

Reviewer comments are shown in blue color fonts, and authors' responses are given black color fonts.

General comments:

This paper attempts to apply statistical learning models to classifying rainfall in two river basins of Sri Lanka using the relationship between rainfall in Sri Lanka and El-Nino Southern Oscillation (ENSO), Indian Ocean Dipole (IOD). I believe the predictions in this study is helpful to the management of the water systems considering the impacts of climate change. My comments for this manuscript are as follows:

Response: Thank you for identifying the importance of the study. Shortcomings of the paper identified by the reviewers will be rectified as suggested, for the next submission. Response for the detail comments are given below.

1. Line 114-116. It is mentioned that the data will be tested for normality. If not, the data will be transformed. Can you provide the test results and specify the transformation function?

Response: Shapiro-Wilk's method is used to identify the normality of rainfall anomaly distribution. When the original data does not normally distribute, we carried out several transformation methods, and find the best transformation for particular rainfall data set. For example, Peradeniya NEM normality test results are given below.

Data 1: original data

W = 0.92157, p-value = 0.0005787

Data 2: data transformed by log

W = 0.92625, p-value = 0.000916

Data 3: data transformed by Square

W = 0.70236, p-value = 4.266e-10

Data 4: data transformed by Square root

W = 0.98177, p-value = 0.4632

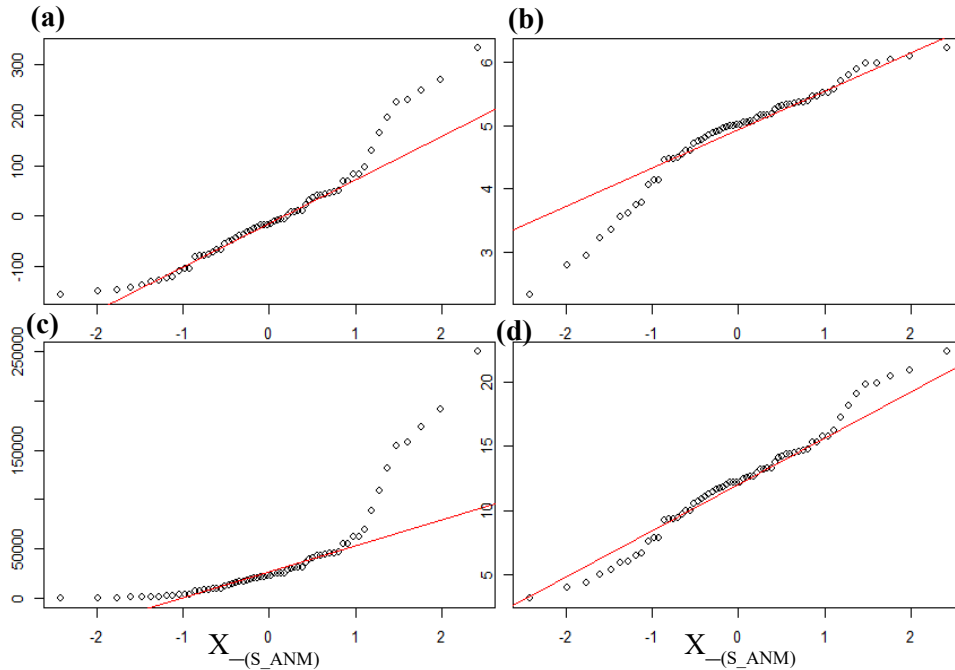


Figure S 1: Peradeniya NEM data qqplots (a) original form (b) log transformation (c) square transformation (d) square root transformation

Further, from data plots (Figure S 1) and the S-W statistic, we conclude that the square root transformed data is closer to being normally distributed than the other forms. In this particular example of Peradeniya NEM season, we used square root transformed anomalies. However, for other data sets, it can be a different transformation function.

2. Line 117: Using one's own definition of the anomaly is okay, but I think it is necessary to justify the cutting values of the three classes of anomalies.

Response: Definition of precipitation anomaly is a choice. Extreme seasonal precipitation were defined statistically in different manner in various studies using statistical thresholds (Easterling et al., 2000; Jentsch, Kreyling, & Beierkuhnlein, 2007; Knapp et al., 2015; Smith, 2011). 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 (Figure S 2). We deemed this a reasonable choice for our analysis.

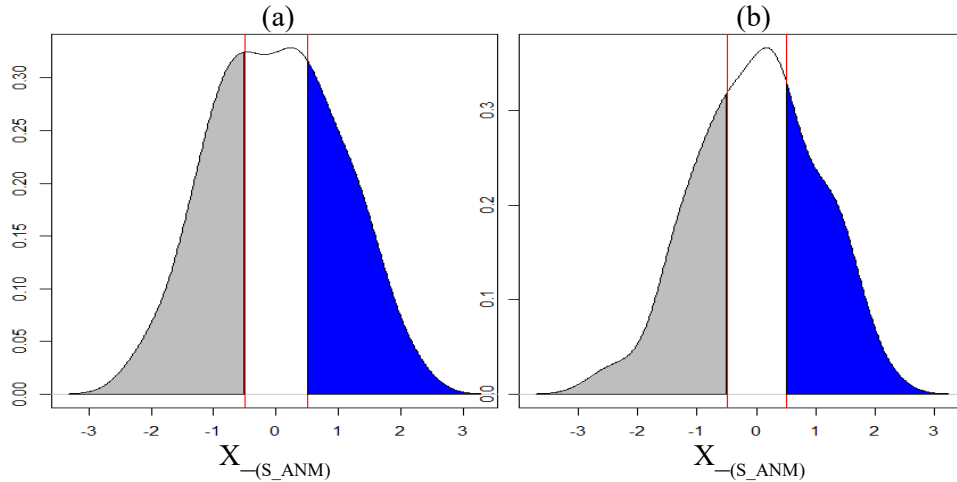


Figure S 2: (a) Norton Bridge SWM rainfall anomaly distribution (b) Manampitiya NEM rainfall anomaly distribution

3. a) Section 2.3: Before applying the three models, can you provide the reasons why these three particular models were selected to perform the analysis?

Response: We have explored several statistical techniques to identify the relationship between rainfall anomalies and ENSO and IOD climate indices. We have used linear discriminant analysis, quadratic discriminant analysis, regression and classification tree models, Naïve Bayes, KNN and Random Forest methods. Study results indicate that quadratic discriminant analysis, classification tree models and Random Forest models are given higher accuracy than the other methods for this application.

b) Also, I believe a bit more details about the three models are necessary in order to help more readers to understand how these models work. If possible, can you use graphs to illustrate the models?

Response: A brief outline of the Quadratic Discriminant Analysis method, Classification tree and Random Forest model are given below. More information about the methods can be obtained from the references as indicated in the paper.

2.3.1 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, Witten, Hastie, & Tibshirani, 2013; Löwe, Madsen, & McSharry, 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

$$Pr(Y = k|X = x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^K \pi_l f_l(x)} \quad \text{Eq.(1)}$$

In Eq (1), the posterior probability $Pr(Y = k|X = x)$ indicates that observation $X = x$ belongs to the k th class. For p predictors, the multivariate Gaussian distribution density function is defined for every class k (Eq.(2)).

$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k)\right) \quad \text{Eq.(2)}$$

In Eq.(2), Σ_k is the covariance matrix and μ_x is the mean vector. The covariance matrix (Σ_k) and mean (μ_x) for each class are estimated from the training data set (Eq.(3), Eq.(4)).

$$\mu_k = \frac{1}{N_k} \sum_{i:y_i=k} x_i \quad \text{Eq.(3)}$$

$$\Sigma_k = \frac{1}{(N_k - 1)} \sum_{i:y_i=k} (x_i - \mu_k)^T (x_i - \mu_k) \quad \text{Eq.(4)}$$

Substituting a Gaussian density function for the k^{th} class (Eq.(2)) into Bayes theorem and taking the log values, the quadratic discriminant function is derived (Eq.(5)). Prior probabilities for class k (π_k) is calculated by the frequency of data points of class k in the training data (Eq.(6)). For a total number of N points in the training observations, N_k is the number of observations belong to k th class.

$$\delta_k(x) = -\frac{1}{2} (x - \mu_x)^T \Sigma_k^{-1} (x - \mu_x) + \log \pi_k \quad \text{Eq.(5)}$$

$$\pi_k = \frac{N_k}{N} \quad \text{Eq.(6)}$$

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 $k \cdot p \cdot (p + 1) / 2$. 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.”

2.3.2 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.(7)). However, for the tree-growing, the Gini index (G) is considered as the criterion for splitting into regions (Eq.(8))

$$E = 1 - \max_k(\hat{p}_{mk}) \quad \text{Eq.(7)}$$

$$G = \sum_{k=1}^K \hat{p}_{mk} (1 - \hat{p}_{mk}) \quad \text{Eq.(8)}$$

In Eq.(7) and Eq.(8), \hat{p}_{mk} represents the fraction of observations in the m^{th} class that belong to the k^{th} 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.

2.3.3 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).

[4. Please give more details with regard to the parameter tuning in building the random forest model.](#)

Response: We carried out the study for 100 to 5000 number of trees to build the random forest model. Highest accuracy is given by 500 numbers of trees.

5. Can you specify the programming tools you used to perform the analysis?

Response: We used R programming language to carry out the classification. R packages: caret, tree, randomForest, fitdistribplus, devtools and quantreg are used for the studies.

6. Line 186, it would be better to increase the font size of Fig.3.

Response: I will improve the Fig.3 in the final submission as given below.

