
1975
1980
1985
1990
Year

Figure 3(c)

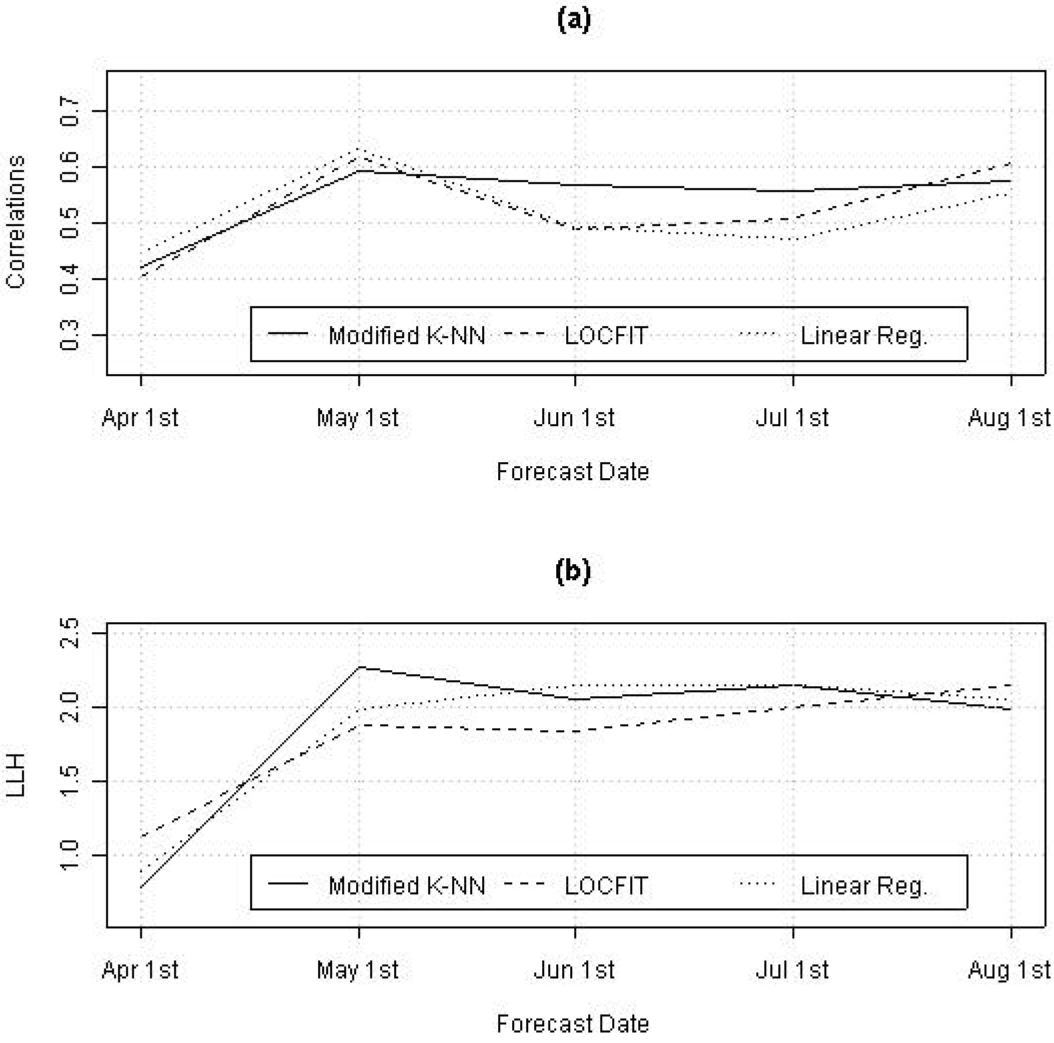

(a)
Correlations
0.3
0.4
0.5
0.6
0.7
Modified K-NN
LOCFIT
Linear Reg.
Apr 1st
May 1st
Jun 1st
Jul 1st
Aug 1st
Forecast Date
(b)
LLH
1.0
1.5
2.0
2.5
Modified K-NN
LOCFIT
Linear Reg.
Apr 1st
May 1st
Jun 1st
Jul 1st
Aug 1st
Forecast Date

Figure 4

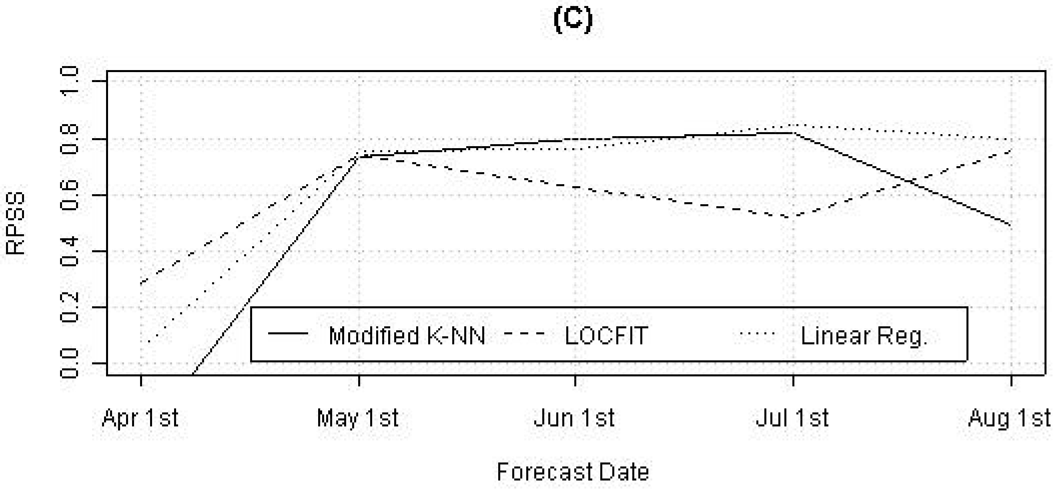

(C)
RPSS
1.0
0.8
0.6
0.4
0.2
0.0
Modified K-NN
LOCFIT
Linear Reg.
Apr 1st
May 1st
Jun 1st
Jul 1st
Aug 1st
Forecast Date

Figure 4

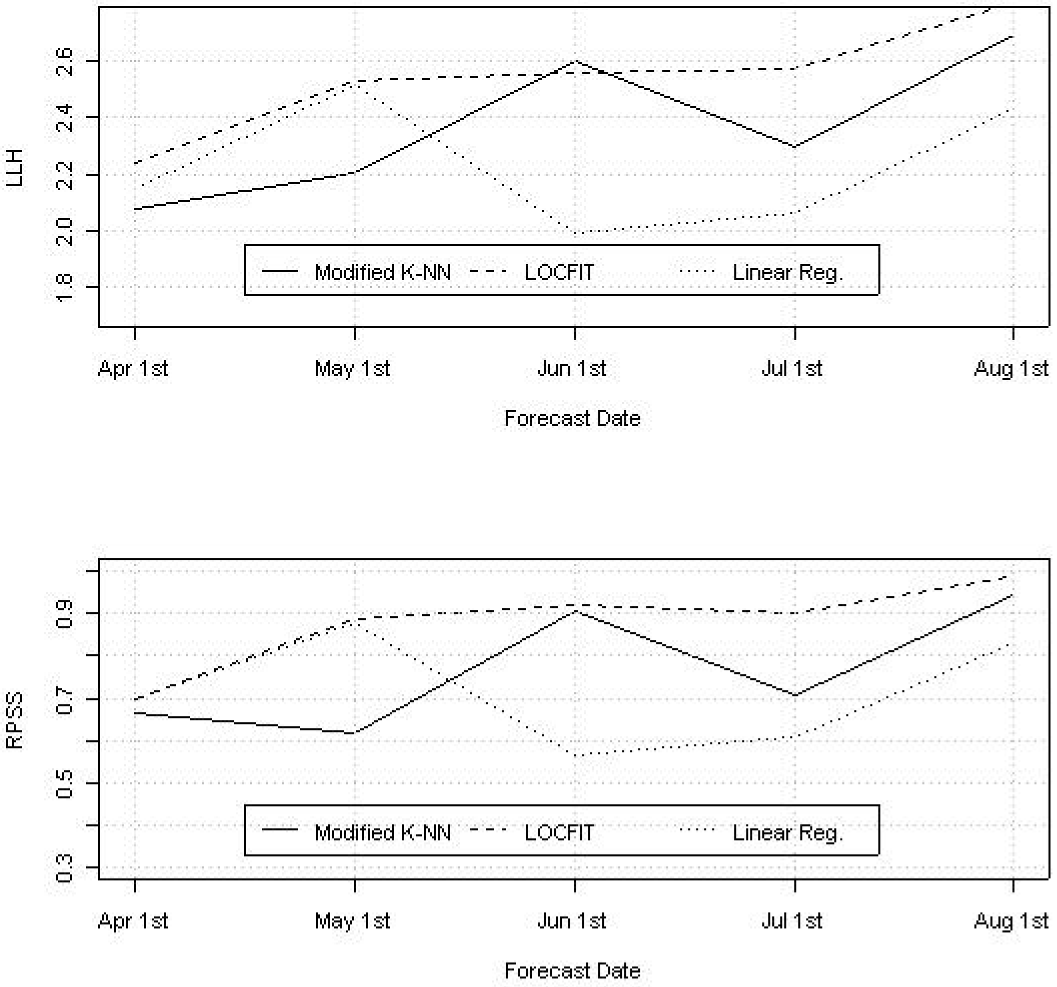

LLH
1.8
2.0
2.2
2.4
2.6
Modified K-NN
LOCFIT
Linear Reg.
Apr 1st
May 1st
Jun 1st
Jul 1st
Aug 1st
Forecast Date
RPSS
0.3
0.5
0.7
0.9
Modified K-NN
LOCFIT
Linear Reg.
Apr 1st
May 1st
Jun 1st
Jul 1st
Aug 1st
Forecast Date

Figure 5(a)

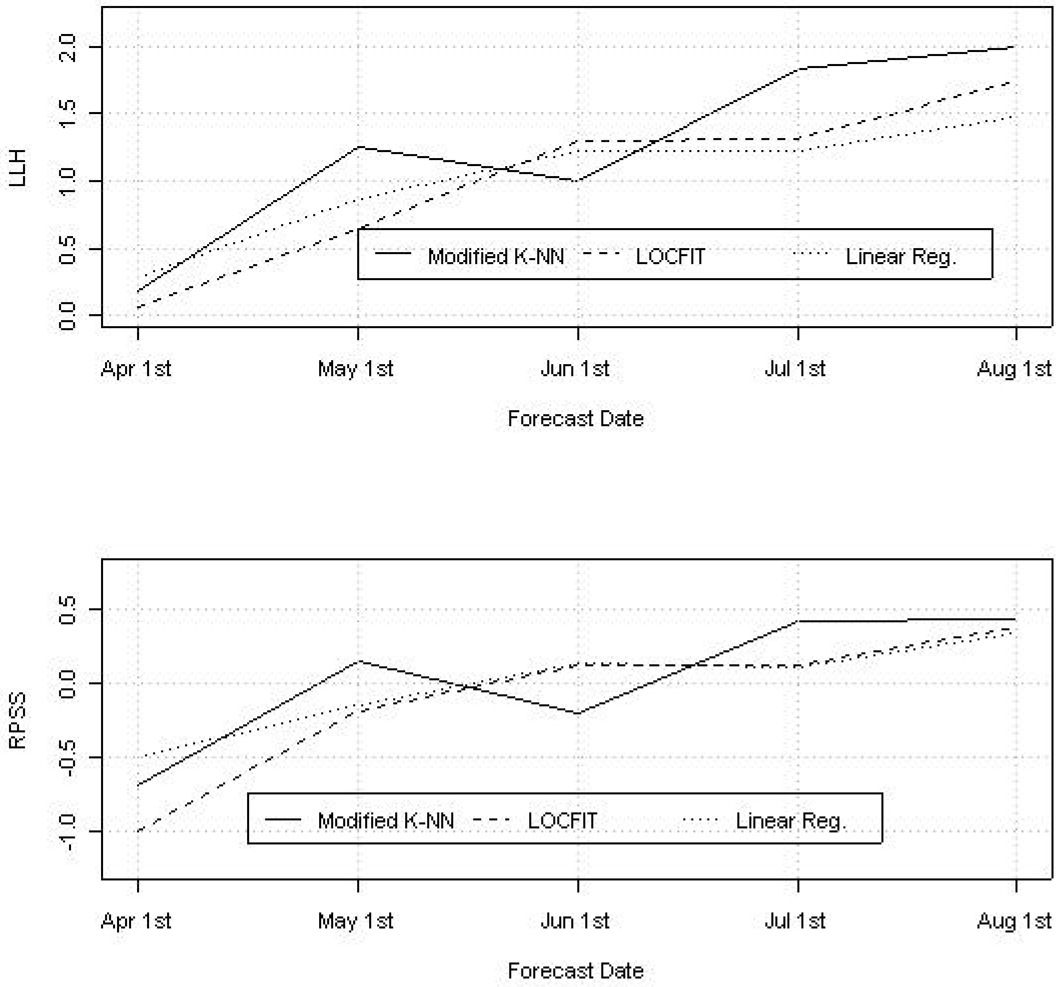

LLH
0.0
0.5
1.0
1.5
2.0
Modified K-NN
LOCFIT
Linear Reg.
Apr 1st
May 1st
Jun 1st
Jul 1st
Aug 1st
Forecast Date
RPSS
-1.0
-0.5
0.0
0.5
Modified K-NN
LOCFIT
Linear Reg.
Apr 1st
May 1st
Jun 1st
Jul 1st
Aug 1st
Forecast Date

Figure 5(b)

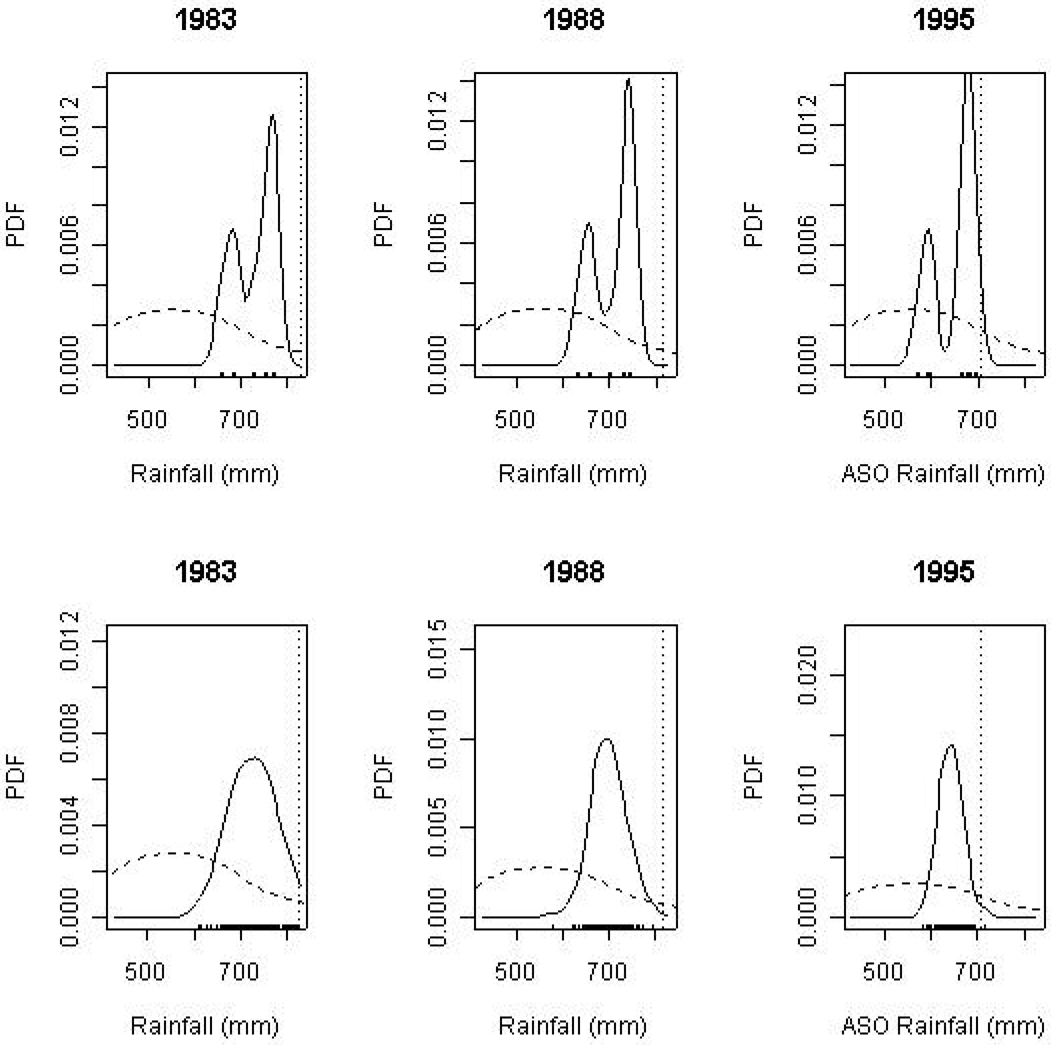

1983
1988
1995
PDF
0.000
0.006
0.012
500
700
Rainfall (mm)
Rainfall (mm)
ASO Rainfall (mm)
1983
1988
1995
0.000
0.004
0.008
0.012
0.000
0.005
0.010
0.015
0.000
0.010
0.020
Rainfall (mm)
Rainfall (mm)
ASO Rainfall (mm)

Figure 6

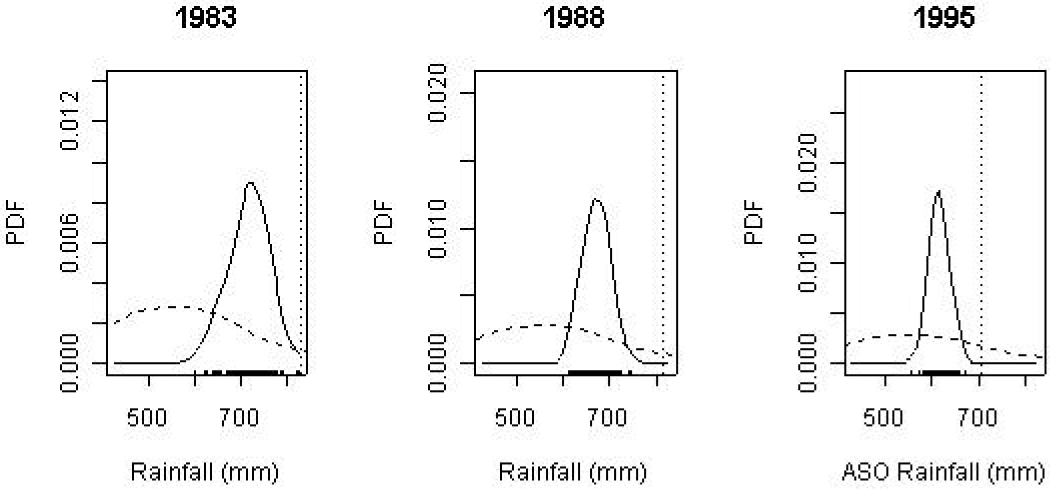

1983
PDF
0.000
0.006
0.012
500
700
Rainfall (mm)
1988
PDF
0.000
0.010
0.020
500
700
Rainfall (mm)
1995
PDF
0.000
0.010
0.020
500
700
ASO Rainfall (mm)

Figure 6

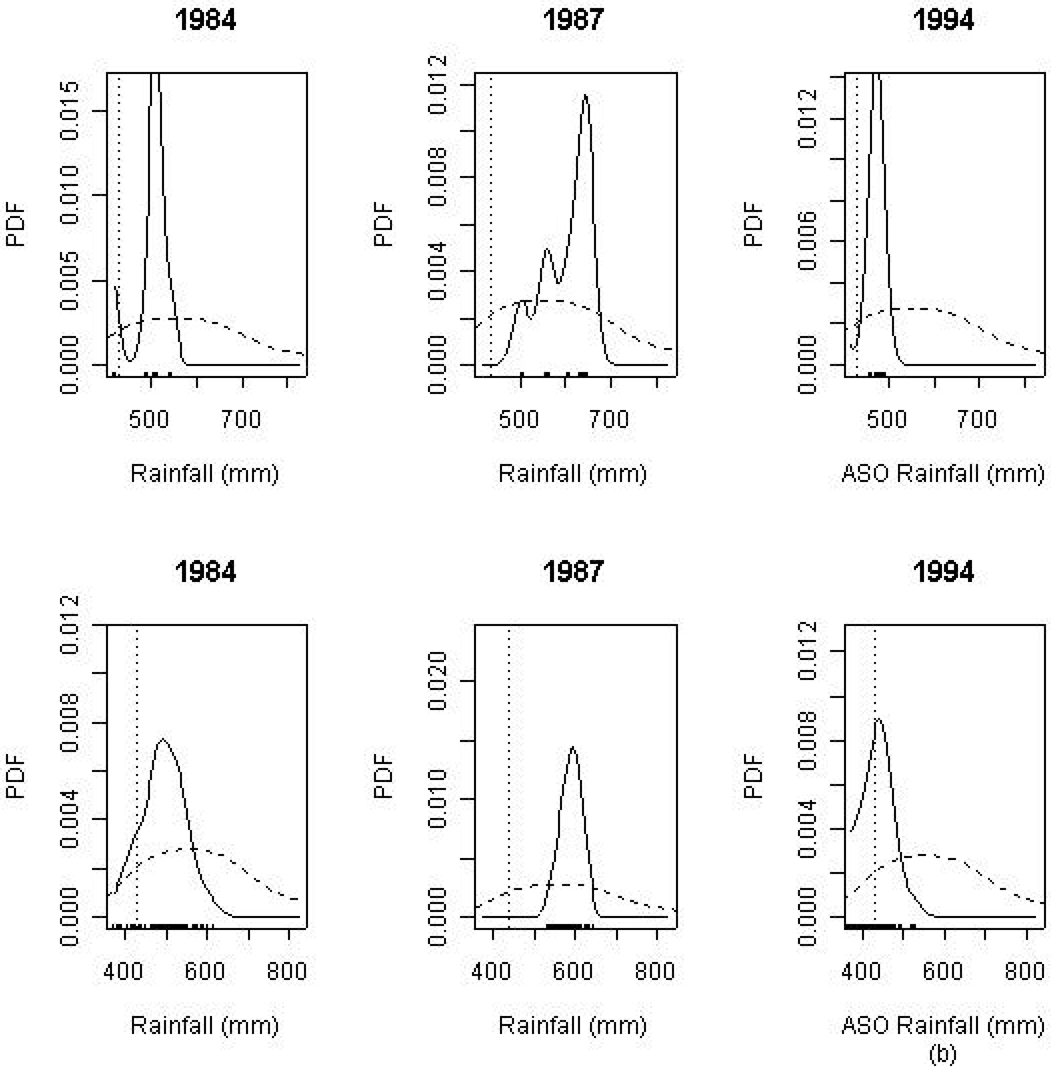

1984
PDF
0.000
0.005
0.010
0.015
500
700
Rainfall (mm)
1987
PDF
0.000
0.004
0.008
0.012
500
700
Rainfall (mm)
1994
PDF
0.000
0.006
0.012
500
700
ASO Rainfall (mm)
1984
PDF
0.000
0.004
0.008
0.012
400
600
800
Rainfall (mm)
1987
PDF
0.000
0.010
0.020
400
600
800
Rainfall (mm)
1994
PDF
0.000
0.004
0.008
0.012
400
600
800
ASO Rainfall (mm)
(b)

Figure 7

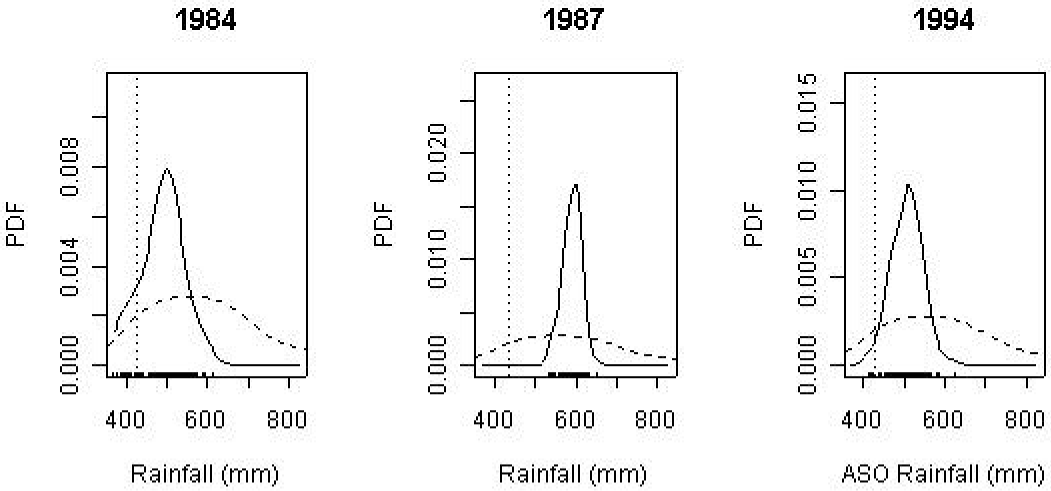

1984
1987
1994
PDF
0.000
0.004
0.008
0.000
0.010
0.020
0.000
0.005
0.010
0.015
400
600
800
Rainfall (mm)
Rainfall (mm)
ASO Rainfall (mm)

Figure 7