Identifying ENSO Influences on Rainfall with Classification Models: Implications for Water Resource Management of Sri Lanka

Thushara De Silva M.[1,3], George M. Hornberger[1,2,3]

[1]Department of Civil and Environmental Engineering, Vanderbilt University, Nashville, Tennessee, USA.
[2]Department of Earth and Environmental Science, Vanderbilt University, Nashville, Tennessee, USA.
[3]Vanderbilt Institute for Energy and Environment, Vanderbilt University, Nashville, Tennessee, USA.

Correspondence to: Thushara De Silva M. (thushara.k.de.silva@vanderbilt.edu, thushara.k.de.silva@ieee.org)

Abstract. Seasonal to annual forecasts of precipitation patterns are very important for water infrastructure management. In particular, such forecasts can be used to inform decisions about the operation of multipurpose reservoir systems in the face of changing climate conditions. Success in making useful forecasts often is achieved by considering climate teleconnections such as the El-Nino-Southern Oscillation (ENSO), Indian Ocean Dipole (IOD) as related to sea surface temperature variations. We present a statistical analysis to explore the utility of using rainfall relationships in Sri Lanka with ENSO and IOD to predict rainfall to Mahaweli and Kelani, river basins of the country. Forecasting of rainfall as classes; flood, drought and normal are helpful for the water resource management decision making. Results of these models give better accuracy than a prediction of absolute values. Quadratic discrimination analysis (QDA) and classification tree models are used to identify the patterns of rainfall classes with respect to ENSO and IOD indices. Ensemble modeling tool Random Forest is also used to predict the rainfall classes as drought and not drought with higher skill. These models can be used to forecast the areal rainfall using predicted climate indices. Results from these models are not very accurate; however, the patterns recognized provide useful input to water resources managers as they plan for adaptation of agriculture and energy sectors in response to climate variability.

1 Introduction

The spatial and temporal uncertainty of water availability is one of the major challenges in water resource management. Understanding patterns and identifying trends in seasonal to annual precipitation are very important for water infrastructure management. In particular, forecasts that incorporate such information can be used to inform decisions about the operation of multipurpose reservoir systems in the face of changing climate conditions.

Success in making useful forecasts often is achieved by considering climate teleconnections such as the El-Nino-Southern Oscillation (ENSO) as related to sea surface temperature variations and air pressure over the globe using empirical data (Amarasekera et.al., 1997; Denise et.al., 2017; Korecha and Sorteberg, 2013; Seibert et.al., 2017). Also, modes of variability of other tropical oceans can be related to regional precipitation (Dettinger and Diaz, 2000; Eden et al., 2015; Maity and Kumar, 2006; Malmgren et al., 2005; Ranatunge et al., 2003; Suppiah, 1996; Roplewski and Halpert, 1996). For example, the effect of the Indian Ocean Dipole (IOD) is identified as independent of the ENSO effect (Eden et al., 2015). Pacific decadal oscillation (PDO), Atlantic multi-decadal mode oscillation (AMO), ENSO, and IOD teleconnections to precipitation have been found by many studies over the globe. Variations of precipitation in the United States are explained by ENSO, PDO and AMO (Eden et al., 2015; National Oceanic and Atmospheric

Administration, 2017; Ward et.al., 2014), in African countries by ENSO, AMO and IOD (Reason et.al., 2006), and in South east Asian countries by ENSO: Indonesia (Lee, 2015; Nur'utami and Hidayat, 2016), Thailand (Singhrattna et.al., 2005), China (Cao et al., 2017; Ouyang et al., 2014; Qiu et.al., 2014). Australia (Bureau of Meteorology, 2012; Verdon and Franks, 2005), and central and south Asia (Gerlitz et al., 2016).

The impact of ENSO and IOD on the position of the intertropical convergence zone (ITCZ) has been identified as a primary factor driving south Asian tropical climate variations. South Asian countries get precipitation from two monsoons from the movements of ITCZ in boreal summer (2^0 N) and boreal winter (8^0 S). The South western monsoon (summer monsoon) is during June-August months and the North eastern monsoon (winter monsoon) is during December – February months (Schneider et.al, 2014). Climate teleconnections have been studied for summer monsoons (Singhrattna et. al., 2005; Surendran et.al., 2015) and winter monsoons (Zubair and Ropelewski, 2006), A negative correlation of ENSO with Indian summer monsoon has been identified (Jha et al., 2016; Surendran et al., 2015).

The objective of this study is to explore the climate teleconnection to dual monsoons and inter monsoons. Water resource management decisions typically are based on precipitation throughout the year and it is extremely important to explore the possibility that rainfall might be related to teleconnection indices for which seasonal forecasts are available. Sri Lanka is a South Asian country that gets rainfall from two monsoons and two inter-monsoons. We explore ENSO and IOD climate teleconnection to Sri Lanka precipitation throughout the year. Past studies have identified climate teleconnection linking precipitation to climate indices for several months and monsoon seasons, and shown the importance of these for forecasting rainfall in river basins (Chandimala and Zubair, 2007; Chandrasekara et al., 2003). We extend these analyses across monsoon and inter-monsoon seasons.

Although rainfall anomalies may be correlated strongly with teleconnection indices, the scatter in the data can be large, making predictions from regression models have high uncertainty. However, water managers may act on information about whether rainfall is expected to be abnormally low or high. Seasonal precipitation is generally forecasted in broad categories. For example, the US National Weather Service forecasts seasonal precipitation as above normal, below normal, and normal (National Oceanic and Atmospheric Administration, 2018). The International Research Institute for Climate and Society also forecasts seasonal precipitation as above, below and near normal (International Research Institute for Climate Society, 2018). We chose to follow a similar approach and investigate river basin rainfall teleconnections to climate indices with classification models. If reasonably accurate relationships can be developed, they will be useful for water resources management. For example, in Sri Lanka decisions about allocations of water for irrigation and hydropower could be improved with estimates of when low rainfall seasons are likely.

2 Hydrometeorology and climatology of the study area

Sri Lanka is an island in the Indian Ocean (latitude $5^o\,55'$ N - $9^o\,50'$ N, longitudes $79^o\,40'$ E – $81^o\,53'$ E). Mean annual rainfall varies from 880 mm to 5500 mm across the island. The rainfall distribution is determined by the monsoon system of the Indian Ocean interacting with the elevated land mass in the interior of the country. The country is divided into three climatic zones according to the rainfall distribution: humid zone (wet zone) (annual rainfall > 2500 mm), intermediate zone (2500 mm < rainfall < 1750 mm) and arid zone (dry zone) (rainfall < 1750 mm) (Department of Agriculture Sri Lanka, 2017).

Sri Lanka, a water-rich country, has 103 river basins varying from 9 km^2 to 10448 km^2. A large fraction of the water resources management infrastructure of the country is associated with the Mahaweli and Kelani river basins. The catchment areas of the Mahaweli and Kelani are 10448 km^2 and 2292 km^2 respectively. The two rivers start from the central highlands. Mahaweli, the longest river, travels to the ocean 331 km in the eastern direction and the Kelani 145 km in the western direction. Average annual discharge volume for the Mahaweli and Kelani basins are 26368 10^6 m^3 and 8660 10^6 m^3 respectively (Manchanayake and Madduma Bandara, 1999). The Kelani river basin is totally inside the humid zone whereas the Mahaweli river basin migrates through all three climate zones (Fig.1).

The temporal pattern of rainfall in Sri Lanka can be divided into four seasons as follows.

(1) Generally low precipitation across the country from the Northeast monsoon (NEM), which gets most precipitation during January to February. The arid zone of the country gets significant precipitation from the NEM, while humid zone gets very little rainfall during this period.

(2) The whole country gets precipitation from the first inter-monsoon (FIM) during March to April months. However, rainfall during this period is not very high across the country.

(3) The highest precipitation for the country is from the South western monsoon (SWM) during May to September. However, only the humid zone gets high precipitation during this season.

(4) The whole country gets precipitation from the second inter-monsoon (SIM) during October to December. Generally, precipitation from SIM is higher than FIM.

The time period of NEM and SIM are generally considered as December to February and October to November respectively (Department of Meteorology Sri Lanka, 2017; Malmgren et.al, 2003; Ranatunge et al., 2003). However, considering the bulk amount of water received from the monsoon, we consider January and February as the period of NEM and October to December as the period of SIM.

Reflecting the rainfall seasons, the country has two agriculture seasons “Yala” (April - September) and “Maha” (October - March). Because the arid zone gets minimal precipitation during the SWM, the agricultural systems (165,000 ha) developed under the Mahaweli multipurpose project depend on irrigation water during the Yala season. The country depends on stored water to drive hydropower year round. The Mahaweli and Kelani hydropower plants of 810 MW and 335 MW capacity serve as peaking and contingency reserve power to the power system (Ceylon Electricity Board, 2015). Management of reservoir systems is done to cater both to irrigation and hydropower requirements.

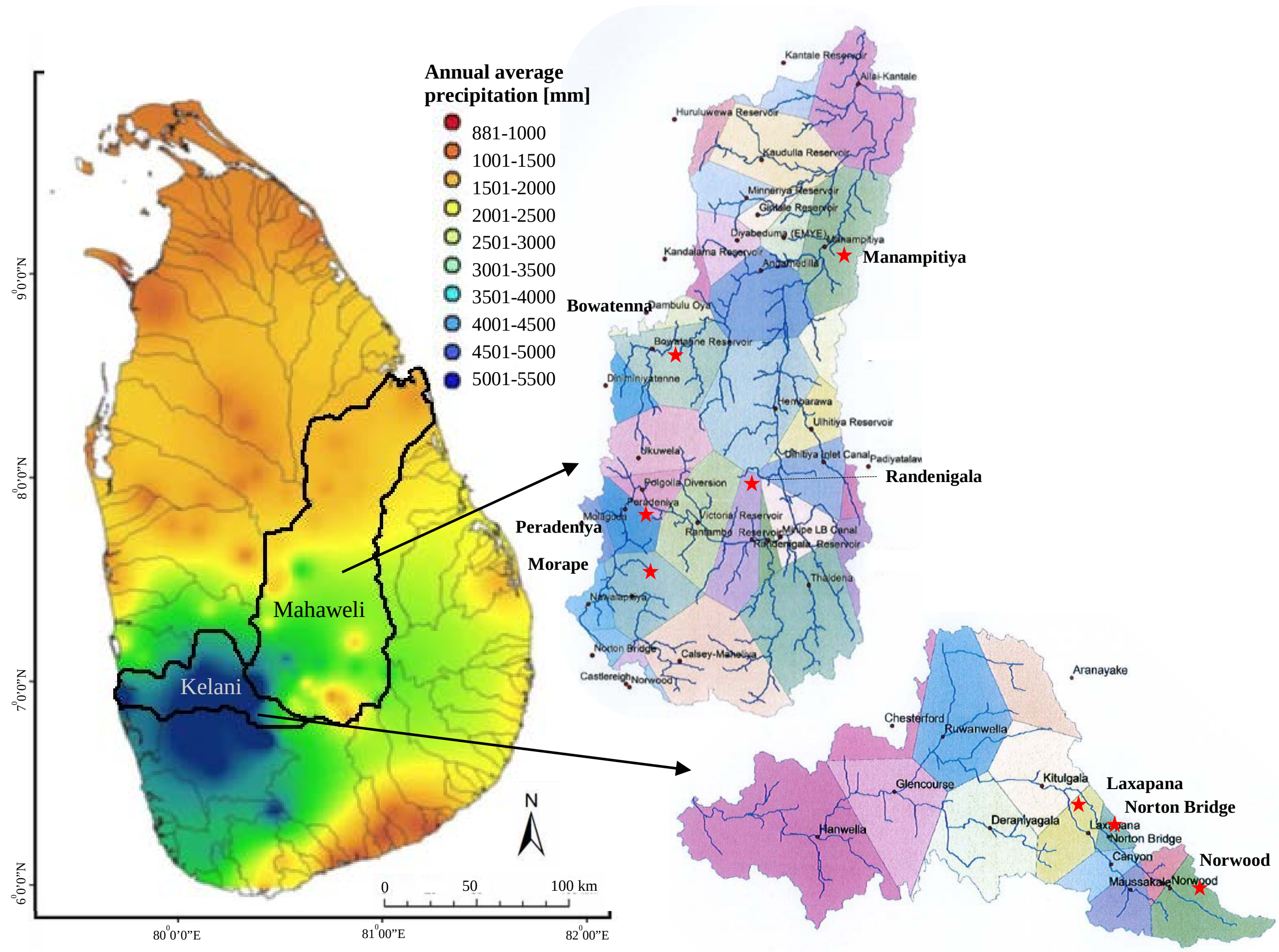


Figure 1 Mahaweli and Kelani river basins of Sri Lanka

2.1 Sub-basin rainfall (Areal rainfall)

Monthly rainfall data for years 1950 - 2013 are used for the study (Ceylon Electricity Board, 2017). River basin rainfall was calculated using the Thiessen polygon method (Viessman, 2002). The Mahaweli river basin is divided into 16 Thiessen polygons and the Kelani river basin is divided into 11 Thiessen polygons (Fig. 1). Since this study does not aim to explore rainfall across sub-basins, we do not use digital elevation maps to define the sub-basins. Considering the importance of sub-basins for the reservoir catchment and for water use, eight sub-basins are selected for analysis. Morape, Randenigala, Peradeniya, Manampitiya and Bowatenna represent the Mahaweli major reservoir catchments and irrigation tanks, and Norton Bridge, Norwood and Laxapana represent the Kelani basin reservoir catchments. The catchment of the major Mahaweli river reservoir cascade (Kotmale, Victoria, Randenigala, Rantambe, Bowatenna) is represented by Morape and Peradeniya located in the humid zone and by Randenigala and Bowatenna located in the intermediate zone. The arid zone major irrigation catchments of the Mahaweli are represented by Manampitiya. The catchment of the Kelaniya reservoir cascade (Norton Bridge and Moussakele) in the humid zone is represented by Laxapana, Norton Bridge and Norwood.

We calculate the rainfall for the four seasons, NEM, FIM, SWM and SIM for 64 years of historical data. Rainfall anomalies are calculated by reducing the seasonal mean rainfall (Eq.(1)) and standardized anomalies are calculated by dividing the rainfall anomalies by the standard deviation (SD) (Eq.(2)).

$$X_{ANM} = (X - \bar{X}_t) \tag{1}$$

$$X_{S\_ANM} = (X - \bar{X}_t)/SD_t \tag{2}$$

Where, $\bar{X}_t$ is the average of seasonal rainfall, X_{ANM} is the rainfall anomaly and X_{S_ANM} is the standardized rainfall anomaly.

Standardized rainfall anomalies are divided into three classes as dry, average and wet (Table 1). A normality test for the rainfall data classes is done using the Shapiro-Wilk test. If the rainfall data are not normally distributed, log (e), square root or square functions are used to transform the data into normally distributed data sets (Fig. A 1). Extreme seasonal precipitation has been defined statistically in different ways using statistical thresholds (Easterling et al., 2000; Jentsch et.al., 2015; Smith, 2011). We use 0.5 as a threshold to define three classes, which results in fairly evenly distributed data across the three classes (Fig. A 2).

Table 1 Rainfall anomaly classification

Class	Range
dry	$X_{S\_ANM} < -0.5$
average	-0.5 <= X_{S_ANM}<0.5
wet	0.5 <= X_{S_ANM}

2.2 ENSO & IOD indices

The ENSO phenomenon is represented by MEI, NINO34, NINO3, NINO4 indices, and the Indian Ocean dipole phenomenon is represented by DMI index. NINO34, NINO3, NINO4 indices are based on tropical sea surface temperature anomalies (National Center for Atmospheric Research, 2018) and the Multivariate ENSO Index (MEI) is based on sea-level pressure, zonal and meridional components of the surface wind, sea surface temperature, surface air temperature, and total cloudiness fraction of the sky (National Oceanic and Atmospheric Administration, 2017). The Indian Ocean Dipole (IOD) is an oscillation of sea surface temperature in the equatorial Indian ocean between Arabian sea and south of Indonesia (Bureau of Meteorology Australia, 2017). IOD is identified as relevant to the climate of Australia (Power et.al., 1999) and countries surrounded by the Indian ocean in southern Asia (Chaudhari et al., 2013; Maity and Nagesh Kumar, 2006; Qiu et al., 2014; Surendran et al., 2015). The Dipole Mode Index (DMI) is used to represent the IOD capturing the west and eastern equatorial sea surface temperature gradient.

Data used for the analyses are NINO34, NINO3, NINO4, MEI monthly data from years 1950 – 2013, (National Oceanic and Atmospheric Administration, 2017; National Center for Atmospheric Research, 2018), and the DMI monthly data from years 1950-2013 (HadISST dataset, Japan Agency for Marine-Earth Science and Technology 2017). Because we analyzed the data in rainfall seasons, values of the climate indices over the season are averaged. For example for the NEM season, the MEI value is the average of January and February monthly values and for the SWM season, DMI is the average of May, June, July and September values.

3 Methods

Seasonal values of MEI and DMI were used as the predictors to classify seasons into the three rainfall classes. The total data set is divided into 75 % for training the model and 25 % for testing model performance. Quadratic discriminant analysis (QDA) and classification trees were selected for the analyses. A random forest model also was applied to investigate the reliability of a cross-validated statistical forecast tool based on an advance estimate of MEI and DMI. We used the R programming language to carry out the statistical analyses. R packages: caret, tree, randomForest, fitdistriplus, devtools and quantreg are used for the studies.

3.1 Quadratic Discriminant Analysis (QDA)

The mathematical formulation of QDA can be derived from Bayes theorem assuming that observations from each class are drawn from a Gaussian distribution (James et.al., 2013; Löwe et.al., 2016).

The prior probability π_k represents the randomly chosen observation coming from kth class with density function $f_k(x)$. Bayes theorem states that

$$Pr(Y = k|X = x) = -\frac{\pi_k f_k(x)}{\sum_{l=1}^{K} \pi_l f_l(x)} \tag{3}$$

In Eq.(3), the posterior probability $Pr(Y = k|X = x)$ indicates that observation $X = x$ belongs to the kth class. For p predictors, the multivariate Gaussian distribution density function is defined for every class k (Eq.(4)).

$$f_k(x) = -\frac{1}{(2\pi)^{p/2}|\sum_k|^{1/2}}\, exp\left(-\frac{1}{2}(x - \mu_k)^T \textstyle\sum_k^{-1}(x - \mu_k\,)\right) \tag{4}$$

In Eq.(2), $\sum_k$ is the covariance matrix and μ_x is the mean vector. The covariance matrix ($\sum_k$) and mean (μ_x) for each class are estimated from the training data set (Eq.(5), Eq.(6)).

$$\mu_k = -\frac{1}{N_k} \sum_{i:y_i=k} x_i \tag{5}$$

$$\textstyle\sum_k = -\frac{1}{(N_k - 1)} \sum_{i:y_i=k} (x_i - \mu_k)^T (x_i - \mu_k) \tag{6}$$

Substituting a Gaussian density function for the k^{th} class (Eq.(4)) into Bayes theorem and taking the log values, the quadratic discriminant function is derived (Eq.(7)). Prior probabilities for class k (π_k) is calculated by the frequency of data points of class k in the training data (Eq.(8)). For a total number of N points in the training observations, N_k is the number of observations belong to kth class.

$$\delta_k(x) = -\frac{1}{2}\,(x - \mu_x)^T \textstyle\sum_k^{-1}(x - \mu_x\,)\; + log\,\pi_k \tag{7}$$

$$\pi_k = -\frac{N_k}{N} \tag{8}$$

Covariance, mean and prior probability values are inserted into the discriminant function ($\delta_k(x)$) together with the state variables (Eq.(5)). The corresponding class is selected according to the largest value of the function. The number of parameters to be estimated for the QDA model for k classes and p predictors is $k.p.(p + 1)\,/\,2$. For this study, the QDA model output is the probability that an observation of a climate category will fall into each of the rainfall classes.

3.2 Classification Tree model

For the classification tree model the predictor space is divided into non-overlapping regions ($R_1 .. R_j$). A classification tree predicts each observation as belonging to the most commonly occurring class of the training data regions (James et.al., 2013). Recursive binary splitting is used to grow the classification tree.

Classification error rate, Gini index and cross-entropy are typically used to evaluate the quality of particular split (James et.al., 2013), and in our study we used the first two indices. Classification error rate (E) gives fraction of observation that do not belong to the most commonly occurring class of the training data regions (Eq.(9)). However, for the tree-growing, the Gini index (G) is considered as the criterion for splitting into regions (Eq.(10))

$$E = 1 - max_k(\hat{p}_{mk}) \tag{9}$$

$$G = \sum_{k=1}^{K} \hat{p}_{mk}\,(1 - \hat{p}_{mk}) \tag{10}$$

In Eq.(9) and Eq.(10), $\hat{p}_{mk}$ represents the fraction of observations in the mth class that belong to the kth class. The Gini index is considered as a measure of node purity of the tree model, since small values of the index indicate that node has a higher number of observations from a single class.

The complexity of the trees are adjusted using a pruning process to produce more interpretable results. Complex trees reduces training error by overfitting the training data. Simple trees can be interpreted well, however, selecting a model which can find the pattern of data is important. In order to achieve the low classification error (training error + testing error), pruning technique is used. First, grow the very large tree, and sub tree is obtained by removing the weak links of the tree. Using tuning parameter to examine the trade-off between complexity of tree and the training error, and defining minimum samples for a node, maximum depth of the tree, and maximum number of terminal nodes are some of the pruning methods (Analytical Vidhya Team, 2016). For this study, we defined the maximum number of nodes to obtain the simple tree (pruned tree).

Tree models give the probability that an observation falls into each of the three rainfall classes. The predicted class is assigned based on the highest probability. Tree models handle ties of probability values by randomly assigning the class.

3.3 Random Forest

A random forest is an ensemble learning method used for classification and regression problems. The method is based on a multitude of decision trees based on training data with the final model as the mean of the ensemble (Breiman, 2001). Individual trees are built on a random sample of the training data with several predictors from the total number of predictors. Individual trees are built from the bootstrapped training data set.

There are some features, which can be tuned to make the better performed random forest model. Maximum number of predictors from the total predictors for individual trees, maximum number of trees, maximum node size of the trees and minimum sample leaf size are some of these features (Analytical Vidhya Team, 2015). In our study, we use the maximum number of trees as the main tuning parameters.

In a random forest model the importance of the variable is measured as the decrease in node impurity from the splits over the variable. This value is calculated by averaging the Gini index over the multitude of trees with a larger value indicating high importance of the predictor (James et.al., 2013).

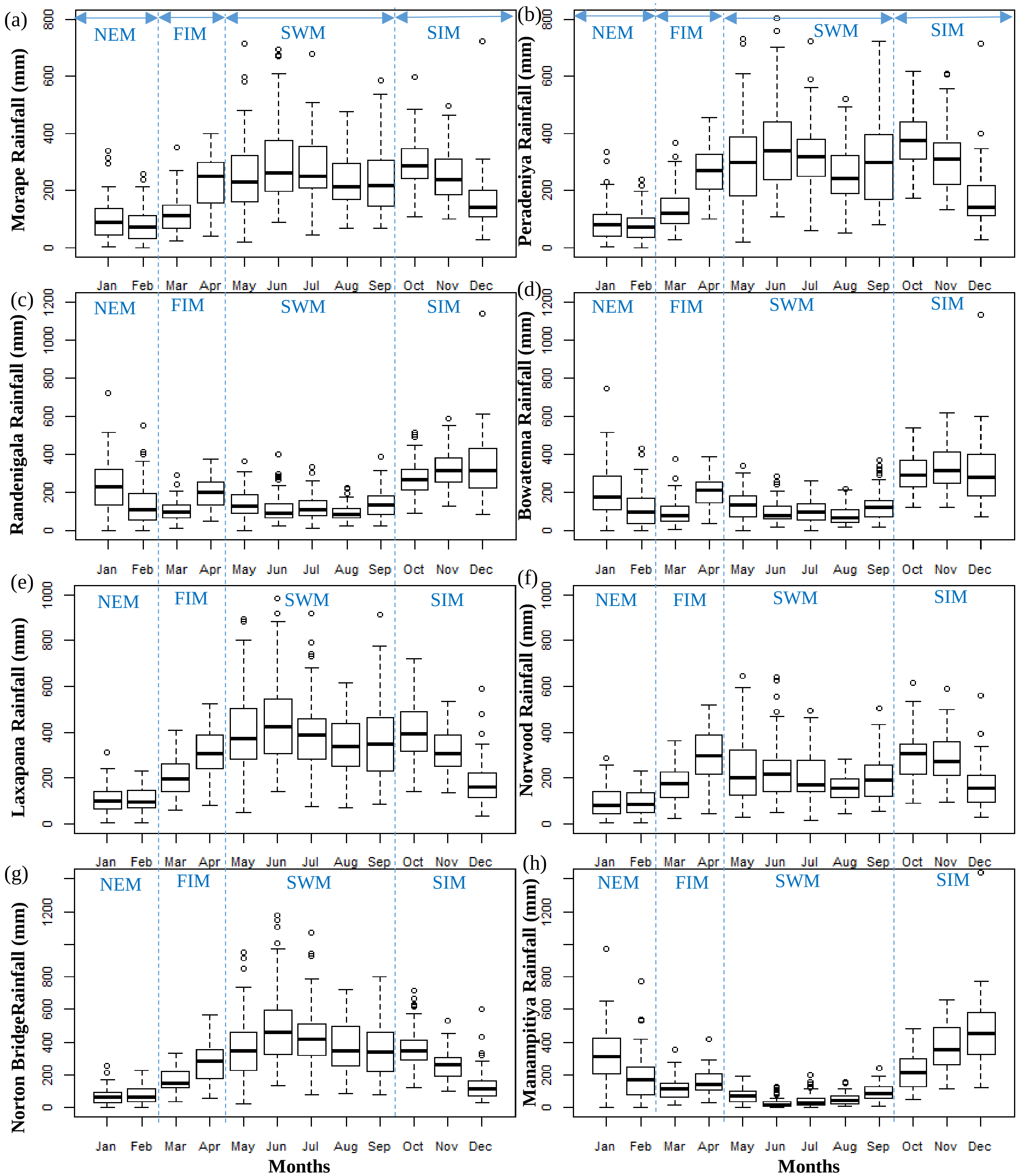


Figure 2 Sub basin Rainfall for (a) Morape, (b) Peradeniya,(c) Randenigala, (d) Bowatenna, (e) Laxapana (f) Norwood, (g) Norton Bridge, and (h) Manampitiya. Rainfall seasons are North East Monsoon (NEM), First Inter-Monsoon (FIM), South West Monsoon (SWM), and Second Inter-Monsoon (SIM)

4 Results

Monthly rainfall boxplots of eight sub basins over the year for 1950 - 2013 illustrate the seasonal and the spatial variation of rainfall patterns (Fig. 2). The largest fraction of total rainfall in the arid zone occurs at the end of the SIM (December) and during the NEM (January - February) with correspondingly high variability whereas there is little rainfall in the arid zone during the SWM (May - September) with correspondingly little variability (Fig. 2 (h)). The intermediate zone receives approximately 60% of total rainfall from the SIM and NEM. Although the variability of the rainfall is low in the intermediate zone, high rainfall can occur in all seasons (Fig. 2 (c) and (d)). In the humid zone, a large portion of rainfall occurs in SWM and early months of SIM (October-November). High variability of humid zone rainfall is observed at the end of FIM (April), in the SWM (May-September), and at the start of SIM (October) (Fig. 2 (a), (b), (e), (f) and (g)).

Similar to other investigators, we observe several strong correlations between rainfall anomalies and the climate indices (Table A. 1, Table A. 2, and Appendix). Higher correlation values between MEI and rainfall anomalies can be seen compared to the correlation with other ENSO indices (Table A. 1). In addition, rainfall in the SWM is very important for stations in the humid zone of the country which is the source of a large amount of water stored in reservoirs (Table A. 2). Correlation coefficients between SWM rainfall at Norton Bridge are negative and strong, -0.31 for MEI ($p = 0.01$) and -0.37 for DMI ($p < 0.01$). The strength of the correlation notwithstanding, the residuals from a regression model indicate that high uncertainty would attach to any forecast (Fig. 3). Thus, we are led to explore the efficacy of classification methods (Appendix).

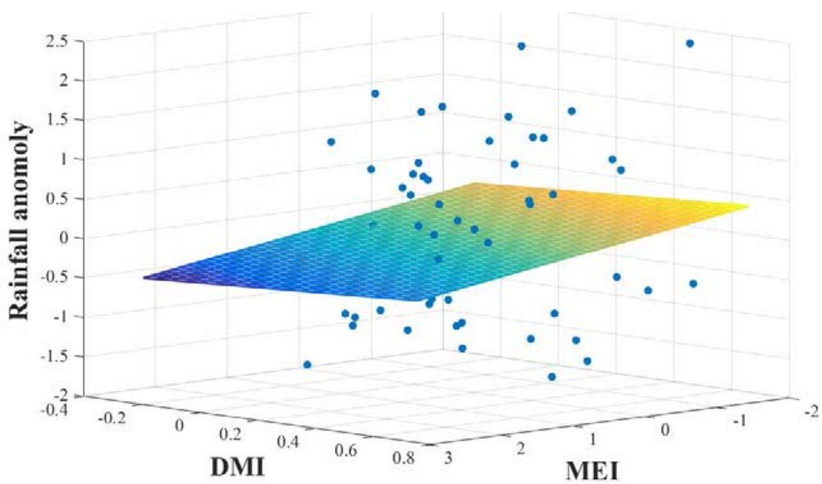


Figure 3 Linear regression of rainfall anomaly on MEI and DMI. High values of MEI and DMI are associated with low values of rainfall.

We present classification results for two sub-basins, one that has the highest rainfall during the NEM, Manampitiya, and one that has the highest rainfall for the SWM, Norton Bridge (Fig. 4). Norton Bridge represents the areal rainfall of reservoir catchments in the wet zone and Manampitiya represents the rainfall that contributes to irrigation tanks in the dry zone. Results of other sub-basins are presented in the supplementary materials (Fig. A 4, Fig. A 5, Fig. A 6, Fig. A 7, Appendix). Because MEI has higher correlation with rainfall anomalies than other ENSO indices, classification was done with only MEI and DMI.

The SWM is a season when the humid zone receives the bulk of rainfall. At Norton Bridge, the occurrences of the dry rainfall anomaly class in the SWM is seen to “clump” in the region of relatively high MEI and DMI. Both the classification tree and the QDA successfully identify the pattern (Fig. 4 (a) and (c)) with an overall accuracy of 73 %, 19 and 16 correct out of 22 occurrences (Table 2). In the arid zone the NEM season is one of the most important for rainfall. At Manampitiya, the MEI provides the primary variable in the classification, with the dry anomaly class being correctly selected in 52 % by tree model and 95 % with the QDA model. The results suggest that it may be possible to identify seasons when it is expected to be anomalously dry. The correct classification of “average” conditions likely has less importance for water managers. We explored classification using two classes, “Dry” and “Not Dry.” In this case, the classification model again correctly classifies 86 % of the anonymously dry cases and gets more than 69 % of the “Not Dry” cases correct (Fig. 5).

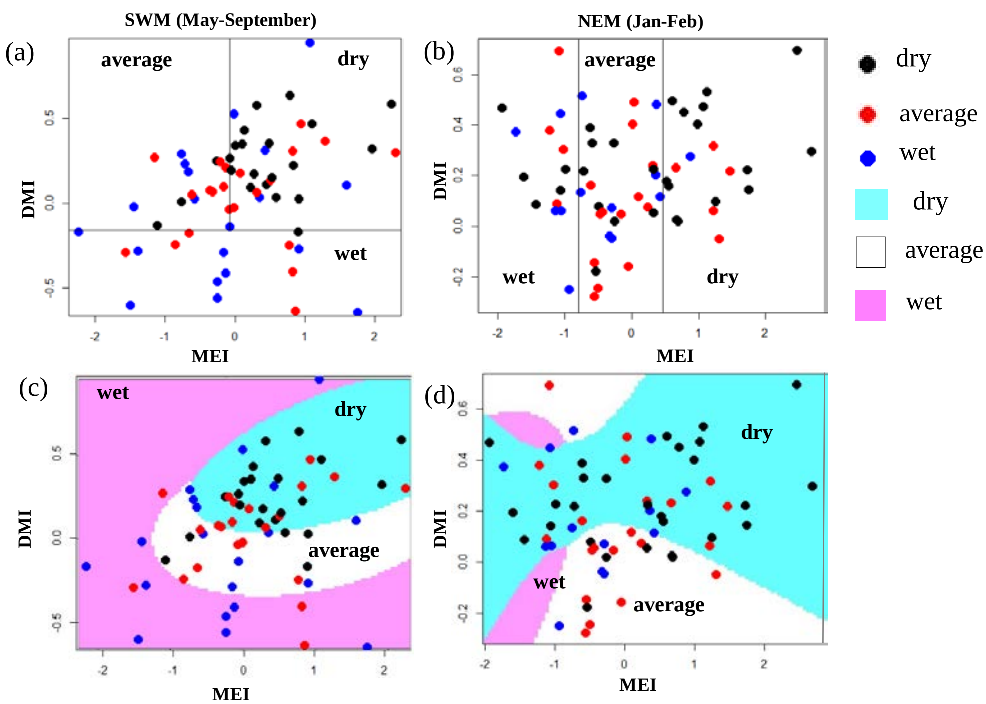


Figure 4 Norton Bridge and Manampitiya rainfall classes (dry, average, wet) identified by ENSO and IOD phenomena. (a) Norton Bridge SWM rainfall classification tree model (b) Manampitiya NEM rainfall classification tree model (c) Norton Bridge SWM rainfall QDA (d) Manampitiya NEM rainfall classification by QDA

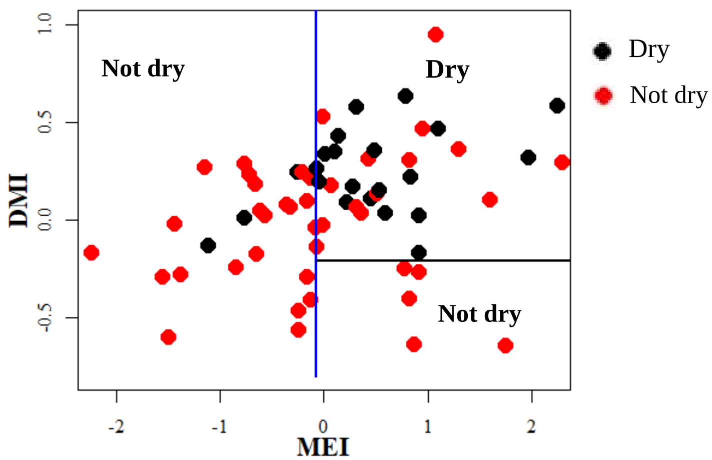


Figure 5 Classification tree for Norton Bridge SWM rainfall using two categories (dry and not dry)

Table 2 Classification model results. Highlighted cells indicate where there may be information content with respect to forecasting either dry or wet anomaly classes as judged by a classification success rate of at least 2/3.

Season	**Manampitiya**			**Norton Bridge**		
	QDA Model			*QDA Model*		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	22/23	11/25	1/16	5/20	25/29	2/15
FIM	9/21	20/24	5/19	3/20	14/23	14/20
SWM	2/21	30/27	2/16	16/22	9/22	9/20
SIM	17/25	13/20	7/19	7/22	15/22	11/20
Season	*Tree Model*			*Tree Model*		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	12/23	9/25	11/16	11/20	18/29	8/15
FIM	9/21	19/24	8/19	13/21	6/23	15/20
SWM	6/21	25/27	7/16	19/22	8/22	9/20
SIM	20/25	0/20	17/19	19/22	5/22	14/20

Classification trees are known to be unstable. That is, small changes in the observations can lead to large changes in the decision tree. The random forest approach overcomes the issue by building a "bag" of trees from bootstrap samples. The robustness of the model can then be checked by considering the "out-of-bag" error. The results of the random forest indicate that predictions of three rainfall anomaly classes using MEI and DMI is not feasible (Table 3). The out-of-bag error rate is close to two thirds, which for three categories is equivalent to a random selection.

Table 3 Results of random forest ensemble classification results

Season	**Norton Bridge**				**Manampitiya**			
	Dry	Normal	Wet	OOB Er	Dry	Normal	Wet	OOB Er
NEM	11/20	12/29	6/15	55%	14/23	10/25	5/16	55%
FIM	7/21	8/23	8/20	64%	10/21	11/24	6/19	58%

SWM	9/22	6/22	8/20	64%	6/21	17/27	5/16	56%
SIM	13/22	9/22	9/20	52%	15/25	8/20	7/19	53%

However, the results of the random forest for a classification as either “Dry” or “Not Dry” suggests that there may be skill in such a prediction. The out-of-bag error rates for this case range from 22 % to 38 % for Norton Bridge and Manampitiya (Table 3) and from 20 % to 39 % across all stations (Table A. 7).

Table 4 Results of random forest ensemble classification results for two rainfall anomaly classes

Season	Norton Bridge			Manampitiya		
	Dry	Not dry	OOB Error	Dry	Not dry	OOB Error
NEM	9/20	36/44	30 %	13/23	33/41	28 %
FIM	5/21	35/43	38 %	8/21	35/43	33 %
SWM	9/22	32/42	36 %	5/16	34/43	39 %
SIM	10/22	36/42	28 %	16/25	34/39	22 %

The QDA method produces results that are promising with respect to identification of extreme dry events as indicated by seasonal rainfall (Table 5).

Table 5 Classification results for extreme dry (very low rainfall) and wet (very high rainfall) seasons.

Class	Range	Norton Bridge SWM		Manampitiya NEM	
		tree	QDA	tree	QDA
Very dry	$X_{S\_ANM} < -1.0$	10/11	10/11	6/11	11/11
dry	$-1.0 <= X_{S\_ANM} < -0.5$	9/11	6/11	5/11	9/10
average	$-0.5 <= X_{S\_ANM} < 0.5$	8/22	9/22	9/25	11/25
wet	$0.5 <= X_{S\_ANM} <= 1.0$	5/11	5/11	1/5	0/5
Very wet	$1.0 <= X_{S\_ANM}$	6/11	6/11	7/11	1/11

5 Discussion

Understanding seasonal rainfall variability across the spatially diverse Mahaweli and Kelani river basins is important for irrigation and hydropower water planning. SWM and SIM are the key rainfall seasons for sub basins in the humid zone (Norton Bridge, Morape, Peradeniya and Laxapana), delivering 80 % of annual rainfall (Fig. 2 (a), (b), (e), (f)). For the arid zone (Manampitiya) and intermediate zone (Randenigala, Bowatenna) sub basins, the major season is SIM, which delivers more than 40 % of annual rainfall (Fig. 2 (c),(d),(h)). The arid zone also gets rainfall during the NEM (24 % of annual rainfall at Manampitiya) and the intermediate zone gets rainfall during the SWM (25 % - 30 % of annual rainfall at Randenigala and Bowatenna).

Climate teleconnection indices are related to rainfall anomalies observed during the two main growing seasons, Yala and Maha. The Maha agriculture season (October-March) depends on rain from SIM and NEM. During El Nino events rainfall increases for the first three months of the Maha season (SIM: October-December) (Fig. A 4, Fig. A 5, Fig. A 6, Fig. A 8) (Ropelewski and Halpert, 1995) and decreases during the last three months (NEM: January-March)(Fig. 4 (b)). In Yala season (April-September), La-Nina events enhance the rainfall during SWM (Fig. 4 (a), (c), Fig. A 4, Fig. A 5, Fig. A 6, Fig. A 8)(Whitaker et.al, 2001). During El Nino events the SWM rainfall is reduced (Fig. 4 (a), (c), Fig. A 8, Fig. A 9) (Chandrasekara et.al, 2017; Chandimala and Zubair, 2007; Zubair, 2003). The El Nino impact during the SWM is not as significant as it is during the NEM season (International Research Institute, 2017a). We find, however, that there is an interaction between two teleconnection indices, MEI and IOD for SWM rainfall. During the Yala season there is a high probability of having a drought when both the IOD and MEI are positive (Fig. 5). Also not having drought is probable when both the IOD and MEI are negative (Fig. 5, Fig. A 8, Fig. A 9).

Classification of wet, average, and dry rainfall anomalies using the MEI and DMI indices is successful. For example, a dry SWM season for Norton Bridge (Table 2) and other humid-zone stations (Table A. 4) is classified correctly with greater than 70 % accuracy with QDA and tree models. However, a random forest approach demonstrates that there is little skill in identifying a full wet-average-dry classification. However, a random forest model using only two rainfall categories shows more than 60 % accuracy in identifying “dry” and “not dry” classes of key rainfall seasons of the humid zone (Table 4, Table A. 7). Similarly, for arid zone locations such as Manampitiya, the dry rainfall class identification for NEM and SIM seasons is about 60 % (Table 4, Table A. 7).

Our statistical classification models can be combined with MEI and DMI forecasts to indicate the season-ahead expectation for rainfall. ENSO forecasts are available from the International Research Institute for Climate and Society (International Research Institute, 2017b) and IOD forecasts are available in the Bureau of Meteorology (BOM), Australian Government (Bureau of Meteorology, 2017). ENSO and IOD predictions are also associated with the uncertainty. Therefore, final forecast accuracy is a combination of the MEI, DMI forecast uncertainties and model’s accuracy rate in each class. Although overall prediction accuracy is not extremely high, a forecast of an anomalously low rainfall season can have value for risk-averse farmers (Cabrera et.al., 2007) and can guide plans for hydropower management (Block and Goddard, 2012).

The electricity and agriculture sectors of Sri Lanka heavily rely on Mahaweli and Kelani river water resources so season ahead forecasts of abnormally low rainfall should be useful for decisions on adaptation measures. For example, water availability of the first three months of a growing season is important for crop selection and the extent of land to be cultivated. Hydropower planning and scheduling of maintenance of the power plants also can benefit from season-ahead forecasts. The damage that can occur due to incorrect rainfall forecasts in the agriculture and energy sectors can be minimized with emergency planning during the season, which is the usual practice.

Although the accuracy of predicting low or not low seasonal rainfall is not very high, decisions based on forecasts that are improvements over climate averages should be an improvement over current practices. The accuracy of statistical models can be improved with longer records, which are important to train the classification models. Also, models can be fine-tuned for important shorter periods such as crop planting months and harvesting months for irrigation water planning.

6 Conclusion

ENSO and IOD phenomena teleconnections with river basin rainfall provide potentially useful information for water resource management. Relationships identified between teleconnection indices and river basin rainfall agree with other research findings. Prediction of seasonal rainfall classes from ENSO and IOD indices can inform water resources managers in reservoir operation planning for both hydropower and irrigation releases.

Code availability

Codes use for the analysis and generate the graphs can be found at:
https://github.com/thusharadesilva/Rainfall_Season_Classification.git

Data availability

Areal rainfall data can be found from Ceylon Electricity Board or Mahaweli Authority of Sri Lanka. It is required to get the data from either of these organization and it is not possible to make the data publically available. We added data of one station with the codes to test the codes, however, name is not mentioned to protect the privacy. Climate indices data are publically available and references are given.

DMI data: Japan Agency for Marine Earth Science and Technology.: SST/DMI data set and

MEI data: National Oceanic and Atmospheric Administration.: El Nino Southern oscilation

NINO3, NINO3.4, NINO4 data: National Center for Atmospheric Research.: NINO SST Indices (NINO 1+2, 3, 3.4, 4; ONI and TNI)

Appendix: Identifying ENSO Influences on Rainfall with Classification Models: Implications for Water Resource Management of Sri Lanka

Normality Testing:

The Shapiro-Wilk's method is used to identify the normality of rainfall anomaly distribution. The Manampitiya NEM normality test results are given below as an example.

Data 1: original data

W = 0.96675, p-value = 0.08185

Data 2: data transformed by square root

W = 0.98772, p-value = 0.7772

Data 3: data transformed by log

W = 0.91577, p-value = 0.0003325

Further, from data plots (Fig. A 1) and the S-W statistic, we conclude that the square root transformed data is closer to being normally distributed than the other forms.

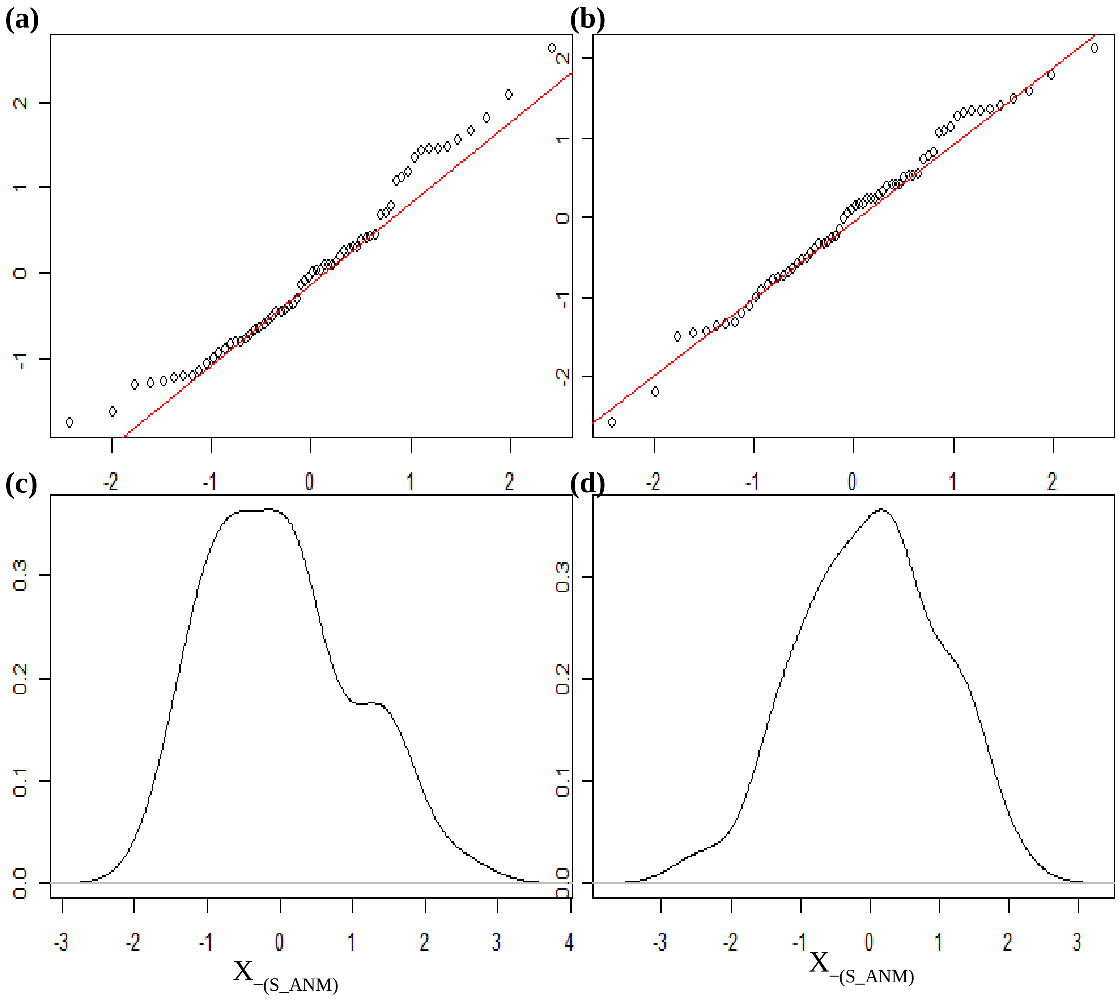


Figure A 1 Manampitiya NEM standardized data (a) original form qqplot (b) square root form qqplot (c) original form density plot (d) square root form density plot

Classification of data

Using 0.5 as a threshold for a normal distribution defines portions of the data that are fairly evenly distributed into three categories – about 31 %, 38 %, and 31 % for a normal distribution (Fig. A 2). We deemed this a reasonable choice for our analysis.

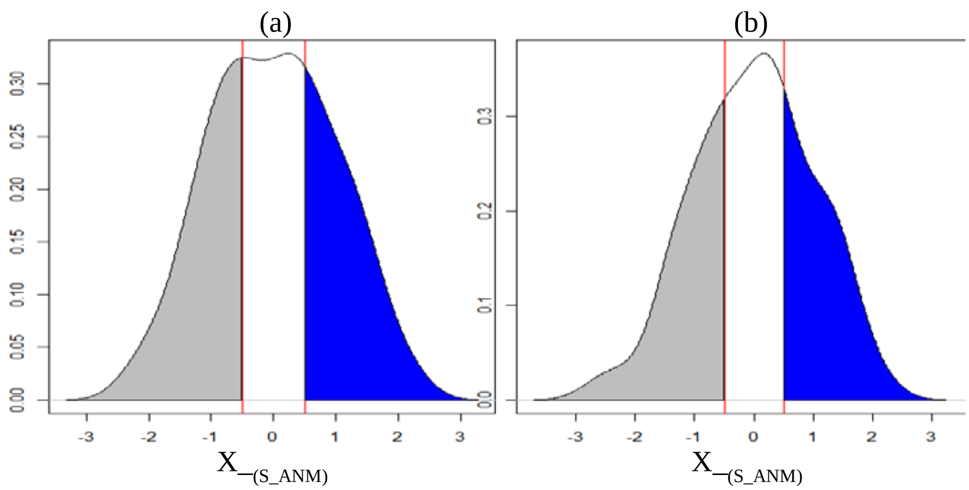


Figure A 2 (a) Norton Bridge SWM rainfall anomaly distribution (b) Manampitiya NEM rainfall anomaly distribution

Correlation analysis with multiple climate indices

We examined the correlation between rainfall anomalies and multiple climate indices to choose the two climate indices MEI and DMI (Fig. A 3, Table A. 1). The ENSO phenomenon is represented by MEI, NINO34, NINO3, NINO4 indices. Correlation analysis indicates that MEI, which is estimated using several climate factors such as sea-level pressure, zonal and meridional components of the surface wind, sea surface temperature, surface air temperature, and total cloudiness fraction of the sky (National Oceanic and Atmospheric Administration, 2017), demonstrates higher correlation with rainfall anomalies in sub-basins for all rainfall seasons compared to the NINO34, NINO3 and NINO4. The Indian Ocean dipole phenomenon is represented by the DMI index, which represents the gradient of the sea surface temperature. Based on the correlation analysis and the content of the indices, we selected MEI as the indicator for ENSO and DMI as the indicator for IOD.

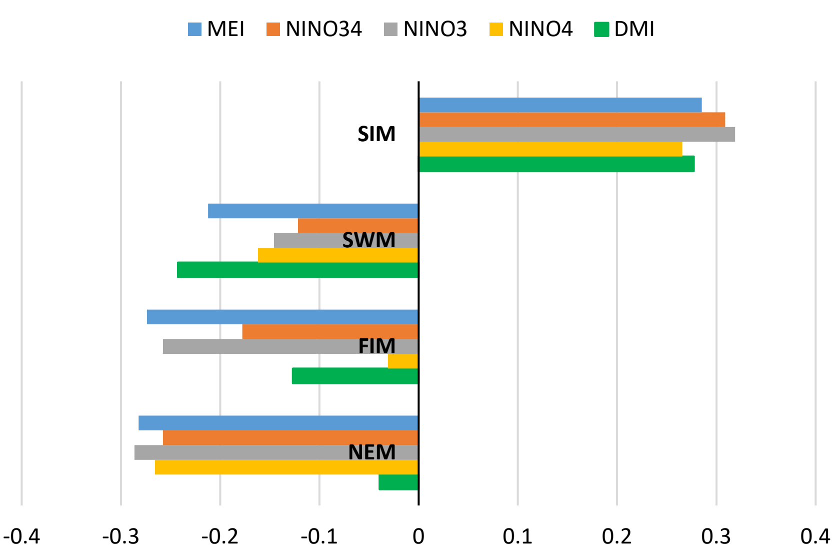


Figure A 3 Correlation between Norwood rainfall anomalies with multiple climate indices

Table A. 1 Correlation analysis of rainfall anomalies and climate indices

Rainfall			**Morape**					**Peradeniya**		
Month	MEI	NINO34	NINO3	NINO4	*DMI*	MEI	NINO34	NINO3	NINO4	*DMI*
NEM	-0.35	-0.35	-0.34	-0.38	*-0.09*	-0.38	-0.40	-0.39	-0.42	*-0.11*
FIM	-0.28	-0.19	-0.28	-0.07	*-0.11*	-0.27	-0.18	-0.30	-0.06	*-0.06*
SWM	-0.35	-0.24	-0.23	-0.26	*-0.29*	-0.35	-0.26	-0.25	-0.27	*-0.31*
SIM	0.21	0.23	0.27	0.19	*0.12*	0.17	0.19	0.21	0.15	*0.09*
Rainfall			**Laxapana**					**Norwood**		
Month	MEI	NINO34	NINO3	NINO4	*DMI*	MEI	NINO34	NINO3	NINO4	*DMI*
NEM	-0.27	-0.26	-0.28	-0.27	*-0.01*	-0.28	-0.26	-0.29	-0.27	*-0.04*
FIM	-0.28	-0.16	-0.27	-0.03	*-0.07*	-0.27	-0.18	-0.26	-0.03	*-0.13*
SWM	-0.3	-0.23	-0.21	-0.25	*-0.31*	-0.21	-0.12	-0.15	-0.16	*-0.24*
SIM	0.1	0.10	0.14	0.06	*0.08*	0.29	0.31	0.32	0.27	*0.28*
Rainfall			**Randenigala**					**Bowatenna**		
Month	MEI	NINO34	NINO3	NINO4	*DMI*	MEI	NINO34	NINO3	NINO4	*DMI*
NEM	-0.30	-0.31	-0.29	-0.34	*-0.11*	-0.35	-0.36	-0.35	-0.38	*-0.2*
FIM	-0.29	-0.23	-0.33	-0.10	*-0.04*	-0.23	-0.17	-0.25	-0.09	*-0.02*
SWM	-0.17	-0.12	-0.09	-0.18	*-0.24*	-0.18	-0.09	-0.05	-0.11	*-0.12*
SIM	0.37	0.38	0.41	0.36	*0.35*	0.35	0.41	0.40	0.40	*0.36*
Rainfall			**Norton Bridge**					**Manampitiya**		
Month	MEI	NINO34	NINO3	NINO4	*DMI*	MEI	NINO34	NINO3	NINO4	*DMI*
NEM	-0.32	-0.30	-0.33	-0.33	*-0.01*	-0.26	-0.28	-0.26	-0.28	*-0.16*
FIM	-0.18	-0.12	-0.21	-0.01	*-0.08*	-0.2	-0.17	-0.31	-0.06	*-0.14*
SWM	-0.31	-0.22	-0.21	-0.22	*-0.37*	-0.07	0.08	0.08	-0.01	*-0.03*
SIM	0.02	-0.02	0.03	-0.04	*-0.15*	0.45	0.46	0.44	0.46	*0.51*

Correlation analysis with MEI and DMI climate indices

Correlation coefficients between rainfall anomalies and MEI and DMI are negative for the NEM, FIM and SWM seasons and positive for the SIM season. Rainfall anomalies correlations to the DMI are not stronger as the correlations to the MEI. However, there are strong correlations for the anomalies of major monsoons to the sub basins and DMI values. For example, wet sub basins (Morape, Peradeniya, Laxapana, Norwood, Norton Bridge) have high correlation coefficient between SWM rainfall anomalies and DMI, while dry zone (Manampitiya) and intermediate zone (Randenigala, Bowatenna) sub basins have high correlation coefficient between NEM and SIM rainfall anomalies.

Table A. 2 Correlation between rainfall anomalies and MEI, DMI indices. High correlation coefficients are highlighted.

Rainfall	**Morape**		**Peradeniya**		**Randenigala**		**Bowatenna**	
Month	MEI	DMI	MEI	DMI	MEI	DMI	MEI	DMI
NEM	**-0.35**	-0.09	**-0.38**	-0.11	**-0.30**	-0.11	**-0.35**	**-0.20**

FIM	**-0.28**	-0.11	**-0.27**	-0.06	**-0.29**	-0.04	**-0.23**	-0.02
SWM	**-0.35**	**-0.29**	**-0.35**	**-0.31**	-0.17	**-0.24**	-0.18	-0.12
SIM	**0.21**	0.12	0.17	0.09	**0.37**	**0.35**	**0.35**	**0.36**
Rainfall	**Laxapana**		**Norwood**		**Norton Bridge**		**Manampitiya**	
Month	MEI	DMI	MEI	DMI	MEI	DMI	MEI	DMI
NEM	**-0.27**	-0.01	**-0.28**	-0.04	**-0.32**	-0.01	**-0.26**	**-0.16**
FIM	**-0.28**	-0.07	**-0.27**	-0.13	**-0.18**	-0.08	**-0.20**	-0.14
SWM	**-0.30**	**-0.31**	**-0.21**	**-0.24**	**-0.31**	**-0.37**	-0.07	-0.03
SIM	0.10	0.08	**0.29**	**0.28**	0.02	-0.15	**0.45**	**0.51**

Classification methods classification tree models, random forest and quadratic discriminant analysis identify the relationship between standardized rainfall anomaly classes (dry, average, wet) and MEI and DMI values (Fig. A 4, Fig. A 5, Fig. A 6, Fig. A 7). Positive values of MEI and DMI values resulted dry or average rainfall class for the NEM, FIM and SWM seasons. However, for SIM rainfall has wet or average class for the positive values of MEI and DMI. Accuracy of model result are high for the dominant monsoon rainfall seasons of each sub basin (Table A. 3, Table A. 4, Table A. 5). Ensemble model approach with random forest has given comparatively lower out-of-bag error rate for the dominant monsoons' rainfall anomaly classification (Table A. 5). For example, wet zone sub basins such as Norton Bridge, Norwood, Laxapana, Peradeniya and Morape random forest error rate is lower for the SWM and SIM seasons. Same as, dry and intermediate sub basins Manampitiya, Randenigala and Bowatenna NEM and SIM rainfall classes accuracy rate is high than other rainfall seasons. Also all three models have higher accuracy rate in identifying dry events and error rate of identifying wet and dry class also less 15 % (Table A. 3, Table A. 4, Table A. 5). Further analysis of two rainfall classes dry and not dry rainfall classes are identified relevant to the MEI and DMI values with classification tree and random forest methods (Fig. A 8, Fig. A 9). Classification tree models for two classes have higher accuracy rate as 65 % - 84 % for eight sub basins (Table A. 6). Random forest out-of-bag error for two classes models are vary between 20 % - 39 % and shows higher skill in identifying rainfall classes for major monsoons of the sub basins (Table A. 7). MEI shows higher variable importance of identifying the rainfall classes compare to the DMI values. Specially, for NEM and SIM which are important to the dry zone sub basins importance of MEI is high in the classification. However, some of the wet zone sub basins shows equal importance of DMI variable in identifying two rainfall classes in FIM and SWM (Fig. A 10).

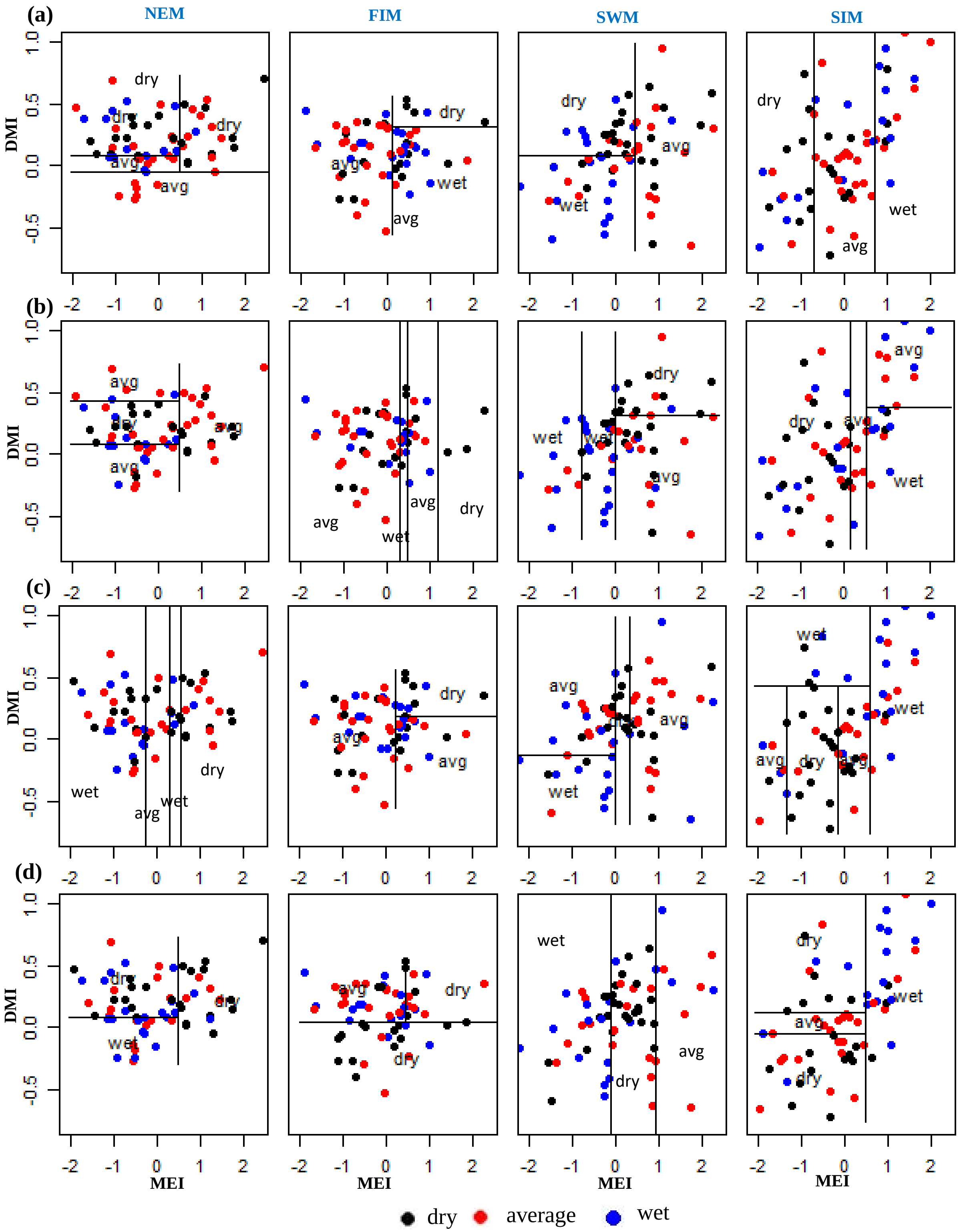


Figure A 4 Identifying relationships between three rainfall classes (dry, average, wet) and MEI and DMI values using classification tree models.(a)Morape (b)Peradeniya (c)Randenigala (d)Bowatenna

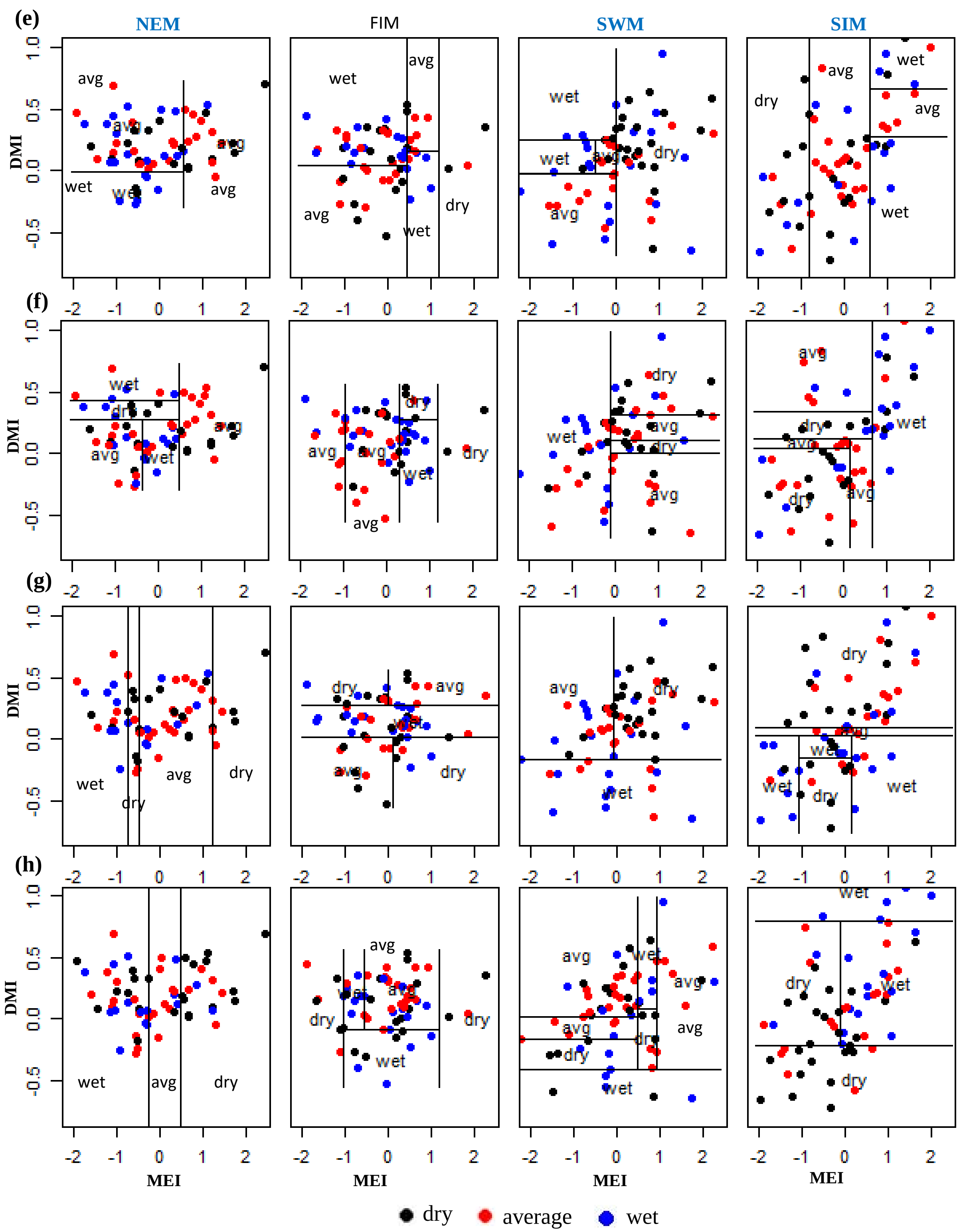
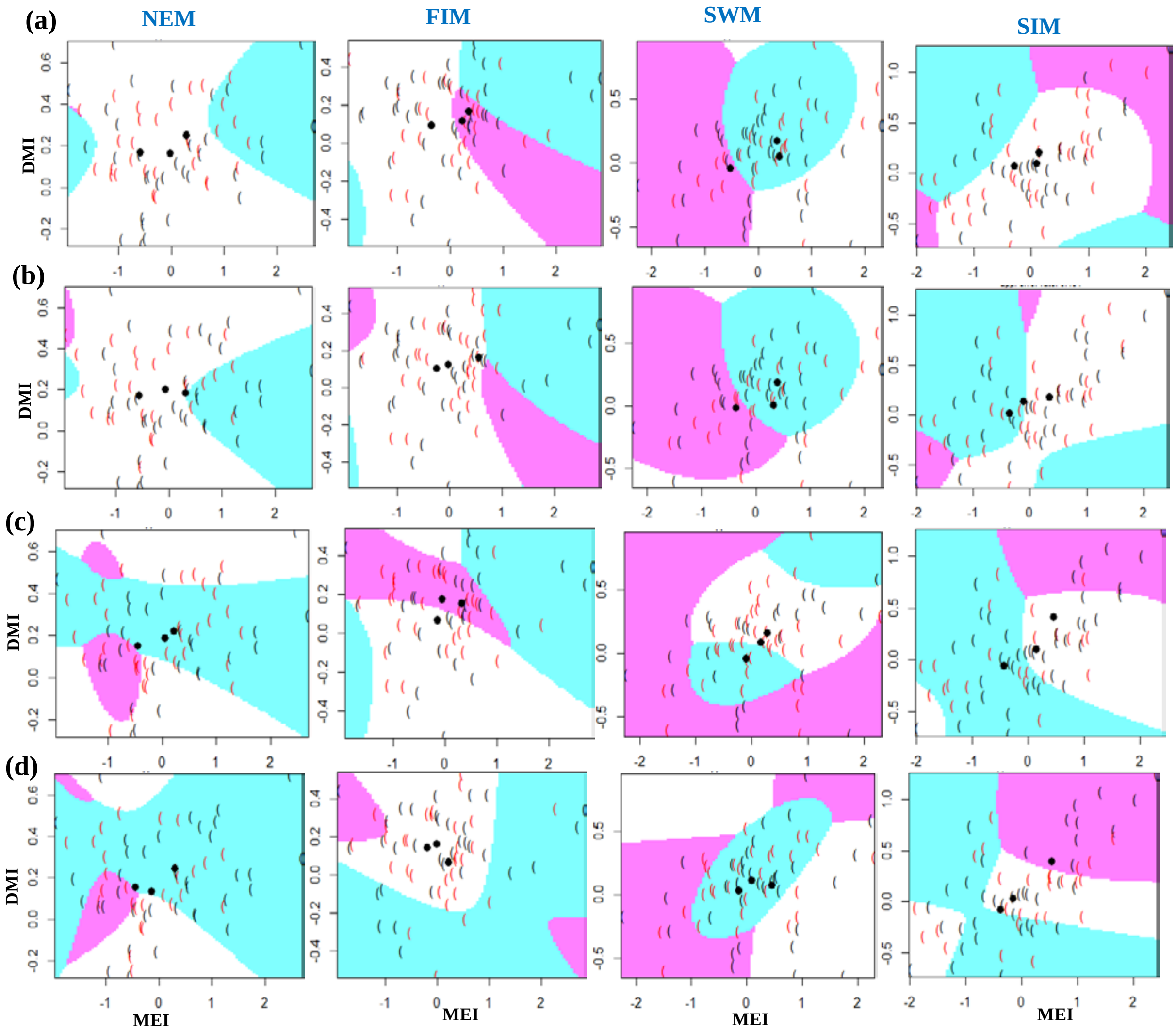


Figure A 5 Identifying relationships between three rainfall classes (dry, average, wet) and MEI and DMI values using classification tree models. (e)Laxapana (f)Norwood (g)Norton Bridge (h)Manampitiya

Table A. 3 Classification tree model results. Highlighted cells indicate where there may be information content with respect to forecasting either dry or wet anomaly classes

Season	**Morape**			**Peradeniya**		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	21/21	13/29	0/14	10/20	24/31	0/13
FIM	5/19	19/25	12/20	5/20	28/28	6/16
SWM	12/24	13/21	12/19	9/23	11/19	18/22
SIM	8/19	18/28	9/17	12/25	16/19	5/20
Season	**Randenigala**			**Bowatenna**		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	11/24	11/25	12/15	24/24	12/19	0/21
FIM	8/20	24/25	3/19	17/21	17/25	0/18
SWM	8/21	23/24	8/19	18/25	6/21	12/18
SIM	14/24	11/21	15/19	17/21	9/26	13/17
Season	**Laxapana**			**Norwood**		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	0/19	24/24	6/21	4/19	22/28	10/17
FIM	2/20	14/26	18/18	7/19	19/21	12/24
SWM	19/23	14/20	8/21	10/20	14/27	11/17
SIM	8/21	22/26	9/17	16/20	15/25	11/19
Season	**Norton Bridge**			**Manampitiya**		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	11/20	18/29	8/15	12/23	9/25	11/16
FIM	13/21	6/23	15/20	9/21	19/24	8/19
SWM	19/22	8/22	9/20	6/21	25/27	7/16
SIM	19/22	5/22	14/20	20/25	0/20	17/19



416 dry average wet (- correctly identified events (- incorrect

417 **Figure A 6** Identifying relationships between three rainfall classes (dry, average, wet) and MEI and DMI values using QDA models.(a) Morape (b) Peradeniya (c) Randenigala (d) Bowatenna

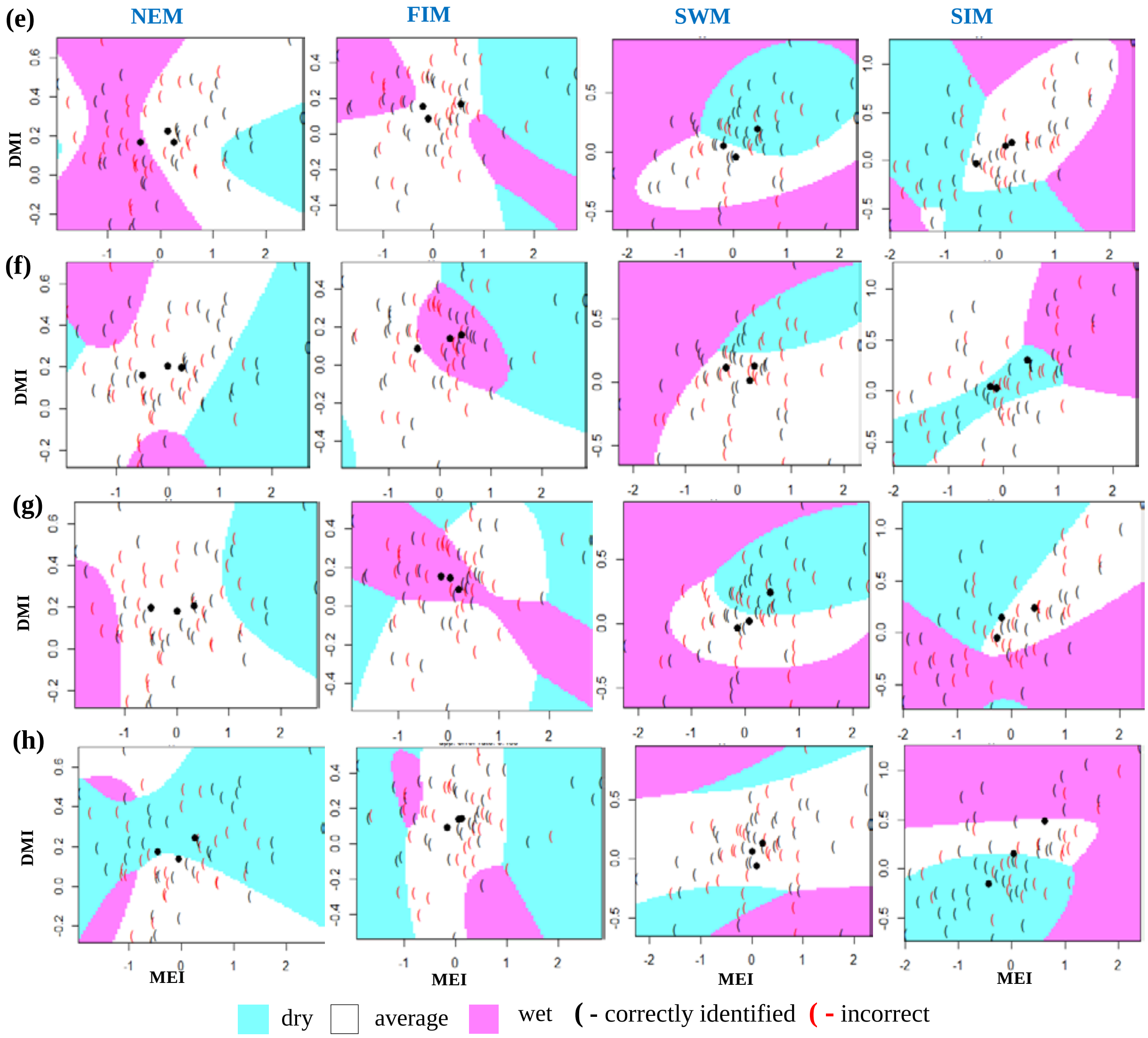


Figure A 7 Identifying relationships between three rainfall classes (dry, average, wet) and MEI and DMI values using classification tree models. (e) Laxapana (f) Norwood (g) Norton Bridge (h) Manampitiya

Table A. 4 Classification QDA model results. Highlighted cells indicate where there may be information content with respect to forecasting either dry or wet anomaly classes

Season	**Morape**			**Peradeniya**		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	6/21	28/29	0/14	10/20	28/31	0/13
FIM	7/19	22/25	9/20	5/20	28/28	2/16
SWM	19/24	6/21	13/19	20/23	6/19	13/22
SIM	5/19	26/28	2/17	13/25	16/19	4/20
Season	**Randenigala**			**Bowatenna**		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	17/24	8/25	4/15	24/24	9/19	3/21
FIM	8/20	13/25	12/19	9/21	23/25	1/18
SWM	4/21	13/24	8/19	19/25	7/21	8/18
SIM	19/24	16/21	6/19	13/21	15/26	10/17
Season	**Laxapana**			**Norwood**		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	4/19	15/24	14/21	8/19	23/28	6/17
FIM	4/20	22/26	8/18	6/19	16/21	13/24
SWM	20/23	13/20	10/21	6/20	19/27	8/17
SIM	9/21	22/26	3/17	11/20	13/25	8/19
Season	**Norton Bridge**			**Manampitiya**		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	5/20	25/29	2/15	22/23	11/25	1/16
FIM	3/20	14/23	14/20	9/21	20/24	5/19
SWM	16/22	9/22	9/20	2/21	26/27	6/16
SIM	7/22	15/22	11/20	17/25	13/20	7/19

Table A. 5 Random forest model results. Highlighted cells indicate where there may be information content with respect to forecasting either dry or wet anomaly classes

Season	Morape			Peradeniya		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	12/21	12/29	5/14	9/20	17/31	5/13
FIM	8/19	14/25	10/20	7/20	17/28	6/16
SWM	11/24	6/21	11/19	11/23	1/19	13/22
SIM	8/19	16/28	2/17	5/25	9/19	6/20

Season	Randenigala			Bowatenna		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	10/24	8/25	4/15	16/24	6/19	11/21
FIM	9/20	8/25	8/19	16/21	14/25	4/18
SWM	9/21	14/24	6/19	14/25	7/21	5/18
SIM	15/24	6/21	7/19	3/21	14/26	11/17

Season	Laxapana			Norwood		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	3/19	11/24	9/21	9/19	16/28	8/17
FIM	1/20	18/26	1/18	8/19	10/21	12/24
SWM	19/23	9/20	4/21	6/20	15/27	4/17
SIM	10/21	12/26	3/17	8/20	14/25	8/19

Season	Norton Bridge			Manampitiya		
	Dry	Normal	Wet	Dry	Normal	Wet
NEM	11/20	12/29	6/15	14/23	10/25	5/16
FIM	7/21	8/23	8/20	10/21	11/24	6/19
SWM	9/22	6/22	8/20	6/21	17/27	5/16
SIM	13/22	9/22	9/20	15/25	8/20	7/19

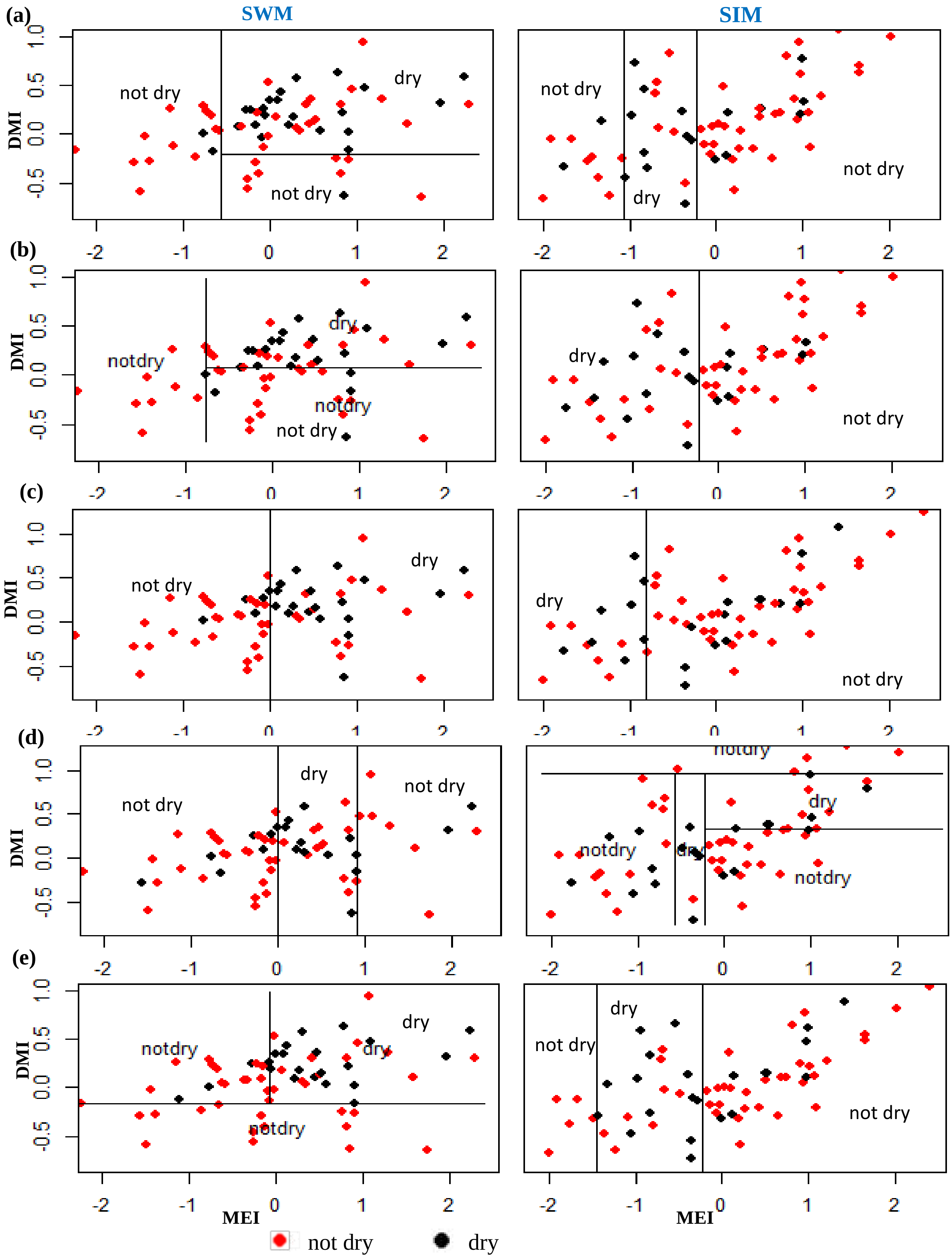


Figure A 8 Identifying relationships between two rainfall classes (dry, not dry) and MEI and DMI values using classification tree models for wet zone sub basins for SWM and SIM seasons. (a) Morape (b) Peradeniya (c) Laxapana (d) Norwood (e) Norton Bridge

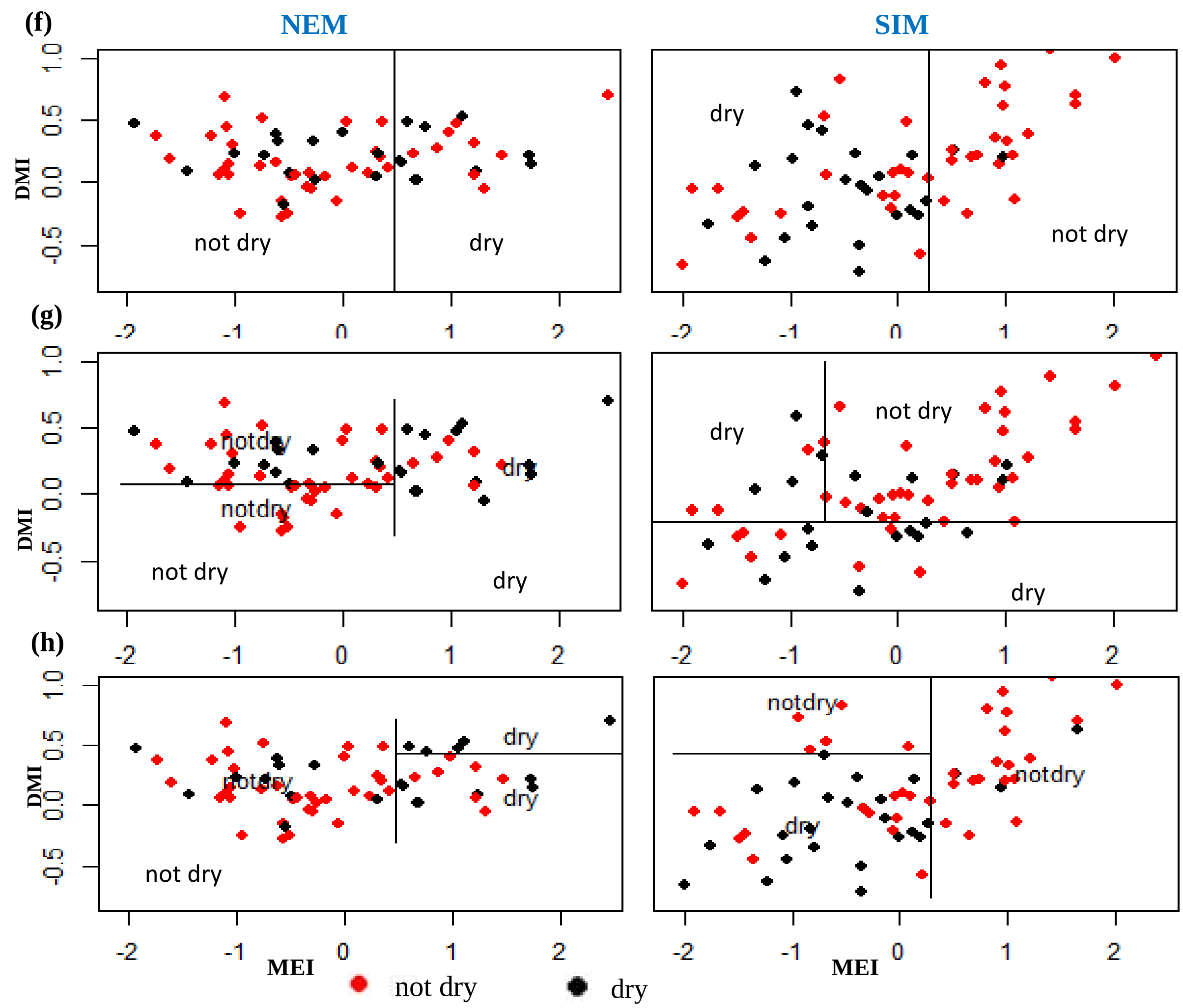


Figure A 9 Identifying relationships between two rainfall classes (dry, not dry) and MEI and DMI values using classification tree models for dry and intermediate zone sub basins for NEM and SIM seasons. (f) Randenigala (g) Bowatenna (h) Manampitiya

Table A. 6 Classification tree model results for major rainfall season to the sub basins.

Season	Morape		Peradeniya		Laxapana		Norwood		Norton Bridge	
	Dry	Not dry	Dry	Not dry	Dry	Not dry	Dry	Not dry	Dry	Not dry
SWM	21/24	22/40	18/23	26/41	19/23	27/41	12/20	34/44	19/22	29/42
SIM	10/19	39/45	12/19	30/45	8/21	36/43	11/20	38/44	13/22	36/42

Season	Randenigala		Bowatenna		Manampitiya	
	Dry	Not dry	Dry	Not dry	Dry	Not dry
NEM	11/24	31/40	14/24	34/40	13/23	34/41
SIM	23/24	22/40	15/21	32/43	22/25	26/39

Table A. 7 Random forest model results.

Season	Morape			Peradeniya		
	Dry	Not dry	OOB Error	Dry	Not dry	OOB Error
NEM	10/21	33/43	33%	8/20	34/44	34%
FIM	5/19	36/45	36%	6/20	37/44	33%
SWM	11/24	29/40	38%	11/23	28/41	39%
SIM	5/19	39/45	33%	5/19	37/45	34%
Season	**Randenigala**			**Bowatenna**		
	Dry	Not dry	OOB Error	Dry	Not dry	OOB Error
NEM	8/24	31/40	39%	15/24	33/40	25%
FIM	6/20	39/44	30%	13/21	38/43	20%
SWM	7/21	38/43	30%	11/25	29/39	38%
SIM	13/24	31/40	31%	6/21	35/43	36%
Season	**Laxapana**			**Norwood**		
	Dry	Not dry	OOB Error	Dry	Not dry	OOB Error
NEM	8/20	37/45	30%	10/19	39/45	23%
FIM	7/20	37/44	31%	8/19	39/45	26%
SWM	12/23	27/41	39%	7/20	37/44	31%
SIM	9/21	34/43	33%	7/20	37/44	31%
Season	**Norton Bridge**			**Manampitiya**		
	Dry	Not dry	OOB Error	Dry	Not dry	OOB Error
NEM	9/20	36/44	30%	13/23	33/41	28%
FIM	5/21	35/43	38%	8/21	35/43	33%
SWM	9/22	32/42	36%	5/16	34/43	39%
SIM	10/22	36/42	28%	16/25	34/39	22%

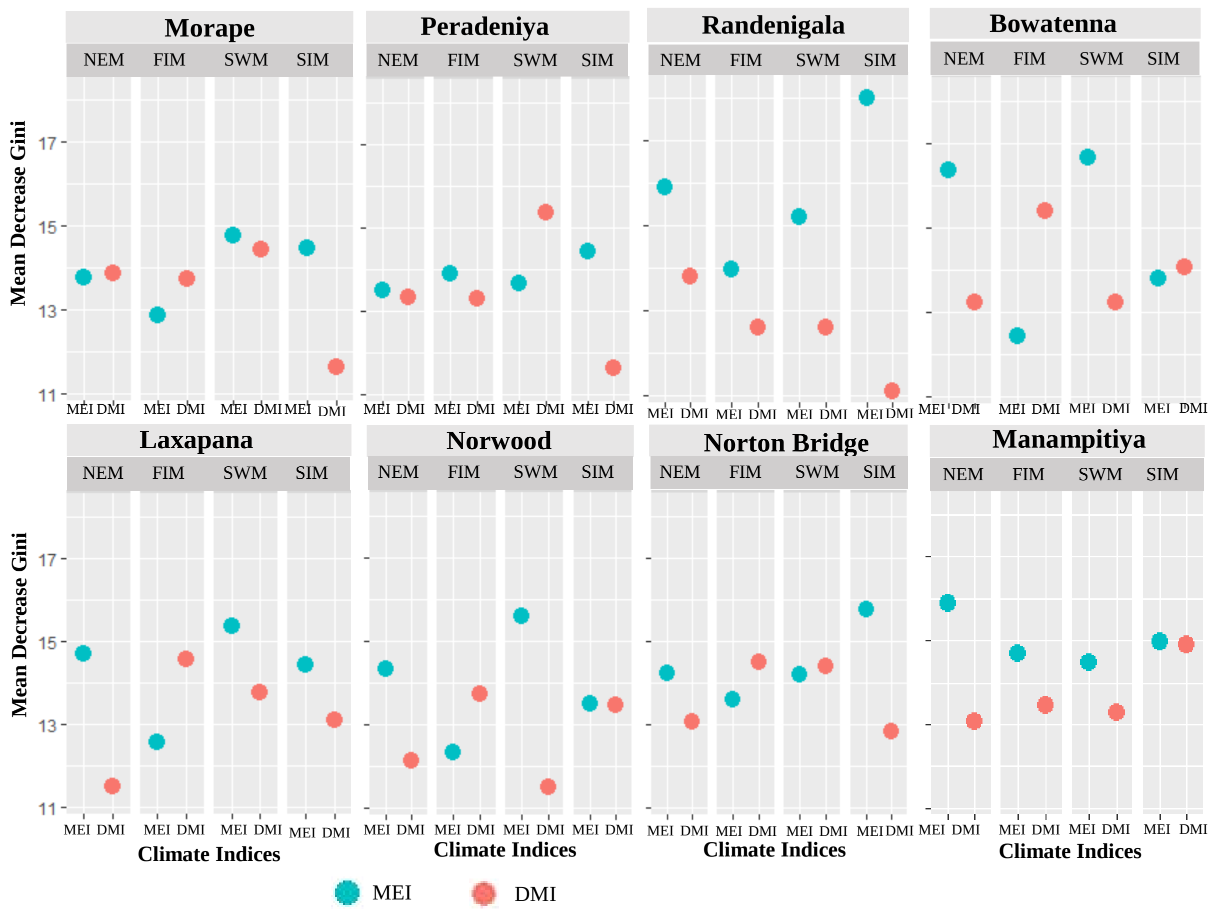


Figure A 10 Random forest importance of variable to identify the dry and not dry classes of rainfall anomalies

Author contributions

TDM and GMH conceptualized the study and TDM carried out the data analysis. TDM prepared the paper with contribution from GMH.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgement

This research is part of a multidisciplinary research initiative called Agricultural Decision-Making and Adaptation to Precipitation Trends in Sri Lanka (ADAPT-SL) at Vanderbilt Institute for Energy and Environment (VIEE). The work is supported by WSC Program Grant No. NSF-EAR 1204685.

References

Amarasekera, K. N., Lee, R. F., Williams, E. R., & Eltahir, E. A. B.: ENSO and the natural variability in the flow tropical rivers. Journal of Hydrology, 200(1–4), 24–39, https://doi.org/10.1016/S0022-1694(96)03340-9, 1997

Analytical Vidhya Team. (2015). Tunning the parameters of your Random Forest model., available at: https://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/, last access: 12 March 12 2018

Analytical Vidhya Team. (2016). A Complete Tutorial on Tree Based Modeling from Scratch., available at: https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/, last access: 12 March 2018

Block, P., & Goddard, L.: Statistical and Dynamical Climate Predictions to Guide Water Resources in Ethiopia, 138(June), 287–298. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000181, 2012.

Breiman, L.: Randomforest2001. Machine Learning, 45(1), 5–32. https://doi.org/10.1017/CBO9781107415324.004, 2001.

Bureau of Meteorology.: Record-breaking La Niña events. Bureau of Meteorology, 26. Retrieved from http://www.bom.gov.au/climate/enso/history/La-Nina-2010-12.pdf, 2012.

Bureau of Meteorology.: Indian Ocean, POAMA monthly mean IOD forecast., available at: http://www.bom.gov.au/climate/enso/#tabs=Indian-Ocean, 2017, last access: 30 March 2017

Cabrera, V. E., Letson, D., & Podesta, G.: The value of climate information when farm programs matter, 93, 25–42. https://doi.org/10.1016/j.agsy.2006.04.005, 2007.

Cao, Q., Hao, Z., Yuan, F., Su, Z., Berndtsson, R., Hao, J., & Nyima, T.: Impact of ENSO regimes on developing- and decaying-phase precipitation during rainy season in China. Hydrology and Earth System Sciences, 21(11), 5415–5426, 2017.

Ceylon Electricity Board.: Long Term Generation Expansion Plan 2015-2034. available at: http://pucsl.gov.lk/english/wp-content/uploads/2015/09/Long-Term-Generation-Plan-2015-2034-PUCSL.pdf, 2015, last access: 1 September 2017.

Ceylon Electricity Board.: River Basin Hydrology Data, System control branch Transmission Division. Sri Lanka, 2017.

Chandimala, J., & Zubair, L.: Predictability of stream flow and rainfall based on ENSO for water resources management in Sri Lanka. Journal of Hydrology, *335*(3–4), 303–312, https://doi.org/10.1016/j.jhydrol.2006.11.024, 2007.

Chandrasekara, S., Prasanna, V., & Kwon, H. H.: Monitoring Water Resources over the Kotmale Reservoir in Sri Lanka Using ENSO Phases. Advances in Meteorology, https://doi.org/10.1155/2017/4025964, 2017.

Chaudhari, H. S., Pokhrel, S., Mohanty, S., & Saha, S. K.: Seasonal prediction of Indian summer monsoon in NCEP coupled and uncoupled model. Theoretical and Applied Climatology, *114*(3–4), 459–477, https://doi.org/10.1007/s00704-013-0854-8, 2013.

Denise, C., Rogers, W., & Beringer, J.: Describing rainfall in northern Australia using multiple climate indices, 597–615. https://doi.org/10.5194/bg-14-597-2017, 2017.

Department of Agriculture Sri Lanka.: Climate zones of Sri Lanka., available at: https://www.doa.gov.lk/images/weather_climate/Climatezone.jpg, last access: 7 November 2017

Department of Meteorology Sri Lanka.: Climate of Sri Lanka., available at: http://www.meteo.gov.lk/index.php?option=com_content&view=article&id=94&Itemid=310&lang=en, last access:

7 November 2017

Dettinger, M. D., & Diaz, H. F.: Global Characteristics of Stream Flow Seasonality and Variability, Journal of Hydrometeorology, *1*(4), 289–310. https://doi.org/10.1175/1525-7541(2000)001<0289:GCOSFS>2.0.CO;2, 2000.

Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., & Mearns, L. O.: Climate extremes: Observations, modeling, and impacts. Science, *289*(5487), 2068–2074. https://doi.org/10.1126/science.289.5487.2068, 2000.

Eden, J. M., Van Oldenborgh, G. J., Hawkins, E., & Suckling, E. B.: A global empirical system for probabilistic seasonal climate prediction. Geoscientific Model Development, *8*(12), 3947–3973. https://doi.org/10.5194/gmd-8-3947-2015, 2015.

Gerlitz, L., Vorogushyn, S., Apel, H., Gafurov, A., Unger-Shayesteh, K., & Merz, B. A.: Statistically based seasonal precipitation forecast model with automatic predictor selection and its application to central and south Asia. Hydrology and Earth System Sciences, 20(11), 4605–4623. https://doi.org/10.5194/hess-20-4605-2016, 2016.

International research institute.: ENSO resources, El-Nino teleconnections & La-Nina teleconnections., available at: http://iri.columbia.edu/our-expertise/climate/enso/, 2017a, last access: 1 January 2017.

International research institute.: IRI ENSO forecast. available at: http://iri.columbia.edu/our-expertise/climate/forecasts/enso/current, 2017b, last access: 1 January 2017.

International Research Institute for Climate Society.: IRI Seasonal Precipitation Forecast., available at: http://iridl.ldeo.columbia.edu/maproom/Global/Forecasts/NMME_Seasonal_Forecasts/Precipitation_ELR.html, 2018, last access: 12 February 2018.

James, G., Witten, D., Hastie, T., & Tibshirani, R.: Springer Texts in Statistics An Introduction to Statistical Learning - with Applications in R. https://doi.org/10.1007/978-1-4614-7138-7, 2013.

Japan Agency for Marine Earth Science and Technology.: SST/DMI data set. available at: http://www.jamstec.go.jp/aplinfo/sintexf/DATA/dmi.monthly.txt, 2019.

Jentsch, A., Kreyling, J., & Beierkuhnlein, C.: A new generation of events , not trends experiments. Frontiers in Ecology and the Environment, 5(7), 365–374. https://doi.org/10.1890/1540-9295(2007)5[365:ANGOCE]2.0.CO;2, 2007.

Jha, S., Sehgal, V. K., Raghava, R., & Sinha, M.: Teleconnections of ENSO and IOD to summer monsoon and rice production potential of India. Dynamics of Atmospheres and Oceans, 76, 93–104. https://doi.org/10.1016/j.dynatmoce.2016.10.001, 2016.

Knapp, A. K., Hoover, D. L., Wilcox, K. R., Avolio, M. L., Koerner, S. E., La Pierre, K. J., Smith, M. D.: Characterizing differences in precipitation regimes of extreme wet and dry years: Implications for climate change experiments. Global Change Biology, 21(7), 2624–2633. https://doi.org/10.1111/gcb.12888, 2015.

Korecha, D., & Sorteberg, A. Validation of operational seasonal rainfall forecast in Ethiopia. Water Resources Research, 49(11), 7681–7697. https://doi.org/10.1002/2013WR013760, 2013.

Lee, H.: General Rainfall Patterns in Indonesia and the Potential Impacts of Local Seas on Rainfall Intensity. Water, 7(4), 1751–1769. https://doi.org/10.3390/w7041751, 2015.

Löwe, R., Madsen, H., & McSharry, P.: Objective classification of rainfall in northern Europe for online operation of urban water systems based on clustering techniques. Water (Switzerland), 8(3). https://doi.org/10.3390/w8030087, 2016.

Maity, R., & Nagesh Kumar, D.: Bayesian dynamic modeling for monthly Indian summer monsoon rainfall using El Niño-Southern Oscillation (ENSO) and Equatorial Indian Ocean Oscillation (EQUINOO). Journal of Geophysical Research Atmospheres, 111(7), 1–12. https://doi.org/10.1029/2005JD006539, 2006.

Malmgren, B. A., Hullugalla, R., Lindeberg, G., Inoue, Y., Hayashi, Y., & Mikami, T.: Oscillatory behavior of monsoon rainfall over Sri Lanka during the late 19th and 20th centuries and its relationships to SSTs in the Indian Ocean and ENSO. Theoretical and Applied Climatology, 89(1–2), 115–125. https://doi.org/10.1007/s00704-006-0225-9, 2007.

Malmgren, B. A., Hulugalla, R., Hayashi, Y., & Mikami, T.: Precipitatioon trends in Sri Lanka since the 1870s and relationships to El Niño-southern oscillation. International Journal of Climatology, *23*(10), 1235–1252. https://doi.org/10.1002/joc.921, 2003.

Manchanayake, P., & Madduma Bandara, C.: Water Resources of Sri Lanka. Sri Lanka: National Science Foundation, Sri Lanka. Retrieved from http://thakshana.nsf.ac.lk/slstic/NA-202/NA_202.pdf, 1999.

National Center for Atmospheric Research.: NINO SST Indices (NINO 1+2, 3, 3.4, 4; ONI and TNI). Retrieved from https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni, 2018.

National Oceanic and Atmospheric Administration.: El Nino Southern oscilation., available at: http://www.esrl.noaa.gov/psd/enso/past_events.html, 2017.

National Oceanic and Atmospheric Administration.: Three Months Outlook, Official Forecast, Climate Prediction Center, National Weather Services., available at: http://www.cpc.ncep.noaa.gov/products/predictions/long_range/seasonal.php?lead=6, last access: 12 February 2018,

Nur'utami, M. N., & Hidayat, R.: Influences of IOD and ENSO to Indonesian Rainfall Variability: Role of Atmosphere-ocean Interaction in the Indo-pacific Sector. Procedia Environmental Sciences, 33, 196–203. https://doi.org/10.1016/j.proenv.2016.03.070, 2016.

Ouyang, R., Liu, W., Fu, G., Liu, C., Hu, L., & Wang, H.: Linkages between ENSO/PDO signals and precipitation, streamflow in China during the last 100 years. Hydrology and Earth System Sciences, 18(9), 3651–3661. https://doi.org/10.5194/hess-18-3651-2014, 2014.

Power, S., Casey, T., Folland, C., Colman, A., & Mehta, V.: Inter-decadal modulation of the impact of ENSO on Australia. Climate Dynamics, 15(5), 319–324. https://doi.org/10.1007/s003820050284, 1999.

Qiu, Y., Cai, W., Guo, X., & Ng, B.: The asymmetric influence of the positive and negative IOD events on China's rainfall. Scientific Reports, *4*, 4943. https://doi.org/10.1038/srep04943, 2014.

Ranatunge, E., Malmgren, B. A., Hayashi, Y., Mikami, T., Morishima, W., Yokozawa, M., & Nishimori, M.: Changes in the Southwest Monsoon mean daily rainfall intensity in Sri Lanka: Relationship to the El Niño-Southern Oscillation. Palaeogeography, Palaeoclimatology, Palaeoecology, 197(1–2), 1–14. https://doi.org/10.1016/S0031-0182(03)00383-3, 2003.

Reason, C. J. C., Landman, W., & Tennant, W.: Seasonal to decadal prediction of southern African climate and its links with variability of the Atlantic ocean. Bulletin of the American Meteorological Society, 87(7), 941–955. https://doi.org/10.1175/BAMS-87-7-941, 2006.

Ropelewski C.F. and Halpert M.S.: Quantifying Southern Oscillation-Precipitation Relantionships. Journal of Climate., 1995.

Schneider, T., Bischoff, T., & Haug, G. H.: Migrations and dynamics of the intertropical convergence zone. Nature, 513(7516), 45–53. https://doi.org/10.1038/nature13636, 2014.

Seibert, M., Merz, B., & Apel, H.: Seasonal forecasting of hydrological drought in the Limpopo Basin: A comparison of statistical methods. Hydrology and Earth System Sciences, 21(3), 1611–1629. https://doi.org/10.5194/hess-21-1611-2017, 2017.

Singhrattna, N., Rajagopalan, B., Clark, M., & Kumar, K. K.: Seasonal forecasting of Thailand summer monsoon rainfall. International Journal of Climatology, 25(5), 649–664. https://doi.org/10.1002/joc.1144, 2005.

Singhrattna, N., Rajagopalan, B., Krishna Kumar, K., & Clark, M.: Interannual and interdecadal variability of Thailand summer monsoon season. Journal of Climate, *18*(11), 1697–1708. https://doi.org/10.1175/JCLI3364.1, 2005.

Smith, M. D.: The ecological role of climate extremes: Current understanding and future prospects. Journal of Ecology, 99(3), 651–655. https://doi.org/10.1111/j.1365-2745.2011.01833.x, 2011.

Suppiah, R.: Spatial and temporal variations in the relationships between the southern oscillation phenomenon and the rainfall of Sri Lanka, International Journal of Climatology, *16*(12), 1391–1407. https://doi.org/10.1002/(SICI)1097-0088(199612)16:12<1391::AID-JOC94>3.0.CO;2-X, 1996.

Surendran, S., Gadgil, S., Francis, P. A., & Rajeevan, M.: Prediction of Indian rainfall during the summer monsoon season on the basis of links with equatorial Pacific and Indian Ocean climate indices. Environmental Research Letters, 10(9), 094004. https://doi.org/10.1088/1748-9326/10/9/094004, 2015.

Verdon, D. C., & Franks, S. W.: Indian Ocean sea surface temperature variability and winter rainfall: Eastern Australia. Water Resources Research, *41*(9), 1–10. https://doi.org/10.1029/2004WR003845, 2005.

Ward, P. J., Eisner, S., Flo Rke, M., Dettinger, M. D., & Kummu, M. Annual flood sensitivities to el nintild;O-Southern Oscillation at the global scale. Hydrology and Earth System Sciences, 18(1), 47–66. https://doi.org/10.5194/hess-18-47-2014, 2014.

Whitaker, D. W., Wasimi, S. A., & Islam, S.: The El Niño - Southern Oscillation and Long-Range the Forecastong of Flows in the Ganges, International Journal of Climatology , 21(1), 77–87, 2001.

Zubair, L.: El Niño-southern oscillation influences on the Mahaweli streamflow in Sri Lanka. International Journal of Climatology, *23*(1), 91–102. https://doi.org/10.1002/joc.865, 2003.

Zubair, L., & Ropelewski, C. F.: The strengthening relationship between ENSO and northeast monsoon rainfall over Sri Lanka and southern India. Journal of Climate, 19(8), 1567–1575. https://doi.org/10.1175/JCLI3670.1, 2006.