MODELING THE INFLUENCE OF LARGE-SCALE CIRCULATION PATTERNS ON PRECIPITATION AND A MULTIVARIATE DROUGHT ANALYSIS FOR MAURITIUS

FINAL REPORT

C Prakash KHEDUN   Kreshna GOPAL   Anoop K SOHUN

SCINOVA CONSULTING

EXECUTIVE SUMMARY

Mauritius suffers from chronic water shortage problems that can severely impact its economy and the well-being of its population. For instance, in 1998-1999, the island faced a drought that resulted in a 40% decrease in sugar production, and a drop of about MUR 2 billion in GDP. On the social level, the water deficit the country faces is a major problem for residents. Water availability in reservoirs and major aquifers are influenced by precipitation regimes, which are affected by large-scale circulation patterns such as the El Niño Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD). In this study we (i) investigate the relationship between both ENSO and IOD with precipitation, (ii) develop an Artificial Neural Network (ANN) for precipitation prediction based on ENSO and IOD, (iii) develop statistical and time-series models for precipitation forecasting, and (iv) conduct a drought analysis based on multiple precipitation deficit variables (duration, severity and inter-arrival time). Monthly precipitation data for the period between 1961 to 2012 for the Vacoas station are used in this study.

We found some correlation between average winter ENSO indices and precipitation, while the correlation for summer was negligible. Statistically significant correlation was found between average winter precipitation and IOD index. We also found that the correlations fluctuate over time.

With ANNs, we obtained an average winter precipitation prediction accuracy of 86%. Prediction of summer precipitation was less accurate than winter precipitation. Results obtained from ANN were more accurate than those from other statistical techniques, such as linear regression and autoregressive integrated moving average (ARIMA). This may be attributed to the ability of ANNs in capturing the non-linearity in the system. Both ANNs and regression based models predict winter precipitation with remarkably higher precision than summer precipitation.

Standardized Precipitation Index (SPI) is proposed as a simple and effective index that can be used for drought definition. It is a spatially invariant quantity that can be computed to give precipitation deficit at multiple timescales. Based on SPI for six months for Vacoas, we found that drought durations vary between 1 and 9 months with a mean of 2.6 months. The mean inter-arrival time is 15 months. We also identified the severity of all drought events between 1961 and 2012. We computed the conditional drought duration and severity.

The findings from this study can help in more efficient planning and management of scarce water resources on the island. This study should be extended in several ways: (i) cover longer time periods to investigate possible impact of climate change, and (ii) cover more stations across the island, (iii) incorporate other meteorological parameters, and (iv) develop models to predict other key variable such as water levels in reservoirs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1   INTRODUCTION

1.1   BACKGROUND

The water shortage problem that Mauritius regularly faces is alarming. Based on the United Nations' definition [*United Nations*, 2011], Mauritius is a water stressed country. The Water Resources Unit in Mauritius predicts that by 2020, the country will slip into the water scarce category [*Deepchand and Khedun*, 2005]. The droughts in 1998-1999 and 2010-2011 were particularly acute, with alarming low water levels in reservoirs and boreholes compared to other years.

The demand for water in Mauritius has been growing rapidly, with an increasing and more affluent population as well as new towns, hotels, industrial zones, residential and mixed-use properties. Furthermore, the island does not have enough carry-over capacity i.e. most of the water received during a hydrologic year is used within the year itself. Thus, a shortage of rainfall can have immediate and disastrous effects on the well-being of the population and various economic sectors of the island.

1.2   RATIONALE AND OBJECTIVES

There have been numerous solutions proposed, in terms of better water management as well as novel infrastructure and technology. Robust and effective solutions can only be developed with a rigorous scientific understanding of the various facets of this complex problem – from the impact of El Niño/La Niña on precipitation in the Indian Ocean, to the legal system of water ownership, through the applicability of technologies like desalination and cloud seeding in Mauritius.

In this project, we aim at filling an evident gap in the understanding of the water problem in Mauritius: exploit climatic and meteorological data, and the latest research in the field to explain and predict rainfall and the level of water in our reservoirs. In particular, the impact of large scale circulation patterns on precipitation is studied. We adopt an interdisciplinary approach that taps on the extensive work done in hydroinformatics, an emerging research and development area on state-of-the-art information technologies to address problems in hydrology.

We use past climatic data to develop explanatory and predictive models for precipitation. Besides traditional statistical approaches such as time-series analysis, and regression, we are also developing machine learning and statistical pattern recognition tools such as artificial neural networks.

Our objective is to develop a model that uncovers the most likely hypothesis of what determines water levels. We are focusing on precipitation at Vacoas because of its proximity to Mare aux Vacoas, one of the most important reservoirs in Mauritius, where the drop in the water level is particularly acute and alarming. We argue that such models are crucially important because they can shed light on the underlying phenomena. A predictive model can enable medium and long term planning, and facilitate proactive measures for forecasted water shortages and droughts. An explanatory model can help in the development of robust, sustainable and effective solutions.

We hypothesize that there exist a statistically significant relationship between large scale circulation patterns and local hydrological conditions, especially precipitation, which determines water level in the reservoirs. A schematic diagram of the research hypothesis is given in Figure 1.
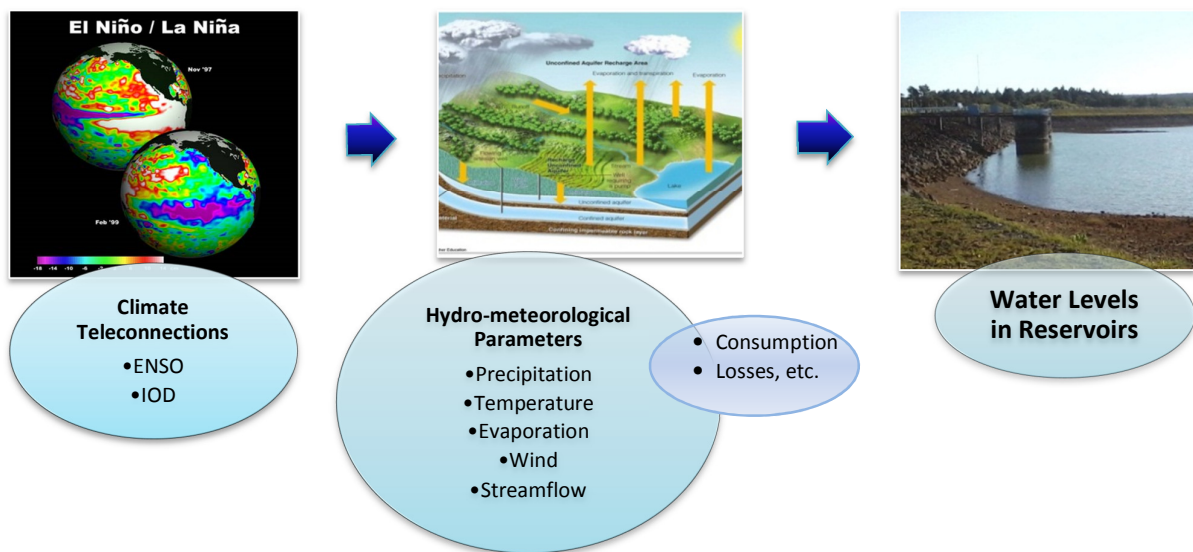
Figure 1. Schematic of research hypothesis

Our objectives in this study are framed around this hypothesis, and consist of two major components: (1) establish the relationship between indices of large-scale circulation patterns (ENSO and IOD) and precipitation in Mauritius, and (2) use these indices to predict precipitation. We also aim to develop a framework for defining drought events based on multiple parameters, such as duration, severity, and inter-arrival time. This framework can be useful in better understanding drought phenomena and determining their conditional probability of occurrence.

One of our initial objectives was to develop a model to establish the relationship between water level in Mare aux Vacoas reservoir to the various factors that affect inflow and outflow of water

(precipitation, water consumption, evaporation, etc.). However, historical data on water level and consumption for the reservoir were not available. Thus, we re-formulated our objectives to focus on precipitation prediction, and also added a new component dealing with multivariate drought analysis.

1.3 OVERVIEW OF METHODOLOGY

1.3.1 Hydroinformatics

Hydroinformatics [*Kumar et al.*, 2006] is an interdisciplinary field for the study of the application of information and communications technologies in addressing the challenges of equitable and efficient use of water for many different purposes. It stems from techniques originating from artificial intelligence, such as artificial neural networks, support vector machines, genetic programming, Markov chains, etc. Hydroinformatics draws from the following disciplines: hydraulics, hydrology, environmental engineering, computer science, statistics, and machine learning. It is used at all stages in the water cycle, from the atmosphere to the ocean, through water supply systems, urban drainage, etc. It provides support for decision making at all levels from governance and policy through management to operations.

1.3.2 Statistics and Time-Series Analysis

Statistics and time-series analysis tools and techniques for analyzing data have been extensively used for establishing patterns and forecasting in atmospheric sciences and hydrology. Coupled with more powerful computational tools, very complex models and routines can be developed that are theoretically sound as well as computational tractable. With the emergence of numerous statistical software (like MATLAB, SAS, Stata, R, etc.), multivariate techniques in the time and frequency have been made more accessible. Extremely large datasets driven by periodic phenomena can now be handled using extensions to time series of classical methods like multivariate regression, analysis of variance, principle components, factor analysis, and cluster analysis [*Silverman and Dracup*, 2000].

2   DATA

2.1   PRECIPITATION

Monthly precipitation data for the Vacoas station was obtained from two sources. Data for the period 1961 to 1990 was obtained from the National Climate Data Center (NCDC) and is part of the Global Historical Climatology Network (GHCN). Data beyond 1990 was not available from the NCDC and was purchased from the Mauritius Meteorological Services (MMS). Figure 2 gives a plot of the monthly precipitation for January 1961 to September 2012, along with the long-term mean (170.54 mm). It is apparent from the plot that the precipitation is not around the long-term mean but is highly variable. For most months, the precipitation is below the long-term mean, and for some months, the precipitation is several magnitudes above the long-term mean.
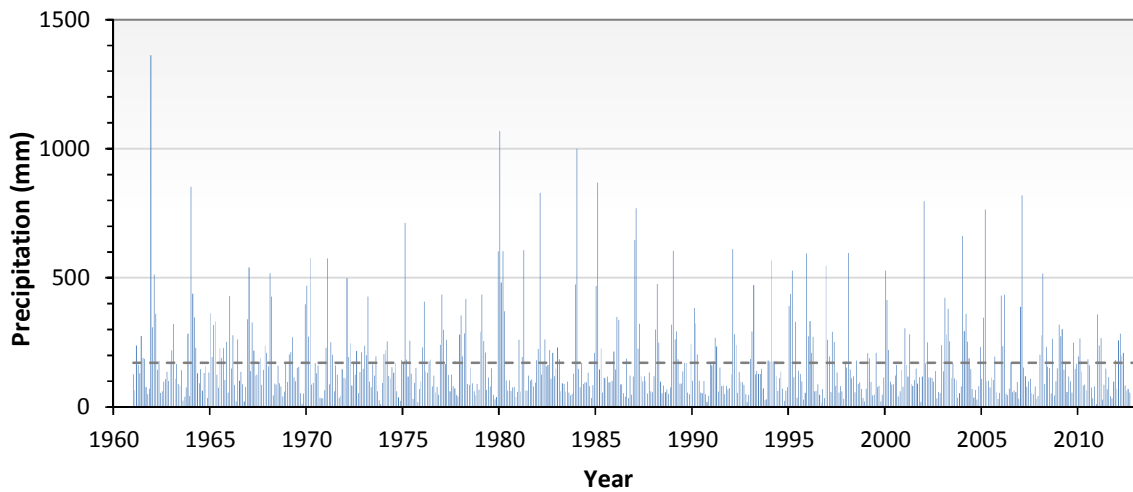
Figure 2. Monthly precipitation at Vacoas for January 1961 to December 1990. The dotted gray line represents the long-term mean (170.54 mm).

Figure 3 is a boxplot for the monthly precipitation for each month. The largest variation in precipitation is noted for February followed by January, which are also the two wettest months. One notable outlier is for December, where the monthly total was 1,362 mm (December 1961). The median for precipitation is generally lower than the long-term mean for each month (Figure 4), which implies a slight positive skewness in monthly precipitation.
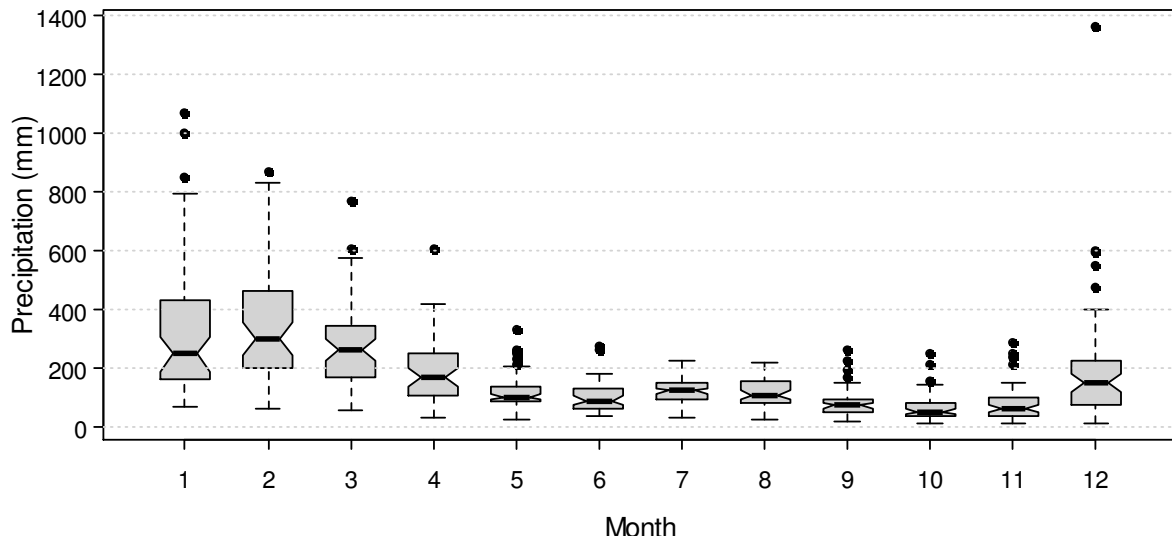
Figure 3. Boxplot of monthly data for Vacoas station

The mean monthly precipitation was computed for the different periods:

- 1961-1990, which is data obtained from the NCDC,

- 1991-2012, which is data purchased from MMS, and

- 1961-2012, the total precipitation series.

Figure 4 shows the mean monthly precipitation for the three periods. Precipitation at Vacoas is unimodal, with the period December to April recording the highest precipitation. We note that the mean monthly precipitation for the period 1991-2012 is considerably lower than the period 1961-1990. Table 1 gives the mean monthly precipitation for the three periods considered and the percentage decrease in rainfall for 1991-2012 with respect to 1961-1990. The largest decrease is noted for December (31%) followed by August (23%), and January (20%).

The percentage change in mean precipitation is atypical, and should not be hastily attributed to climate change without further investigation. There seem to be a shift in the precipitation time series. Such shift may be due to multiple factors, for example, a change in the type of recording instrument, change in the location of the recording instrument, or a change in the nature of the surroundings of the instrument. *Booneeady* [2010] illustrates such shifts in temperature recorded at the Pamplemousses station. Further analysis, coupled with historical information from MMS

may help identify the location in time and cause of the shift. For the purpose of our study, however, we assume that the data is homogeneous and stationary.
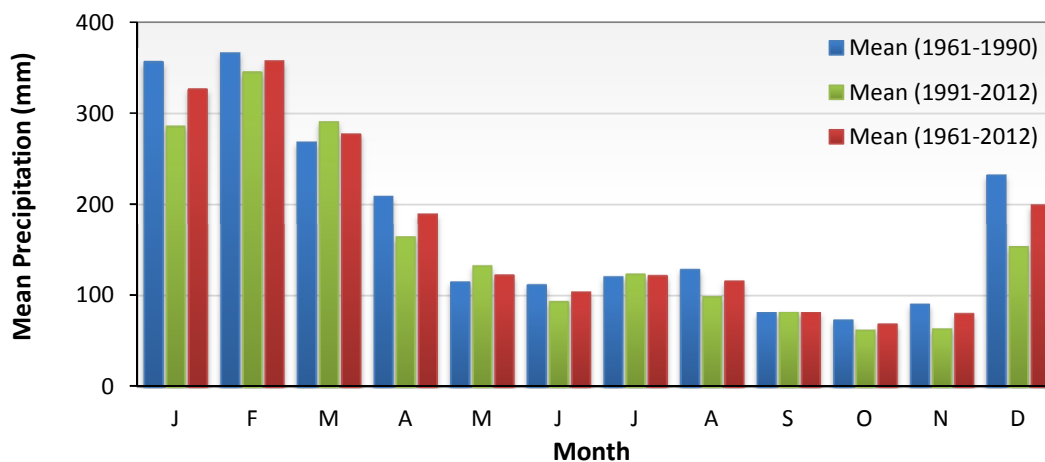
Figure 4. Mean monthly precipitation for the period 1961-1990, 1991-2012, and 1961-2012

Table 1. Mean monthly precipitation for the period 1961-1990, 1990-2012, and 1961-2012 and percentage decrease in mean precipitation between 1961-1990 and 1990-2012

Month	Mean (1961-1990)	Mean (1991-2012)	Mean (1961-2012)	% Decrease
January	356.51	286.395	326.85	19.67
February	366.30	345.277	357.41	5.74
March	268.13	291.450	278.00	−8.70
April	209.10	163.909	189.98	21.61
May	114.80	132.400	122.25	−15.33
June	111.72	93.468	104.00	16.34
July	120.44	123.277	121.64	−2.36
August	128.29	98.814	115.82	22.97
September	80.37	81.309	80.77	−1.17
October	72.23	61.343	67.75	15.07
November	90.79	62.671	79.21	30.97
December	232.13	153.214	199.64	34.00

Several indices are available for ENSO, namely the Southern Oscillation Index (SOI), Niño 3, Niño 4, Niño 3.4, the Multi ENSO Index (MEI), among others. The Niño 3.4 index, proposed by *Barnston et al.* [1997] has been employed in a number of hydro-meteorological studies as it has both maximum correlation with the core ENSO phenomenon and strongest influence on remote teleconnection events. It is the area-averaged sea surface temperature anomaly (SSTA) over the region bounded by 5°N–5°S and 120°W–170°W, straddling the Niño 3 and Niño 4 regions. Monthly data for the Niño 3.4 index was obtained from the International Research Institute (IRI) on Climate and Society Data Library.

2.2.1 Niño Indices

Four Niño regions (1, 2, 3, and 4 as shown in Figure 6) along the equatorial Pacific have been chosen to monitor SST. The warming in these regions are not uniform and no single region can capture the whole ENSO phenomenon. *Barnston et al.* [1997] investigated the correlation between the index from each region with remote hydro-meteorological variables and proposed the Niño 3.4 index. The latter has the highest correlation between the core ENSO phenomenon and remote teleconnection events. A plot of the Niño 3.4 index for the period January 1961 to July 2009 is given in Figure 6. Positive (negative) SSTAs are shown in red (blue). Sustained positive (negative) anomalies are indicative of El Niño (La Niña) conditions.

A number of different definitions have been used to define El Niño (La Niña) events. *Mackay* [2003] suggested the use of the Niño 3.4 index for the definition of El Niños, and "El Niño can be said to occur if 5-month running means of sea surface temperature (SST) anomalies in the Niño 3.4 region (5°N–5°S, 120°–170°W) exceed 0.4°C for 6 months or more." The reverse of this definition can be adopted for the definition of La Niña events. The Niño 3.4 index plotted in Figure 6 is overlain with a 5-month moving average and the ±0.4 thresholds.

Other definitions, with slight variations to the one proposed by *Mackay* [2003], have since been proposed and have been adopted in different regions of the world. To our knowledge, no official definition for El Niño or La Niña has been developed or adopted for Mauritius.
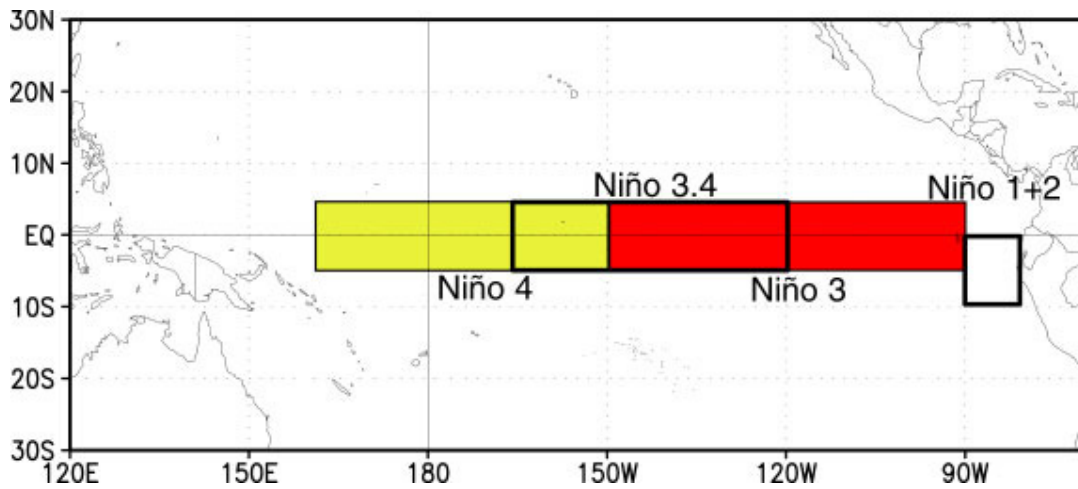
Figure 5. Niño regions in the Pacific Ocean (Source: National Oceanic and Atmospheric Administration)
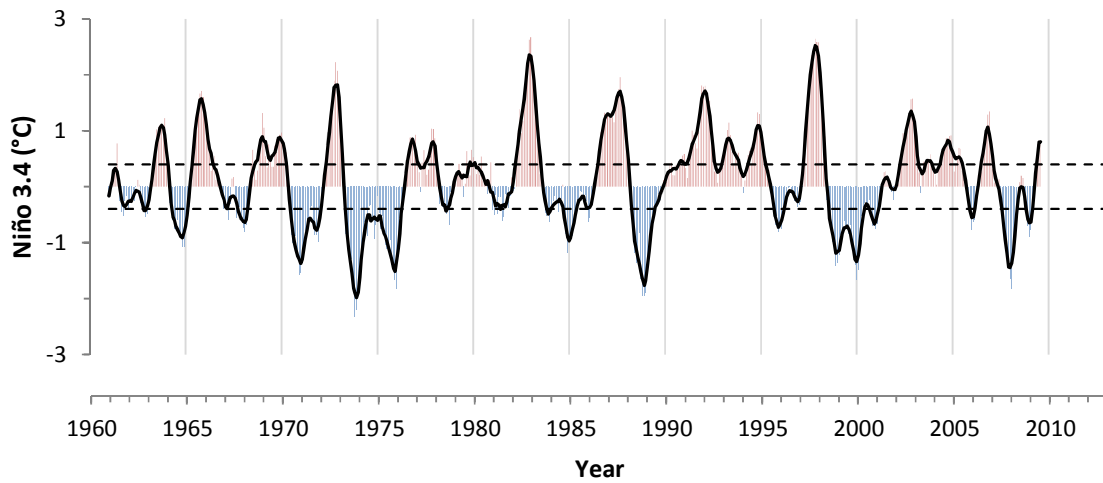
Figure 6. Time series of Niño 3.4. The Niño 3.4 series is overlain with a 5-month centered moving average series and ±0.4 thresholds.

2.2.2 Southern Oscillation Index

The Southern Oscillation Index (SOI) is an index that shows the variation in the Southern Oscillation. It is the difference in mean sea level pressure anomalies at Tahiti and Darwin, Australia (Figure 7). Data for the SOI was obtained from the National Center for Atmospheric Research (NCAR) Climate and Global Dynamics (CGD).

Figure 8 shows a plot of SOI for January 1961 to December 1990. The data is smoothed with a 13-month centered moving average to highlight the positive and negative phases.
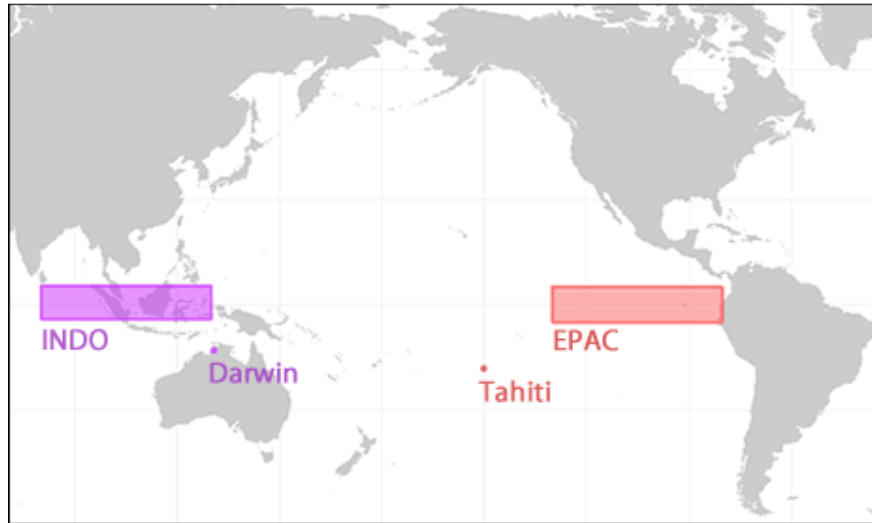
Figure 7. Location of Tahiti and Darwin, Australia (Source: National Oceanic and Atmospheric Administration)

Note that the SOI is negatively correlated with the Niño 3.4 index, i.e. an El Niño (La Niña) will be represented by positive (negative) Niño 3.4 indices and the SOI will be negative (positive).
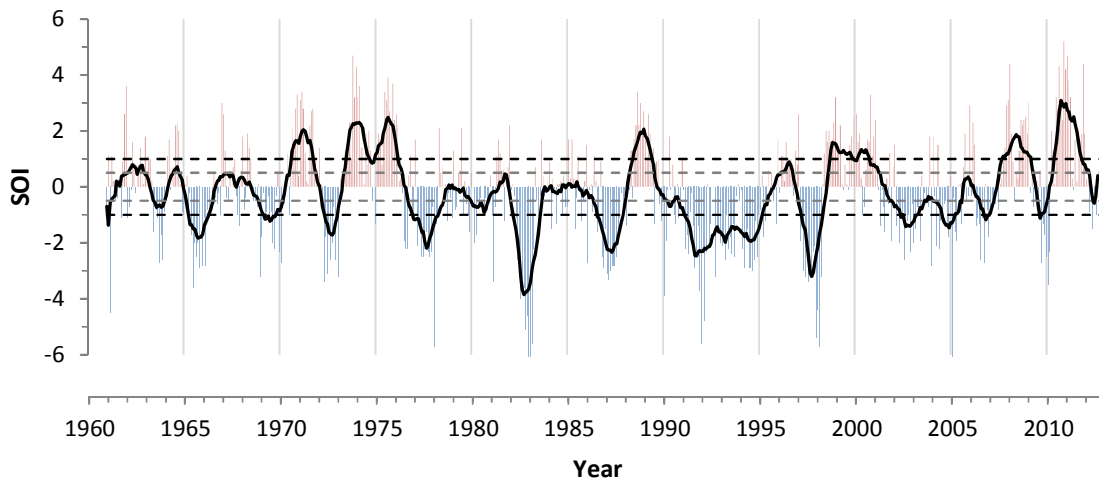
Figure 8. Time series of the Southern Oscillation Index (SOI). The SOI series is overlain with a 13-month centered moving average series and ± 0.5 and ± 1 thresholds.

The strength of El Niño and La Niña events can be gauged based on the scale given in Table 2.

Table 2. Criteria for El Niño (La Niña) Strength (Source: Desert Research Institute)

El Niño (La Niña) Strength	SOI Value
Stronger El Niño	Strongly Negative SOI (−1.00 or less)
Moderate El Niño	Mildly Negative SOI (−0.50 or less)
Neither Niño	Neither (SOI between -0.50 and +0.50)
Moderate La Niña	Mildly Positive SOI (+0.50 or more)
Stronger La Niña	Strongly Positive SOI (+1.00 or more)

2.3 INDIAN OCEAN DIPOLE

IOD is represented by anomalous SST gradient between the western equatorial Indian Ocean (50°E-70°E and 10°S-10°N) and the south eastern equatorial Indian Ocean (90°E-110°E and 10°S-0°N). The location of the west and east poles are shown in Figure 9. The gradient is known as the Dipole Mode Index (DMI) and was obtained from the Japan Agency for Marine-Earth Science and Technology (JAMSTEC). Figure 10 is a plot of the DMI for the period January 1961 to September 2010. Data beyond the latter date was not available.
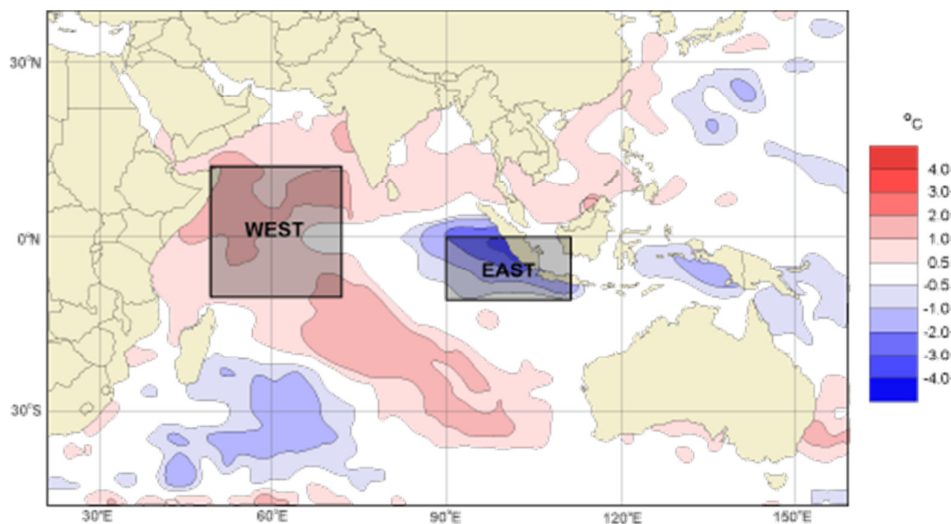
Figure 9. Map showing the location of the west and east poles of the IOD, marked with black boxes. (Source: Bureau of Meteorology, Australia)
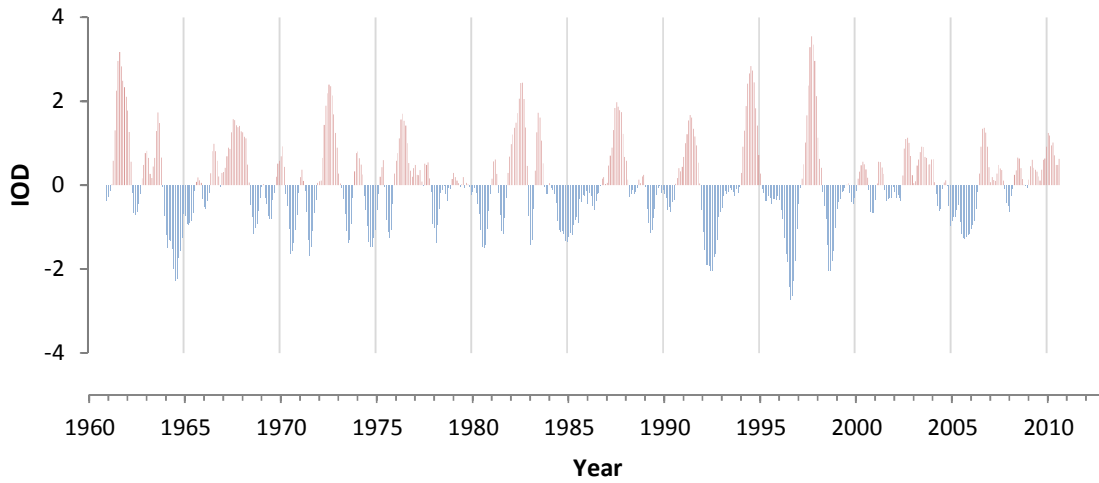
Figure 10. Time series for IOD

2.4 ASSOCIATION OF ENSO AND IOD WITH PRECIPITATION

Pearson correlation is used to determine the correlation between climate indices and precipitation. The Pearson correlation coefficient, ρ_{xy}, is a measure of linear association between two time series: x and y and is given by

$$\rho_{xy}(k) = \frac{E[(x_i - \mu_x)(y_{i+k} - \mu_y)]}{\sigma_x \sigma_y} \tag{1}$$

where $\rho_{xy}(k)$ is the cross-correlation for lag k between time series x_i and y_i with means μ_x and μ_y and standard deviation σ_x and σ_y respectively, and $E[\cdot]$ is the expectation operator. The range for $\rho_{xy}(k)$ is [-1, 1], with larger $|\rho_{xy}|$ implying greater ability of x to predict y.

The correlation coefficient can be used as a statistical test of independence to help make inferences about the degree of association between variables. The null hypothesis is that the two time series x_i and y_i are independent and identically distributed (iid) normal random variables ($\rho_{xy} = 0$). The test statistics t is defined as

$$t = \frac{\rho_{xy}\sqrt{n-2}}{\sqrt{1 - \rho_{xy}^2}} \tag{2}$$

16

The null hypothesis is rejected if $|t| > t_{crit}$, where t_{crit} is from the Student's t distribution with $n - 2$ degrees of freedom with a probability of exceedance of $\alpha/2$ [*Hirsch et al.*, 1992]. The form of the alternate hypothesis determines whether a one-sided or two-sided test is performed.

2.4.1 Correlation between Large-scale Circulation Patterns and Precipitation

Mauritius has two seasons: winter and summer. Winter is from May to October and summer is from November to April. November is also the beginning of the hydrological year. In order to assess the influence of ENSO and IOD on precipitation at Vacoas, the precipitation series was divided into two series, winter and summer, and the average precipitation for each year was computed. The Niño 3.4, SOI and IOD series were similarly divided into two sets.

The correlation between Niño 3.4, SOI and IOD with average winter and summer precipitation were computed. The null hypothesis is that the seasonal average precipitation is not dependent on the indices. The magnitude and sign of the correlation coefficient indicates the existence, strength, and nature of any association between the index representing the phenomenon under study and precipitation recorded.

Correlation between ENSO and Precipitation

Figure 11 is a scatter plot of average winter precipitation anomalies against average winter SOI, and average summer precipitation anomalies against average summer SOI. Precipitation anomaly is the difference between the average precipitation for each year and the long term mean for the season. No trend is apparent in the summer plot, but a small negative trend, above average rainfall with negative SOI, is visible in the winter plot.
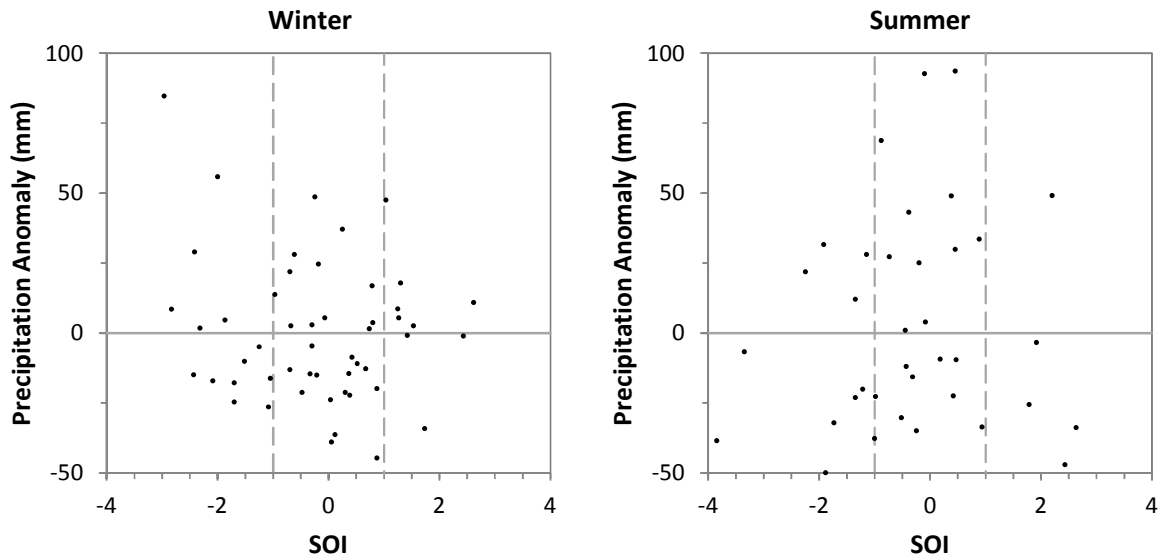
Figure 11. Scatter plot of average precipitation anomaly and average SOI for winter and summer. The two vertical lines show the ± 1 thresholds.

Table 3 gives the correlation coefficient and *p*-values for the correlation between the two ENSO indices and precipitation.

Table 3. Correlation coefficient between ENSO indices and precipitation anomaly at Vacoas. Numbers in bracket are the *p*-values of the correlation.

Index	Period	Winter	Summer
Niño 3.4	1961-1990	0.237 (0.208)	0.031 (0.875)
	1991-2009	0.065 (0.797)	0.296 (0.219)
	1961-2009	0.161 (0.276)	0.075 (0.610)
SOI	1961-1990	−0.329 (0.076)	−0.015 (0.941)
	1991-2012	0.042 (0.858)	−0.348 (0.113)
	1961-2012	−0.184 (0.198)	−0.118 (0.411)

The first thing to note is that the correlation between Niño 3.4 and precipitation is of opposite sign compared to SOI and precipitation. This is because Niño 3.4 and SOI are negatively correlated as explained above.

There exist a small correlation between winter indices and precipitation, for the 1961-1990 series, while there is no apparent correlation between the indices and summer precipitation. The correlation not statistically significant at an alpha of 0.05 but statistically significant at an alpha of 0.1. The correlation between SOI and winter precipitation is stronger than that between Niño 3.4 and precipitation. Even though both phenomena are related, stronger correlation with SOI has often been observed. One possible explanation can be that SOI is an Indo-Pacific phenomenon and recorded closer to Mauritius while Niño 3.4 is recorded further away from Mauritius. Further, precipitation may respond better or faster to atmospheric fluctuations than to changes in SST. Thus SOI may be a better index for the purpose of this study. Given that the phenomenon is cause and effect, we will investigate the effect of lagging SOI with respect to precipitation.

The fact that there is a correlation between winter SOI and precipitation anomaly in the 1961-1990 series but no correlation with the 1991-2012 or 1961-2012 series again prompts to the fact that the series may be inhomogeneous.

Correlation between IOD and Precipitation

Figure 12 is a scatter plot of average winter precipitation anomalies against average winter IOD, and average summer precipitation anomalies against average summer IOD. No trend is visible in the summer plot, but a positive dependence of precipitation with IOD is noted for winter.
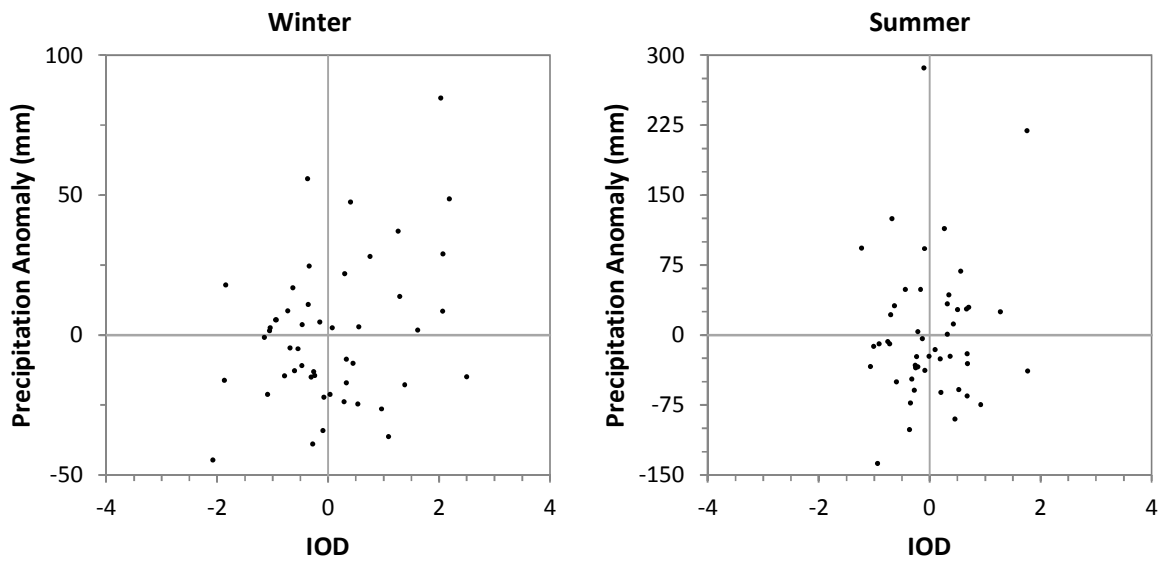
Figure 12. Scatter plot of average precipitation anomaly and average IOD for winter and summer.

Table 4 gives the correlation between average IOD index and average precipitation anomaly for winter and summer. The correlation for three different periods was computed. A positive statistically significant correlation exists between average winter IOD and precipitation anomaly at Vacoas.

Table 4. Correlation coefficient between the IOD index and precipitation anomaly at Vacoas. Numbers in bracket are the *p*-values of the correlation.

Index	*Period*	*Winter*	*Summer*
IOD	1961-1990	0.375 (0.041)	0.108 (0.572)
	1991-2009	0.210 (0.388)	0.135 (0.482)
	1961-2009	0.309 (0.031)	0.124 (0.394)

3 PRECIPITATION PREDICTION USING ARTIFICIAL NEURAL NETWORKS

3.1 METHODOLOGY

An artificial neural network (ANN) is a computational model that is inspired by the structure of biological neural networks in the brain. It consists of interconnected artificial neurons and it processes information using a connectionist approach to computation. ANNs are used to model complex relationships between inputs and outputs and to find non-linear patterns in data. They are in essence non-linear statistical data modeling tools. An ANN is an adaptive system that changes its structure (like weights between connections) based on information provided to the network during a learning phase. It can then be used for prediction of outputs given new input data.

Artificial neural networks have been successfully used in a variety of applications such as medical diagnosis, vehicle control, recognition of faces and objects, e-mail spam filtering, game-playing, handwritten recognition, automated trading systems, etc. ANNs are widely used in hydroinformatics applications like streamflow forecasting, rainfall estimation, subsurface characterization, watershed runoff estimation, etc. [*Govindaraju and Rao*, 2000].

Artificial neural networks are built with densely interconnected units known as neurons. This is analogous to the way biological systems are made up of interconnected neurons. In artificial neural networks, each neuron takes a number of real-valued inputs (possibly the outputs of other units) and produces a single real-valued output, which may become the input to many other units [*Mitchell*, 1997].

In natural systems, information processing often occurs in stages. For example, in human vision, light falling on the retina is first transformed into patterns that the brain can process. Local parts of an image are then recognized by neurons specialized to respond to particular image patterns. The information, or features, extracted from these neurons is then fed into subsequent layers, which correspond to higher cognitive functions.

Similarly, artificial neural networks consist of different stages and these stages can be represented by layers. Each layer consists of a specialized set of neurons, and performs a specific function. Artificial neural networks normally consist of three layers: the input layer, the hidden layer, and the output layer. The input layer simply accepts inputs from the external environment, while the output layer presents the results obtained to the external environment. Hidden layers are various levels of transformation of the data to establish the connection between inputs and

outputs. The number of hidden layers varies; hidden layers may even be absent in certain types of neural networks.

A neuron has a number, i, of inputs x_i and one output y. Associated with each input is weight w_i. There may be an additional parameter w_0, called a bias, which we may view as being the weight associated with an input x_0 and is permanently set to 1. The single neuron is a feed-forward device, that is, the connections are directed from the inputs to the output of the neuron.

The activation, a, of the neuron is calculated according to the inputs x:

$$a = \sum_i w_i\, x_i$$

Then, the output y is set as a function $f(a)$ of the activation. The output is also called the activity of the neuron [*Mackay*, 2003].

3.1.1 Multi-Layer Perceptron

The multi-layer perceptron (MLP) is the most widely used type of neural network. It is a feed-forward neural network, that is, it contains no feedback loops. MLPs usually have 3 layers: an input layer, a hidden layer and an output layer (Figure 13). However, MLPs can also have more than one hidden layer. The figure below shows an MLP with three layers. The input layer has two neurons, the hidden layer has four neurons and the output layer has one neuron.
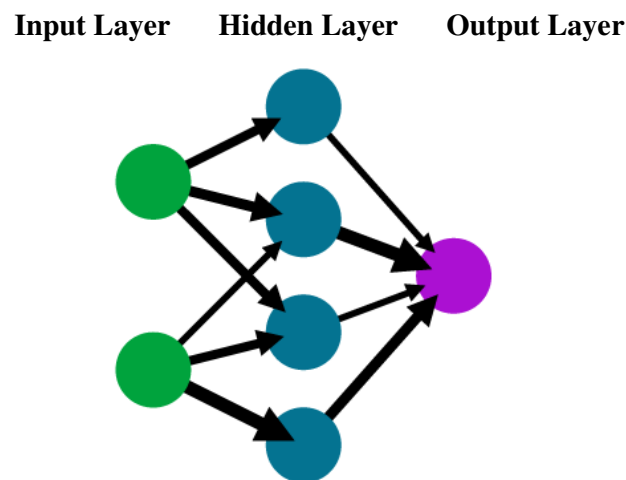
Figure 13. A multi-layer perceptron.

3.1.2 Sigmoid Activation Function

The activation function typically used by MLPs is the sigmoid activation function. Before the output of a neuron is fed to another neuron, the activation of the former neuron is calculated and passed to the sigmoid activation function, *f(x)*, defined as follows (and plotted in Figure 14).

$$f(x) = \frac{1}{1 - e^{-x}}$$
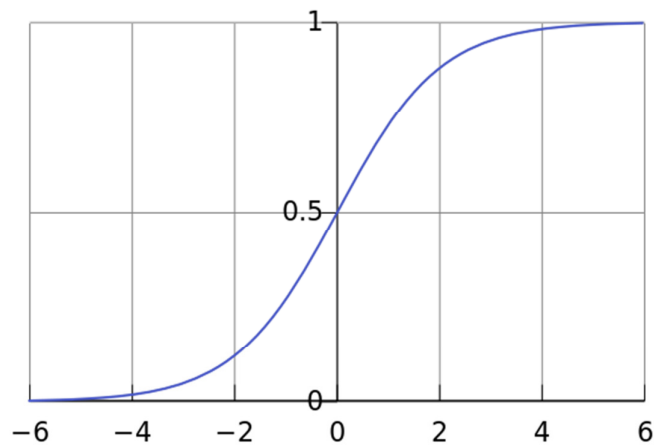
Figure 14. Sigmoid activation function.

3.1.3 Training Algorithms

The true power and advantage of neural networks lie in their ability to capture both linear and non-linear relationships and to learn these relationships directly from the data being modeled. Training algorithms give neural networks the ability to learn. These algorithms enable the neural network to learn patterns presented to the neural network.

Back-propagation is the most widely used training algorithm [*Maier and Dandy*, 2000]. It is a form of supervised learning, where the neural network is given a set of training examples. Each example consists of a pair of input and output and the input is typically representing by a series of values. Other training algorithms, such as genetic algorithms, have also been developed to train MLPs.

3.2 RESULTS

3.2.1 Predicting Precipitation using Multi-Layer Perceptron

We have investigated the use of neural networks for the prediction of precipitation using the data from the Vacoas station. We have used an approach proposed by *Silverman and Dracup* [2000], who successfully applied the technique for long range precipitation prediction in California, USA. We started our tests trying to predict the yearly average precipitation, and then seasonal and monthly precipitation. We devised a series of tests to predict the seasonal average precipitation, showed in Table 5. Example instances that are fed to the neural network are also given.

Table 5. Test scenarios

Test	*Description*	
1	Prediction of average yearly precipitation using previous year's ENSO/IOD values e.g.:	
	Input (12)	SOI values for the months of January to December 2000
	Output	Average precipitation for the year 2001
2	Prediction of average winter precipitation using preceding summer's ENSO/IOD values e.g.:	
	Input (6)	IOD values for the months of November 2000 to April 2001
	Output	Average precipitation for the months May to October 2001
3	Prediction of average summer precipitation using preceding winter's ENSO/IOD values e.g.:	
	Input (6)	Nino 3.4 values for the months of May to October 2000
	Output	Average precipitation for the months of November 2000 to April 2001
4	Prediction of average winter precipitation using preceding year's ENSO/IOD values e.g.:	
	Input (12)	SOI values for the months of January to December 2000
	Output	Average precipitation for the months of May to October 2001
5	Prediction of average summer precipitation using preceding year's ENSO/IOD values	
	Input (12)	Nino 3.4 values for the months of January to December 2000
	Output	Average precipitation for the months November 2000 to April 2001

For each test scenario, 33 to 35 years of data have been used for training, 7 years of data has been used for validation and 7 years of data have been used to test the neural network. The root mean squared errors (RMSE) obtained for the test set are shown in Table 3.

Table 6. The Root Mean Square Error (mm) obtained with test data.

Test	RMSE in precipitation prediction		
	IOD	**Nino 3.4**	**SOI**
1	22.9	12.5	21.1
2	23.4	23.7	21.0
3	34.7	25.0	28.4
4	20.9	23.1	19.5
5	35.2	22.4	30.8

Figure 15 shows that the results obtained when predicting the average winter precipitation are more accurate. This may be because of the fact that large scale circulation patterns seem to influence the precipitation mainly in winter. However, the accuracy of the results may also be because of the relatively low variability of precipitation in winter months. The best results have been obtained using Test 4, which is the prediction of average winter precipitation using the preceding year's SOI values as input.
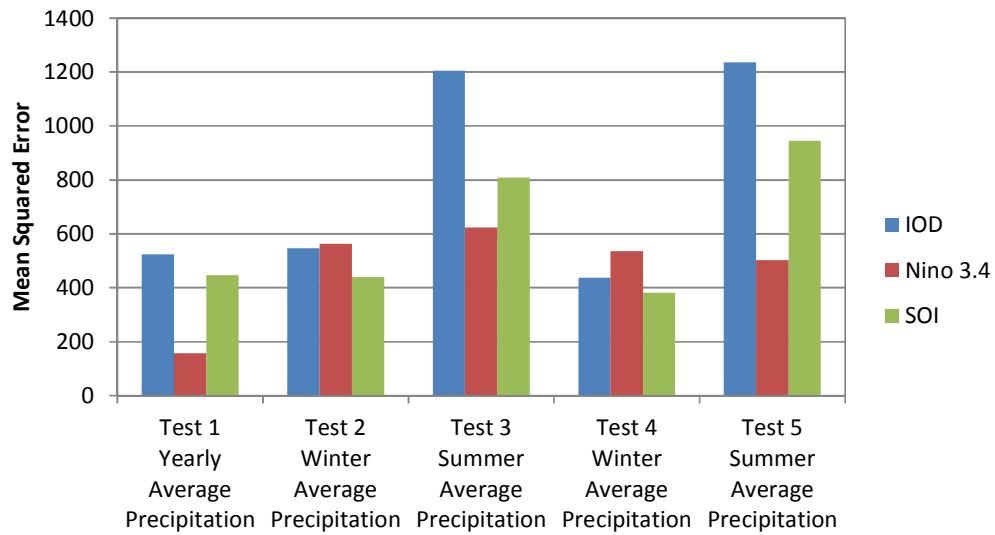
Figure 15. Mean squared error obtained using the different test scenarios.

Figure 16 shows the accuracy to which each year's average winter precipitation was predicted. The accuracies range from 74.8% to 96.2%, with an average of 86%.
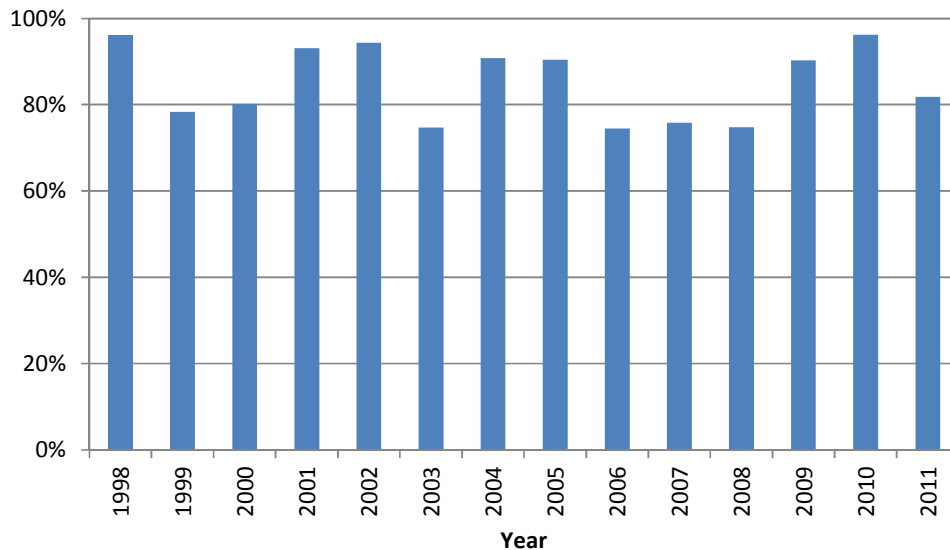
Figure 16. Percentage accuracy of average winter precipitation using SOI.

Based on the results obtained above, we tried to predict the monthly precipitation using SOI values as input (Table 7). The results for monthly precipitation are not as accurate as the results for the average winter precipitation (Table 8 and Figure 17). However, prediction is again more accurate in winter months.

Table 7. Test scenario 6

Test	Description	
6	Prediction of monthly precipitation using previous year's SOI values e.g.:	
	Input (12)	SOI values for the months January to December 2000
	Output	Precipitation for each month

Table 8. Results obtained for test scenario 6

Month	RMSE (mm)
January	71.1
February	108.6
March	162.7
April	60.8
May	37.0
June	22.1
July	26.6
August	36.9
September	64.3
October	64.8
November	21.9
December	59.4
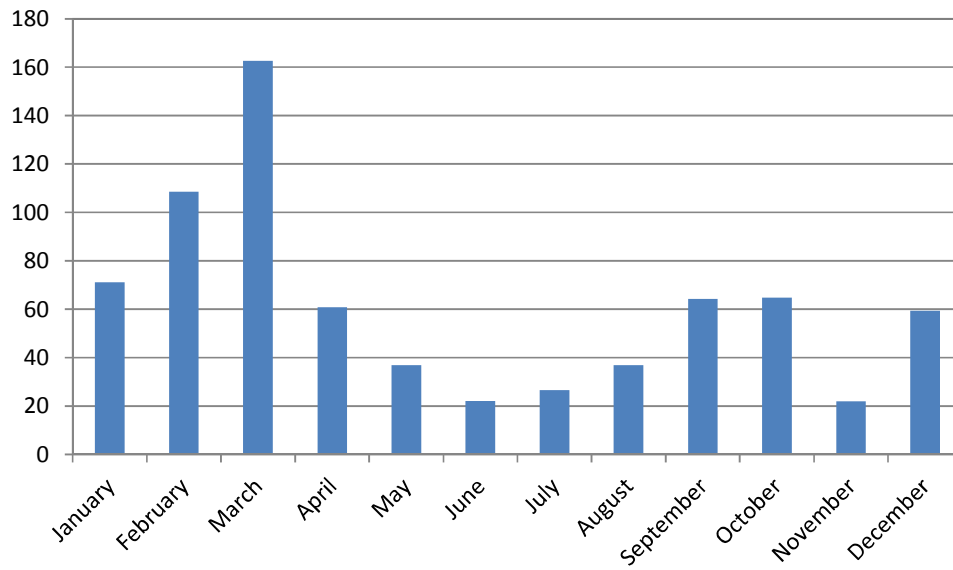
Figure 17. Monthly RMSE in precipitation prediction using SOI as input (Test 6).

4 TIME SERIES ANALYSIS AND FORECASTING

4.1 BACKGROUND

In this chapter, we build statistical and time-series models to establish relationships between precipitation and El Niño Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD). These models are then used to explain and predict precipitation patterns. Our main goal is to uncover short and long term trends, and also predict precipitation, given precipitation data of the past, and also large scale circulation indices such as Niño 3.4, Southern Oscillation Index (SOI) and IOD. Accurate rainfall prediction, in the short, medium and long terms, can be enormously helpful in managing water resources in Mauritius.

There have been numerous studies that relate drought and rainfall to large scale circulation phenomena, such as Southern Oscillation or El Niño [*Cordery and Yao*, 1993; *Pittock*, 1984; *Sun and Furbish*, 1997]. These studies cover various parts of the globe, from Florida to eastern Australia. Qualitative seasonal rainfall forecasts are also made based on the relation between rainfall and large-scale circulation indices, such as SOI. [*Cordery and Yao*, 1993], along with several other studies, establish that there is a fairly strong correlation between SOI and eastern Australian rainfall. *Sun and Furbish* [1997] show that El Niño and La Nina are responsible for up to 40% of annual precipitation variation in Florida.

It is quite usual for rainfall and large-scale circulation indices to have correlation coefficients as high as 0.7. However it has been shown that these relationships may not be consistent over time. *Cordery and Yao* [1993] show that although there are strong, consistent relationships between rainfall and SOI for some seasons at some locations in Australia, for other seasons the correlation coefficient of the relationship may vary from more than 0.6 for long periods down to zero and small negative values for other long periods. *Pittock* [1984] suggests that there is little doubt that a broad pattern of teleconnections exists, but that the influence of particular teleconnection mechanisms waxes and wanes from time to time at particular locations. He suggests that variations in the local spatial patterns are to be expected and these could be caused by statistical fluctuations or stochastic noise in the climatic system.

4.1.1 Time Series

In this work, we are dealing with time series data have a natural temporal ordering and seasonality. Besides yearly cycles, we can also look at the two seasons that are defined for Mauritius: six months of summer, extending from October to April, and six months of winter,

from April to September. In this chapter, we first perform some common time series *analysis* to extract meaningful statistics. We also look into time series *forecasting*, where various models are used to predict future values based on previously observed values.

We assume a stochastic model, where observations closer in time are more likely to be related; we adopt a time-domain method for time series analysis (as opposed to frequency-domain), and parametric approaches such as the autoregressive model. Our analysis covers 50 years, from 1961 to 2009. However, in light of observations made on the non-stationarity of the relationship between rainfall and large scale circulation, we also analyze the data for shorter-term time periods.

4.2 SUMMARY OF DATA

For the study in this chapter, we use monthly data for nearly 50 years (from 1961 to 2009). The variables we use are: precipitation (in mm) at Vacoas, Niño 3.4, SOI, and IOD.

4.2.1 Precipitation

Precipitation data between 1961 and 1990 was obtained from the National Climate Data Center (NCDC), which is part of the Global Historical Climatology Network (GHCN). Data between 1991 and 2009 was obtained from the Mauritius Meteorological Services.

4.2.2 Niño 3.4

The Niño 3.4 index is the area-averaged sea surface temperature anomaly (SSTA) over the region bounded by 5°N–5°S and 120°W–170°W. Monthly data for the Niño 3.4 index was obtained from the International Research Institute (IRI) in the Climate and Society Data Library.

4.2.3 Southern Oscillation Index (SOI)

SOI is an index that shows the variation in the Southern Oscillation. It is the difference in mean sea level pressure anomalies at Tahiti and Darwin, Australia. Data for the SOI was obtained from the National Center for Atmospheric Research (NCAR) and Climate and Global Dynamics (CGD).

4.2.4 Indian Ocean Dipole (IOD)

IOD is represented by anomalous SST gradient between the western equatorial Indian Ocean (50°E-70°E and 10°S-10°N) and the south eastern equatorial Indian Ocean (90°E-110°E and

10°S-0°N). The gradient is known as the Dipole Mode Index (DMI) and has been obtained from the Japan Agency for Marine-Earth Science and Technology (JAMSTEC).

4.3   PRELIMINARY ANALYSIS

Table 9 shows some key statistics for the data series that spans 584 months between January 1961 and August 2009. Figure 18 show the time plots for precipitation, Niño 3.4, SOI and IOD (along with a horizontal lines to indicate the long-term mean).

Table 9. Summary statistics

Key Variable	Mean	Std. Dev.	Minimum	Maximum	Augmented Dickey-Fuller Statistic
Precipitation (mm)	172.9	162.1	11.5	1362.0	-17.82
Niño 3.4	0.1068	0.8604	-2.3298	2.6707	-3.97
SOI	-0.2731	1.8222	-7.6000	4.7000	-11.34
IOD	0.0277	1.0074	-2.7351	3.5479	-3.99
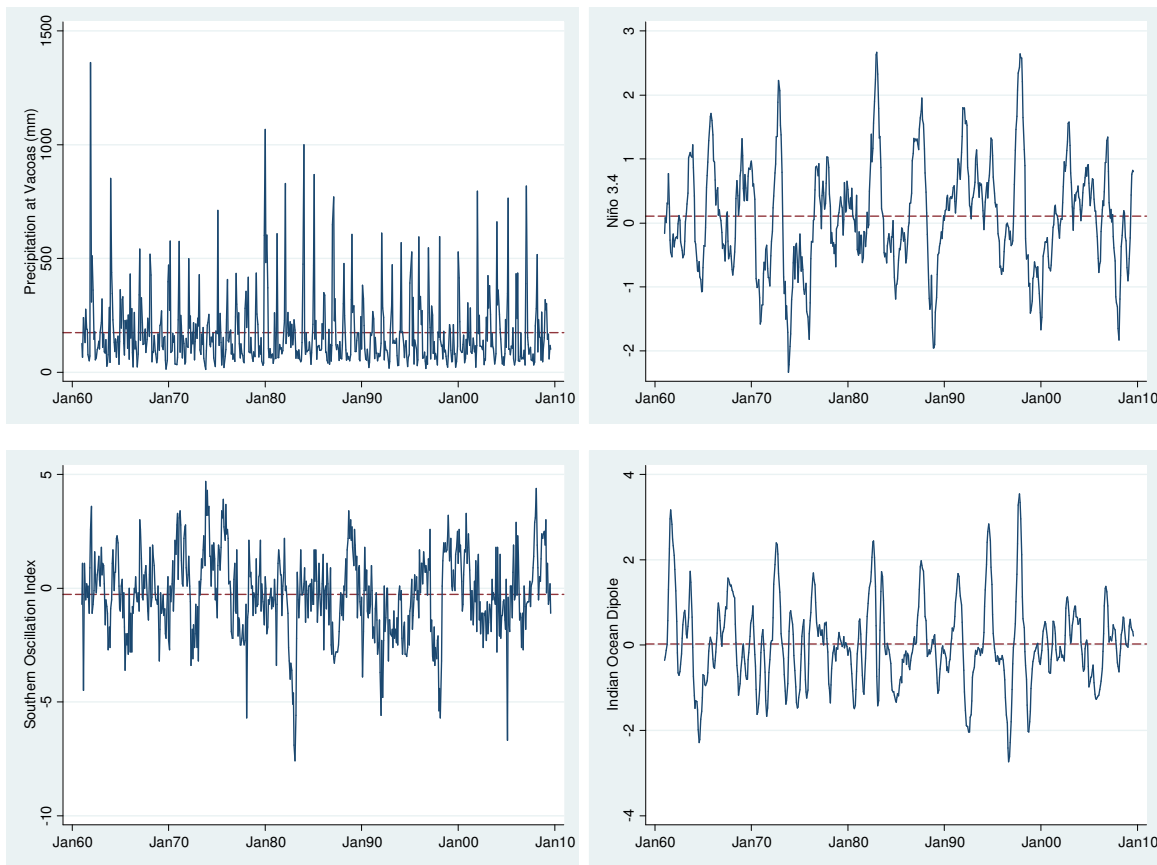
Figure 18. Variation of precipitation at Vacoas (in mm), Niño 3.4, SOI, and IOD with time

Long periods of below the long-term mean precipitation and short periods of high precipitation is clearly visible. For the period of record plotted, the maximum precipitation recorded is 1,362 mm in December 1961. The minimum monthly precipitation recorded is 11.5 mm in November 1973. The coefficient of variation $C_v = \sigma/\mu$ (where σ is the standard deviation and μ is the mean), is a statistical measure of variability, where $C_v < 1$ indicates less variation, while $C_v > 1$ indicates high variability. The coefficient of variation for precipitation is 0.94.

4.3.1 Stationarity Tests

It can be observed that precipitation, and the three teleconnection indices, are highly variable around the mean. We first determine if the data is stationary (i.e. the statistical properties of the time-series, such as mean, variance, autocorrelation, etc. are all constant) because most statistical forecasting methods are based on the assumption that the time series is stationary, or can be rendered approximately stationary (i.e. "stationarized") through the use of mathematical transformations. The four data series are, in fact, stationary, based on the augmented Dickey–

Fuller test for a unit root in the time series sample. The augmented Dickey–Fuller (ADF) statistic (Table 9) used in this test is a negative number. The more negative it is, the stronger is the rejection of the hypothesis that there is a unit root at some level of confidence. The test statistic for all four time series data shows with high confidence (better than 1% significance level) that the series are stationary, especially precipitation and SOI.

Table 10 breaks down the analysis of the data into four decades. It can be observed that in the period 1991-2000, the mean precipitation is much lower than the long-term mean. Also, for the same period, the mean SOI has the highest negative value and the mean Niño 3.4 has the largest positive value over all the four decades. This seems to indicate a correlation between precipitation and Niño 3.4/SOI for that period only. We shall further investigate how the correlation varies with later in the chapter.

Table 10. Mean of key variables over 50 and 5 x 10-year periods

| | Long-term mean | Mean over 10-year periods | | | |
	1961-2009	1961-1970	1971-1980	1981-1990	1991-2000	2001-2009
Precipitation	172.9	186.3	174.1	177.4	154.3	172.2
Niño 3.4	0.107	0.116	-0.161	0.140	0.280	0.168
SOI	-0.273	-0.153	0.203	-0.552	-0.674	-0.177
IOD	0.028	0.094	-0.035	0.046	-0.004	0.038

Figure 19 shows the variation of mean precipitation over time. Here, a moving window of five years is used to track the moving average. The average precipitation fluctuates quite significantly with time, although a slight downward trend is noticeable. However, the 50-year period is too small to ascertain any long-term trend, and here we assume that precipitation is stationary.
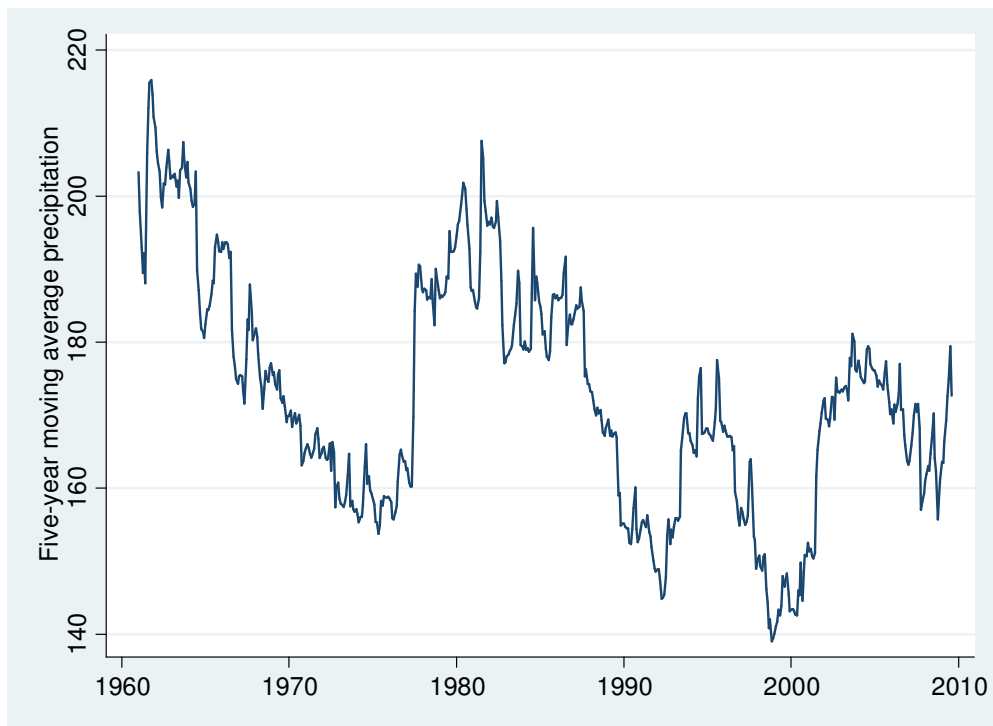
Figure 19. Moving average over five years of precipitation (in mm) at Vacoas.

4.3.2 Correlation between Precipitation & Large Scale Circulation

Table 11 shows various pairwise correlations over a 50-year period among the variables used in this study. These correlations are calculated by comparing the variables for the same months i.e. without considering any lag.

We observe that there seems to be practically no correlation between precipitation and the large scale pattern indices in the long term. On the other hand, indices for the large-scale patterns are fairly well correlated (Figure 20 to Figure 22), especially Niño 3.4 and SOI. This is expected; Niño 3.4 and SOI capture the same phenomenon, and a strong (inverse) correlation is expected. IOD captures information on a different large scale circulation pattern, and it is interesting to see some limited correlation between Niño 3.4 and IOD.

Table 11. Correlation coefficients

	Niño 3.4	SOI	IOD
Precipitation	-0.025	-0.002	0.035
Niño 3.4		-0.733	0.378
SOI			-0.275
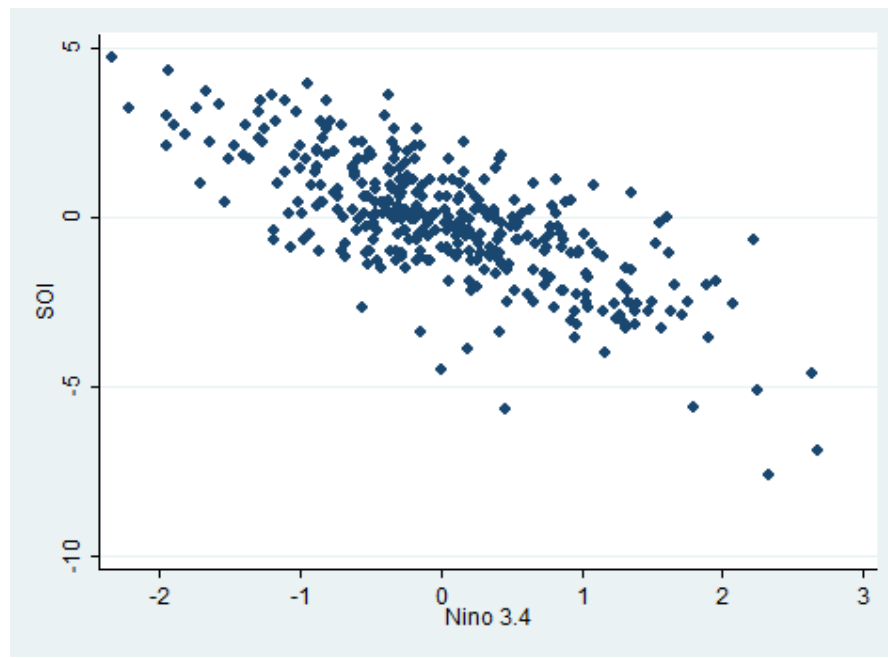
Figure 20. Scatter plot of SOI and Niño 3.4, showing negative correlation between the two (correlation coefficient = -0.7333).
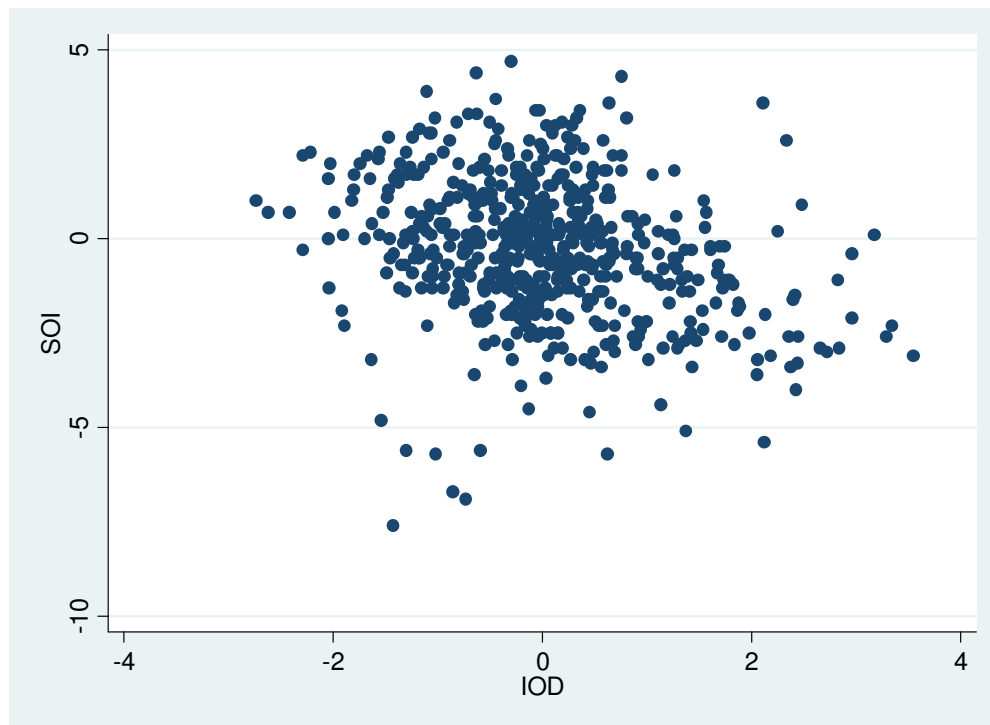
Figure 21. Scatter plot of SOI and IOD, showing slight negative correlation between the two (correlation coefficient = -0.2747).
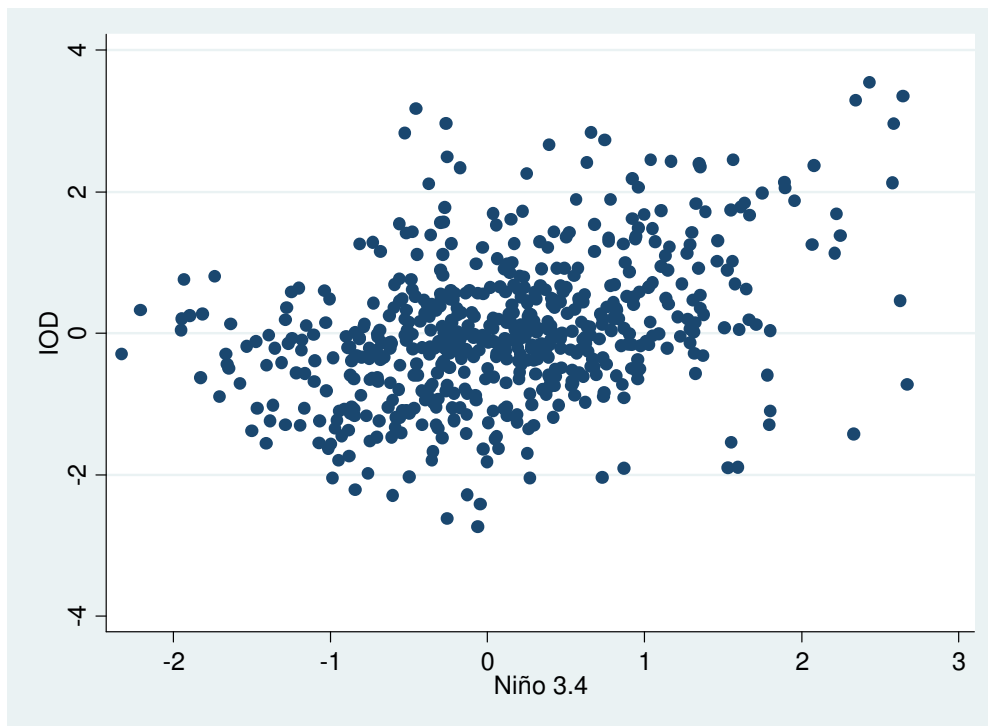
Figure 22. Scatter plot of Niño 3.4 and IOD, showing slight positive correlation between the two (correlation coefficient = 0.3781).

4.3.3 Seasonal Correlations

Our analysis so far has not revealed any significant correlation between precipitation and indices for telecommunication patterns. In this section, relationships between rainfall and large scale circulation indices are examined on a seasonal basis. Seasonal data is used because it gives better correlations, and also because forecasting of precipitation may be possible with lagged relations (precipitation lagging behind El Niño, as we are inclined to expect).

Mauritius has two seasons: winter and summer. Winter is from May to October and summer is from November to April. November is also the beginning of the hydrological year. In order to assess the influence of ENSO on precipitation at Vacoas, the precipitation, Niño 3.4, SOI and IOD series were averaged (over six months) for the two seasons, winter and summer. Table 3.4 reveals much stronger correlations among the variables, remarkably more in winter than in summer.

Table 12. Correlations for summer and winter

	Precipitation-Niño 3.4	Precipitation-SOI	Precipitation-IOD	Niño 3.4-SOI	Niño 3.4-IOD	SOI-IOD
Summer	0.075	-0.079	0.158	-0.923	0.231	-0.062
Winter	0.161	-0.174	0.307	-0.893	0.549	-0.607

Figure 23 shows how the correlations between precipitation and the three indices (for winter) vary with time for a moving window of 10 years between 1960 and 2010. The correlation precipitation and SOI is the strongest, although unstable. It is mostly negative, and the closest correlation occurs with a coefficient of -.99. The correlation between precipitation and both Niño 3.4 and IOD are predominantly positive, as expected. But there is a lot of variation, with the maximum correlation coefficient for a 10 year period is 0.78 for IOD and 0.70 for Niño 3.4.
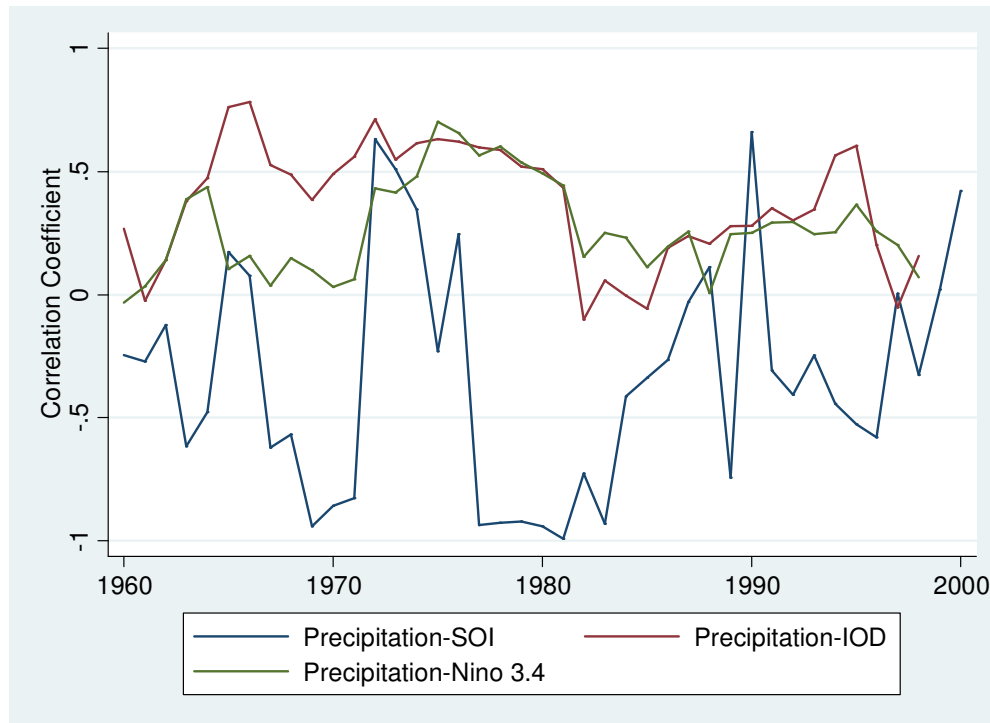
Figure 23. Correlation coefficient between precipitation and teleconnection indices (for moving 10-year window)

Time series data can typically be decomposed into several components: trend, seasonal, cyclical and an idiosyncratic component. The idiosyncratic portion of a series represents all those factors other than trend, seasonal, and cyclical. Intuitively, the idiosyncratic component is like an error term, or the rough part of the series. Data smoothing is commonly so that the trend, seasonal, and cyclical components can be identified. In single moving average smoothing methods, the past observations are weighted equally. Exponential smoothing assigns exponentially decreasing weights as the observations get older. In other words, recent observations are given relatively more weight in forecasting than the older observations. In exponential smoothing, there are one or more smoothing parameters to be determined (or estimated) and these choices determine the weights assigned to the observations.

In this section, we use the triple exponential smoothing method for rainfall forecasting. This method takes into account seasonal changes as well as trends, as proposed by Charles Holt and Peter Winters (Holtz 1957; Winters 1960). The method calculates a trend line for the data as well as seasonal indices that weight the values in the trend line based on where that time point falls in the seasonal cycle. In essence, the method minimizes the root mean square error by finding optimal values for three parameters: the data smoothing factor, the trend smoothing factor the seasonal change smoothing factor.

Figure 24 shows the actual and predicted values of monthly precipitation using Holt-Winters exponential smoothing. The root mean squared error obtained with this method is 129.2 mm.
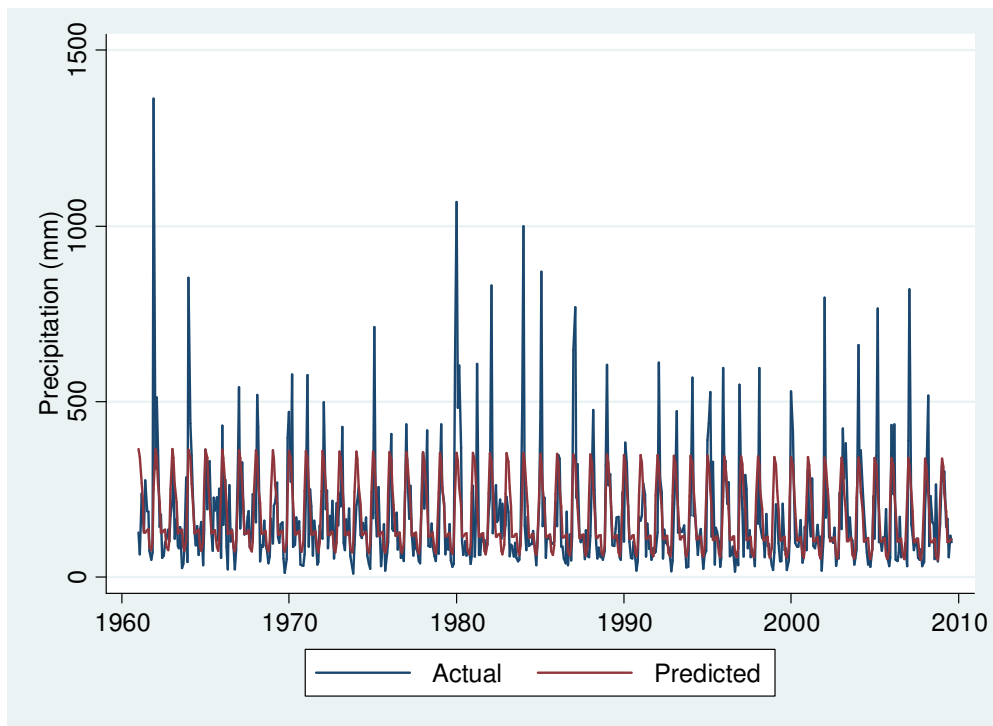
Figure 24. Actual and predicted precipitation using Holt-Winters exponential smoothing (root mean square error = 129.2mm)

4.5 REGRESSION MODELS

We now discuss regression models that were used for prediction and forecasting. The dependent variable is precipitation and the independent variables are lagged precipitation and indices for large scale circulation (Niño 3.4, SOI and IOD). Regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables - that is, the average value of the dependent variable when the independent variables are fixed.

Autoregressive moving average (ARMA) models provide a description of a (weakly) stationary stochastic process in terms of two polynomials, one for the auto-regression and the second for moving averages. Given a time series of data, the ARMA model is a tool for understanding and, perhaps, predicting future values in this series. The model consists of two parts, an autoregressive (AR) part and a moving average (MA) part [*Box et al.*, 2008),].

An autoregressive integrated moving average (ARIMA) model is a generalization of an ARMA model. These models are fitted to time series data either to better understand the data or to predict future points in the series. ARIMA models form an important part of the Box-Jenkins

approach to time-series modeling. We use this model to see if the precipitation can be predicted based only on data of previous years, and then improved by introducing the three teleconnection indices.

Figure 25 shows the autocorrelation of monthly precipitation for lags from 1 month to 15 months. The lengths of the line are indications of the strength of the autocorrelation at various lags. There is strong autocorrelation as well as a pattern over lag, which suggests that the autocorrelations are not random.
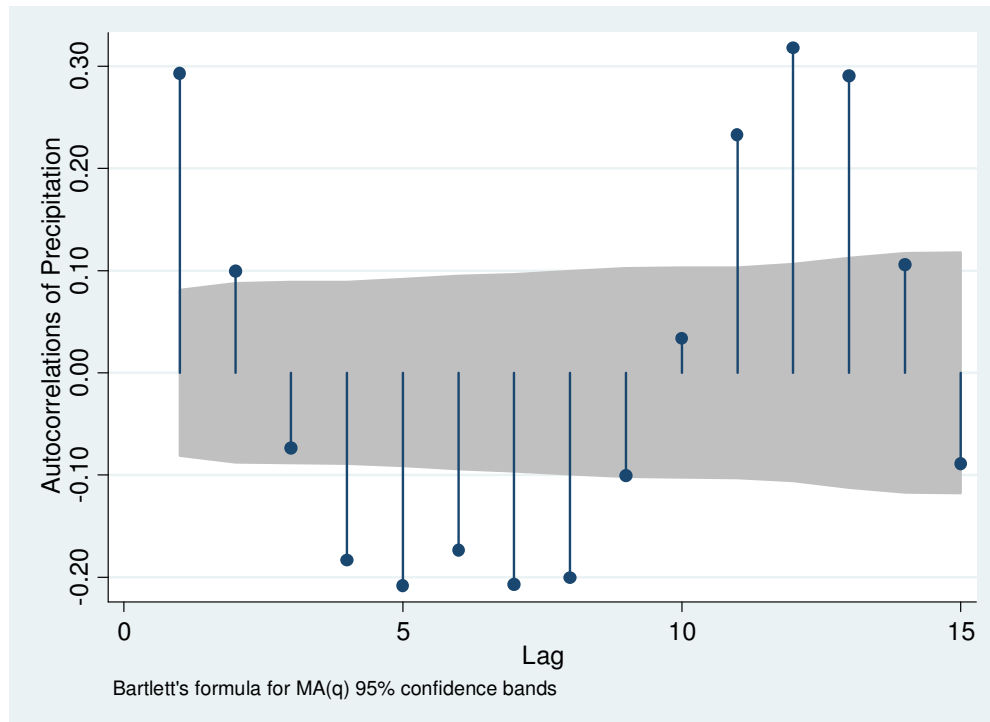
Figure 25. Autocorrelations of precipitation at Vacoas with different lags (in months)

Figure 26 and Figure 27 compare actual and predicted precipitation, based on a simple linear regression and an ARIMA model respectively, where precipitation is the dependent and lagged precipitation as well as the large scale circulation indices are independent variables. The root mean square error is 129 mm and 136.8 mm respectively.
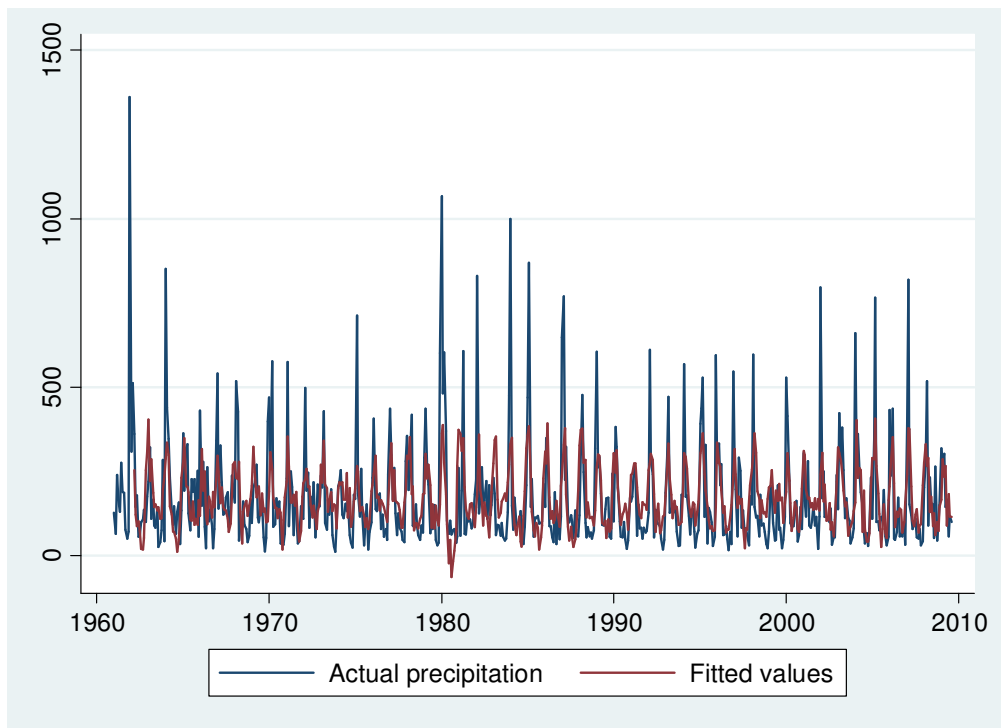
Figure 26. Predicted vs. actual monthly precipitation at Vacoas (in mm) based on linear regression (Root mean square error = 129mm)
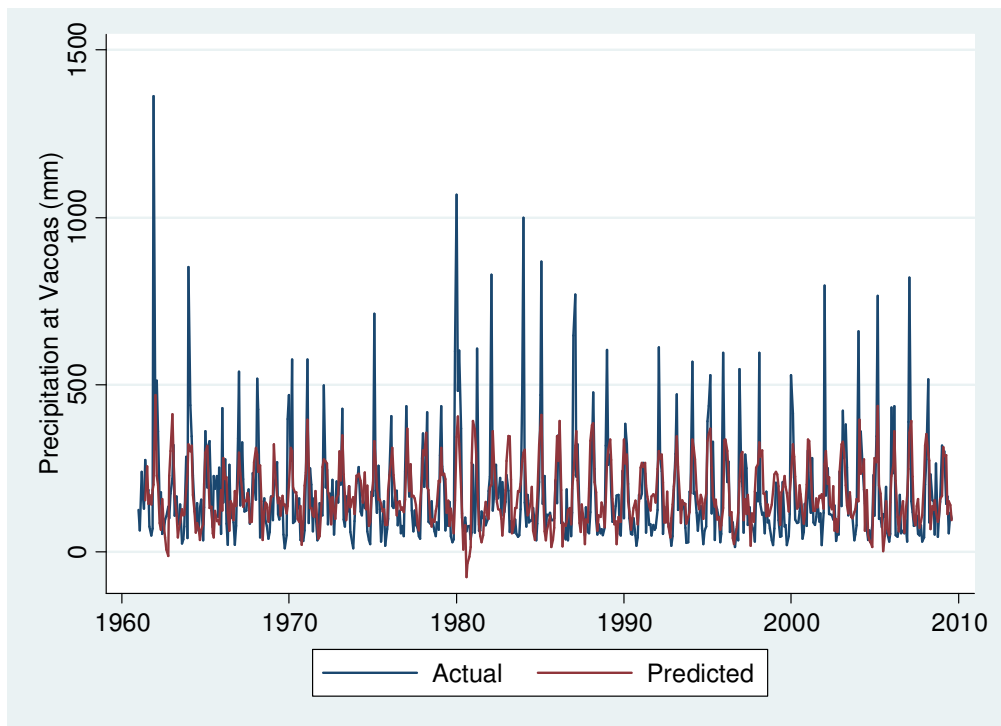
Figure 27. Actual vs. predicted precipitation based on an ARIMA model (Root mean square error = 136.8 mm)

41

There are few observations that we would like to highlight:

For both the simple linear regression and ARIMA models, we used lagged precipitation (up to 12 months) with and without factoring in Niño 3.4, IOD and SOI indices. The latter indices improves the monthly prediction, albeit by a small percentage (5-10%), but consitently so. Here we take the average prediction values independent of the month and season. As we shall see later, the precdiction are much more accurate in winter than in summer.

We tested the optimal lags for the three indices: our results indicate that for both Niño 3.4 and SOI, the last three months of data seem relevant in determining the current precipitation value. For IOD, the last five months of data gives the best prediction of current month's precipitation.

As previous results indicate, the accuracy of prediction varies largely from month to month. Figure 28 shows the root mean square error values for each month. There is a very clear trend of incrementally decreasing error to a minimum at the peak of winter (in June-July) and increasing to a maximum at the peak of summer (in January-February). It is interesting to see that both the correlation and the ability to predict is heavily dependent on the month of the year.

Figure 28. Monthly root mean square error in prediction of precipitation based on regression methods. Rainfall is much more predictable in winter than in summer.

5 DROUGHT ANALYSIS

5.1 CHOOSING A DROUGHT INDEX

Droughts in Mauritius, at least in the public's perception, are identified when the water level in the major reservoirs fall below the long term mean. This is probably the simplest and easiest way to identify droughts, but it is not necessarily the most effective way, as there is no forewarning mechanism, except for the noticeable decline in water level when r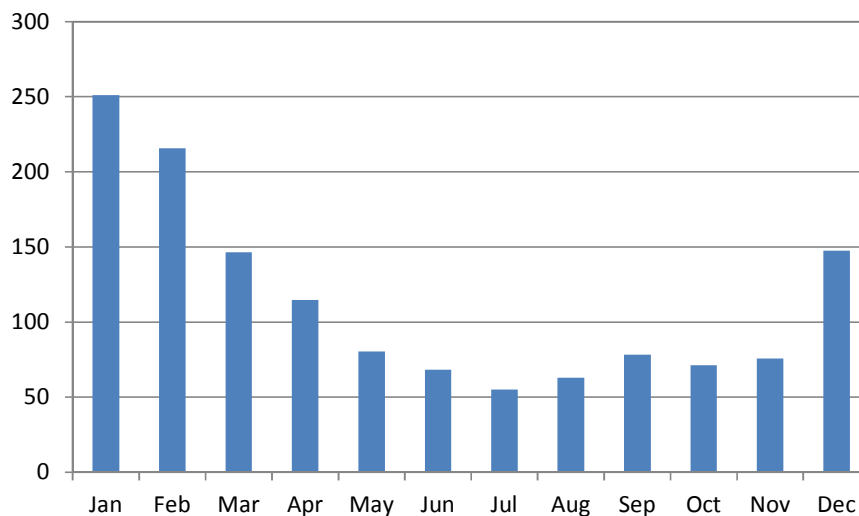ecords are examined. Another possible approach is to monitor rainfall, and the lack thereof may provide a warning, on the possible state of future water availability.

Droughts, however, are not confined to rainfall deficits alone. Numerous definitions for drought exist and each is more adapted to one particular field than another. A lack of precipitation leads to a meteorological drought, which is defined as the degree of dryness, with respect to a normal or average condition, and the duration of the dry event. A meteorological drought is the onset of a drought event. Agricultural droughts follow meteorological droughts and are linked to precipitation deficits, soil water deficits, and decline in groundwater levels etc. A hydrological drought is the result of an extended period of rainfall deficit which translate into low flows in streams and low water levels in reservoirs, which may subsequently have a number of socio-economic implications.

A suite of indices (e.g. Palmer Drought Severity Index, Palmer Hydrological Index, Standardized Precipitation Index, Statistical Z-score, Effective Drought Index, etc.) have been developed for drought identification. Some of based on a single hydro-meteorological parameter, while others may incorporate more than one parameter, and can be a more efficient index for drought assessment. It is imperative that a study on assessing the most important parameter(s) for drought identification be conducted for Mauritius in order to determine which index may be most appropriate for the local context.

Given that in this study, data is scarce, we will use the Standardized Precipitation Index (SPI), which is based on precipitation only.

5.2 STANDARDIZED PRECIPITATION INDEX

SPI was proposed by *McKee et al.* [1993]. The motivation behind this method was to have an index that can be compared across different climatic regions. It is based solely on precipitation and gives a better representation of abnormal wetness and dryness than the Palmer Drought Severity Index, for example.

The SPI has been designed to be a spatially invariant quantity that can be computed to give precipitation deficit at multiple timescales [*Guttman*, 1999]. Computing the index for short and long timescales highlights the impacts of drought on various water resources. Runoff and reservoir storage respond much faster to a storm event than soil moisture, groundwater level. SPI are normally calculated for 3, 6, 9, 12, 24, and 48 month time scales.

SPI calculation is based on long-term precipitation record. A probability density function that describes the long term precipitation observation is determined. *Guttman* [1999] explains that the use of different probability distribution will lead to different SPI values. The chosen distribution is then transformed to a normal distribution so that the mean SPI for the location and desired period is zero. Normalizing allows for a consistent representation of wet and dry periods. A positive SPI value indicates precipitation greater than the mean and negative values indicate lower than the mean. A classification system defining drought intensities, developed by *McKee et al.* [1993], is given in Table 13 below. Continuous negative SPI value indicates drought events, which end when the SPI becomes positive. The durations, severities, and inter-arrival times of drought events can thus be determined from a plot of the SPI values.

Table 13. Interpretation of SPI values

		SPI	$\leq$	-2.0	Extremely dry
-2.0	$<$	SPI	$\leq$	-1.5	Moderately dry
-1.5	$<$	SPI	$\leq$	-1.0	Dry
-1.5	$<$	SPI	$<$	1.0	Neutral
1.0	$\leq$	SPI	$<$	1.5	Wet
1.5	$\leq$	SPI	$<$	2.0	Moderately wet
2.0	$\leq$	SPI			Extremely wet

5.2.1 Computing SPI

SPI calculation described below follows the Weekly SPI User's Manual (available from http://greenleaf.unl.edu/downloads/SPI_Manual.pdf). SPI is computed by fitting a probability density function to long term precipitation record and normalizing the values obtained. In the present study the 2-parameter gamma distribution will be used. The parameters α and β are referred to as the shape and scale parameters respectively. The probability density function of the gamma distribution is given below:

$$g(x) = \frac{1}{\beta\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$$

The parameters thus obtained are used to derive the cumulative density function

$$G(x) = \int_0^x g(x)dx = \frac{1}{\beta\Gamma(\alpha)} \int_0^x x^{\alpha-1} e^{-\frac{x}{\beta}} dx$$

Letting $t = \frac{x}{\beta}$, the equation becomes an incomplete gamma function:

$$G(x) = \frac{1}{\Gamma(\alpha)} \int_0^t t^{\alpha-1} e^{-t} dt$$

The gamma function is undefined at $x = 0$. Since precipitation can be equal to zero, the cumulative density function becomes:

$$H(x) = q + (1-q)G(x)$$

where q is the probability of a zero. The cumulative density function is then transformed to a standard normal random variable Z with mean zero and variance of one, which is the value of the SPI.

$$Z = SPI = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \quad for \quad 0 \le H(x) \le 0.5$$

$$Z = SPI = +\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \quad for \quad 0.5 \le H(x) \le 1.0$$

where

$$t = \sqrt{\ln\left(\frac{1}{(H(x))^2}\right)} \quad for \quad 0 \le H(x) \le 0.5$$

$$t = \sqrt{\ln\left(\frac{1}{(1.0 + H(x))^2}\right)} \quad for \quad 0.5 \le H(x) \le 1.0$$

where

$$c_0 = 2.515517 \qquad\qquad d_1 = 1.432788$$

$$c_1 = 0.802853 \qquad\qquad d_2 = 0.189269$$

$$c_2 = 0.010328 \qquad\qquad d_3 = 0.001308$$

5.3   SPI(6) AT VACOAS

SPI for six months [SPI(6)] was calculated for the Vacoas station. The plot of SPI, overlain on monthly precipitation anomaly for the period January 1961 to September 2012 is given in Figure 29. Anomaly for month i is given as precipitation for month i and year j minus mean for month i. Positive (negative) anomalies, which represent above (below) average rainfall are shown in blue (red). It can be seen that precipitation at Vacoas cycles through positive and negative periods and the SPI(6) values closely follow the precipitation pattern.
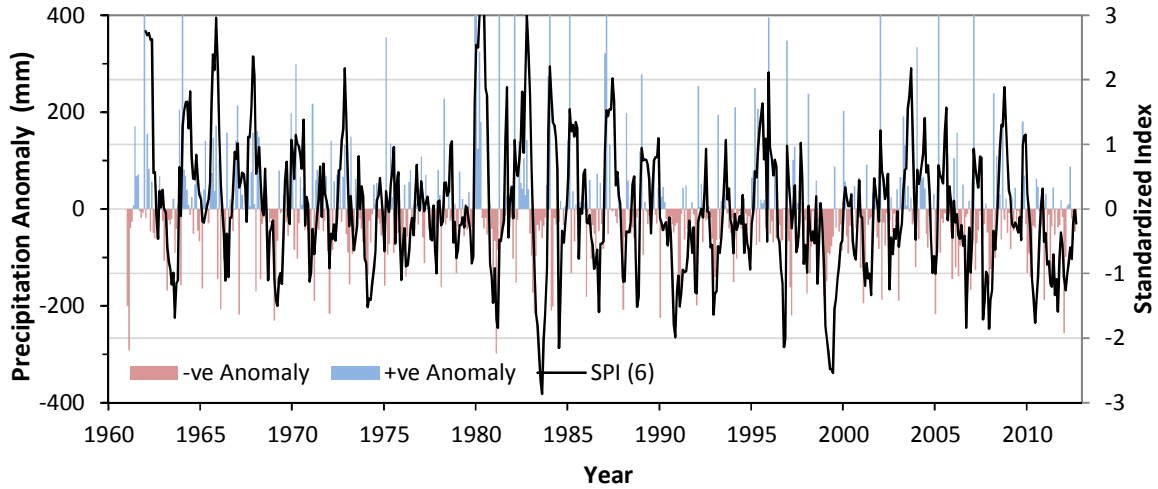
Figure 29. Plot of monthly rainfall anomaly and SPI(6) at Vacoas

5.3.1 Identifying Drought Events

The following information can be extracted from the SPI(6) plot, as illustrated in Figure 30:

- Drought duration

- Drought severity (area under, below the chosen SPI threshold)

- Inter-arrival time which can be arbitrarily defined as the duration between successive droughts or the time between the start of successive drought events. In this study we adopt the former approach.

Figure 30. Drought duration, severity, and interarrival time for drought events from SPI(6)

5.3.2 Drought Duration, Severity and Inter-arrival Time

For the purpose of this study, we assume that a drought event occurs when SPI(6) ≤ −1, and ends when SPI(6) > −1. Thirty four drought events, lasting a total of 88 months were recorded. The basic statistics of each variable (drought durations, severities, and inter-arrival times) thus extracted, from SPI(6) for Vacaos, is given in Table 14.
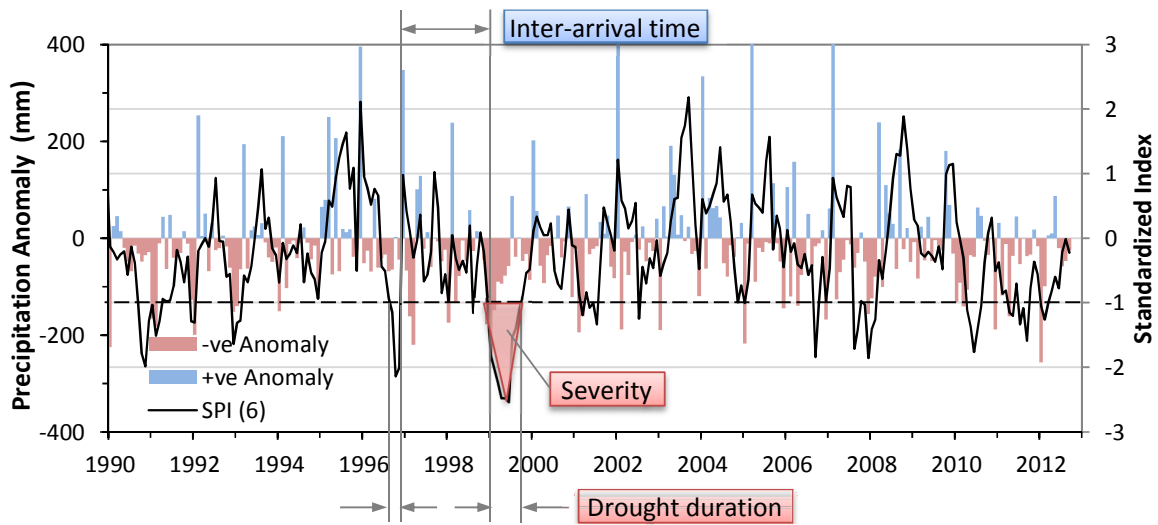
The mean duration of droughts for the period January 1962 to September 2012 is 2.6 months. The minimum drought duration is 1 month, while the maximum is 9 months. This event occurred in January 1990 and lasted until May 1992, and the associated severity is 17.6. The average drought severity is 3.8. The mean drought inter-arrival time is 15.2 months, and the maximum period without a drought is 58 months.

Table 14. Statistics for duration, severity, and inter-arrival time extracted from SPI(6)

	Duration (months)	*Severity[a]*	*Inter-arrival Time (months)*
Number of events	34	34	34
Total	88	129.119	515
Minimum	1	1.003	1
Maximum	9	17.581	58
Average	2.588	3.798	15.147
Standard deviation	1.956	3.746	13.513
Skewness	1.757	2.468	1.189

5.3.3 Marginals for Drought Duration, Severity

The marginal for the drought duration and severity was chosen from a suite of probability distribution functions commonly employed in hydrology. Both graphical and analytical goodness-of-fit tests were used to discriminate between distributions. Duration was found to follow a gamma distribution and severity followed an exponential distribution. The pdf for gamma and exponential is given as:

$$f_d(d|\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} d^{\alpha-1} e^{-d/\beta} \qquad 0 < d < \infty, \qquad \alpha,\beta > 0 \tag{3}$$

$$f_s(s|\lambda) = \frac{1}{\lambda} e^{-s/\lambda} \qquad 0 \le s < \infty, \qquad \lambda > 0 \tag{4}$$

where d is the duration, s is the severity, α and β are the shape and scale parameters of the gamma distribution, respectively, and λ is the parameter for the exponential distribution. The parameters for each distribution were estimated using maximum log-likelihood. The fitted marginal for duration and severity is shown in Figure 31.
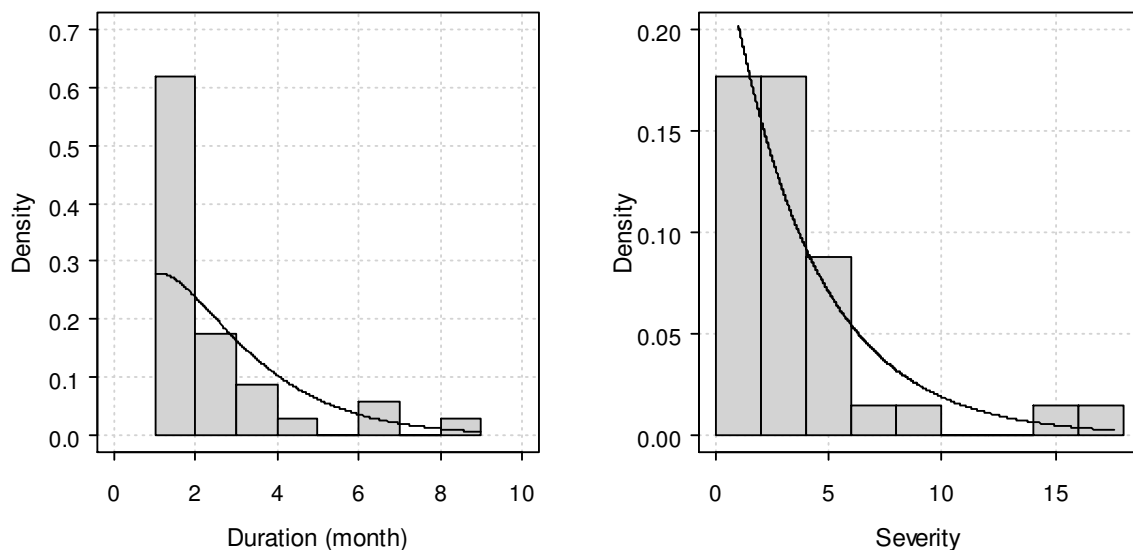

Figure 31. Marginal distributions for duration and severity

5.4 Copula Selection

Copula, which is due to *Sklar* [1959], was used to model the dependence between duration and severity. It allows the combination of different marginal families into a joint distribution. It states that the joint distribution of any randomly distributed variables (X, Y) may be written as

$$H(x, y) = C[F(x), G(y)] \qquad x, y \in R \tag{5}$$

where $F(x)$ and $G(y)$ are the marginal probability distributions and $C = [0,1] \times [0,1] \rightarrow [0,1]^2$, the mapping function is the copula. This implies that a valid probabilistic model for (X, Y) may be obtained when the three components, F, G, and C, are from the following parametric families:

$$F(x; \boldsymbol{\delta}), \qquad G(y; \boldsymbol{\eta}), \qquad C(u, v; \boldsymbol{\theta}) \tag{6}$$

where $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$ are parameter vectors of the marginal distributions and $\boldsymbol{\theta}$ is the parameter vector for the dependence structure. u and v are the quantiles of the uniformly distributed variables $U = F(X)$ and $V = G(Y)$, respectively [*Chowdhary et al.*, 2011]. A number of copula families are available and are categorized into four classes: Archimedean, extreme value, elliptical, and other miscellaneous classes.

The most appropriate copula for modeling the dependence of drought duration and severity at Vacoas was chosen from the following copulas: Gaussian, Student t, Clayton, Gumbel, and Frank. The suitability of the copula was determined from a set of graphical and analytical tests. Gumbel was found to be most suitable for this exercise. Figure 32 gives a plot of the severity versus duration values simulated from the chosen copula overlain with observed ones. It can be seen that the chosen copula is able to simulate the complete range of observations, without any bias at the tails.
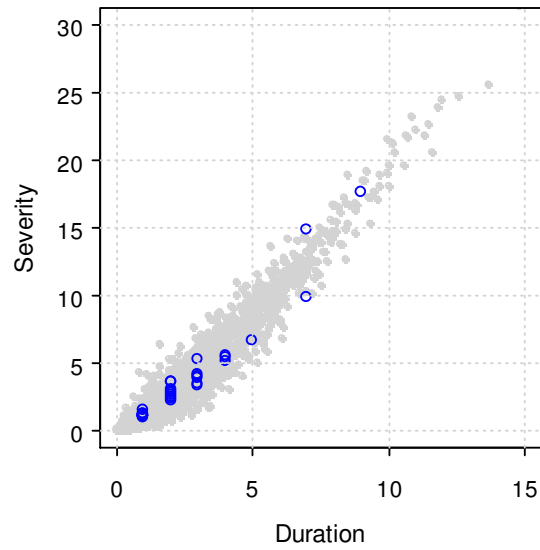
Figure 32. Plot of simulated values (gray) and observed values (blue) for Vacoas.

5.5 CONDITIONAL DISTRIBUTION FOR DROUGHT DURATION AND SEVERITY

Using the chosen copula and associated dependence parameter θ, the conditional distribution can be derived. Following *Shiau* [2006], the probability that duration and severity exceed a certain threshold can be expressed as

$$
\begin{aligned}
P(D \geq d, S \geq s) &= 1 - F_D(d) - F_S(s) + F_{D,S}(d, s) \\
&= 1 - F_D(d) - F_S(s) + C\big(F_D(d), F_S(s)\big)
\end{aligned}
\tag{7}
$$

where $F_D(d)$ and $F_S(s)$ are the cumulative duration and severity frequency distributions, respectively. The conditional drought severity distributions given a drought duration exceeding a certain threshold d', and the conditional drought duration given that drought severity exceeds a certain threshold s', are

$$
P(S \leq s | D \geq d') = \frac{F_S(s) - C\big(F_D(d'), F_S(s)\big)}{1 - F_D(d')}
\tag{8}
$$

$$
P(D \leq d | S \geq s') = \frac{F_D(d) - C\big(F_D(d), F_S(s')\big)}{1 - F_S(s')}
\tag{9}
$$

50

respectively.

The conditional distribution of severity given duration exceeding d' and conditional distribution of duration given severity exceeding s' is given in Figure 33 and Figure 34 respectively. This information can be used for more efficient water resources planning and management.
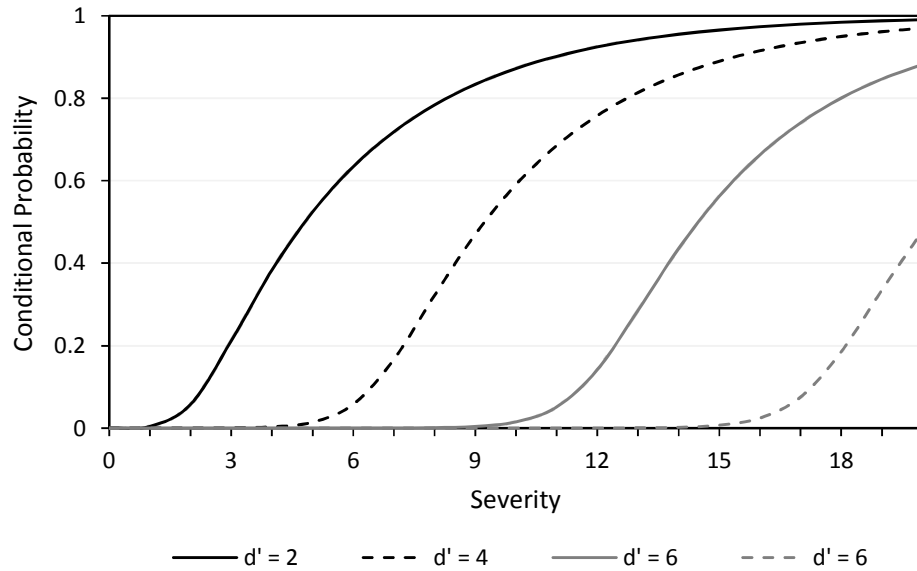

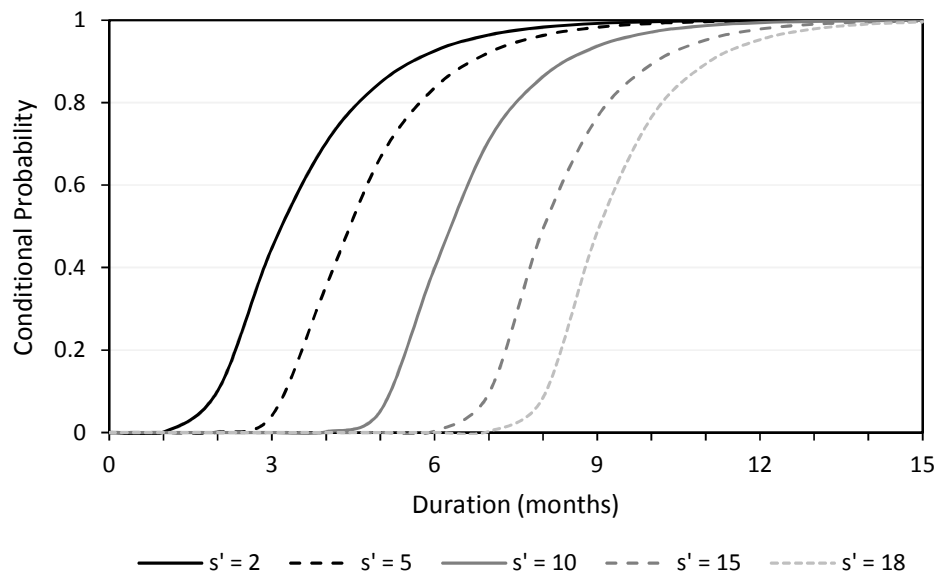**Figure 33. Conditional distribution of Severity given Duration exceeding *d'***

Figure 34. Conditional distribution of Duration given Severity exceeding *s'*

6   DISCUSSION

1.   Mauritius is a water stressed country and is quickly slipping into the water scarce category, as demand for water is growing rapidly. Water availability and allocation is one of the major challenges the country faces. The island does not have enough carry-over capacity i.e. most of the water received during a hydrologic year is used within the year itself. Thus, a shortage of rainfall can have immediate and disastrous effects on the well-being of the population and various economic sectors of the island.

2.   The ability to accurately predict precipitation and drought conditions is a critical factor in the short, medium, and long term water resources management.

3.   Precipitation around the world is known to be influenced by large-scale circulation patterns. In Mauritius, two major teleconnection patterns that are likely to affect rainfall are the El Niño Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD). In this study, our focus was to assess the influence of these circulation patterns on precipitation at Vacoas. The study can be extended to cover the whole island.

4.   The indices we have used are the Niño 3.4 index, the Southern Oscillation Index (SOI), and the Indian Ocean Dipole index for the period 1961 to 2012. Based on the Augmented Dickey-Fuller test, the time series for precipitation and three climate indices are deemed stationary.

5.   Our initial analysis revealed very poor correlation between precipitation and these climate indices. However, when the data series were split into two separate seasons (winter and summer), higher correlations were obtained for winter. For IOD, the correlation for winter is statistically significant at 5% confidence.

6.   Furthermore, we found, using a 10-year moving average window, that the correlations vary markedly with time. It fluctuates from very high correlation (above 0.9) to negligible correlation. This non-stationarity in the correlation between precipitation and climate indices has been in studies pertaining to precipitation in Australia [*Cordery and Yao*, 1993].

7.   Our next aim was to predict precipitation using these teleconnection indices as predictor variables. With artificial neural networks (ANNs), we obtained an average winter precipitation prediction accuracy of 86%. However, prediction of summer precipitation was less accurate than winter precipitation. We also predicted yearly precipitation and

monthly precipitation prediction with reasonable accuracy. Furthermore, monthly precipitation prediction was less accurate than average precipitation prediction for winter.

8. Results obtained from ANN were more accurate than those from other statistical techniques, such as linear regression and autoregressive integrated moving average (ARIMA). This may be attributed to the ability of ANNs in capturing the non-linearity in the system. We should point out that this analysis requires further investigation, and a comprehensive set of tests covering various statistical techniques with different parameters.

9. Both ANNs and regression based models predict winter precipitation with remarkably higher precision than summer precipitation. Two factors may be contributing to this result: (i) the natural high correlation between large-scale circulation pattern in winter, and (ii) the relatively higher variability in summer precipitation.

10. Standardized Precipitation Index (SPI) is proposed as it is a simple and effective index that can be used for drought definition. It is spatially invariant quantity that can be computed to give precipitation deficit at multiple timescales. It can be calculated for different time scales.

11. Based on SPI(6) for Vacoas, we found that drought durations vary between 1 and 9 months with a mean of 2.6 months. The mean inter-arrival time is 15 months. We also identified the severity of all drought events between 1961 and 2012.

12. We determined the probability distributions of duration and severity. Duration was found to follow a gamma distribution and severity followed an exponential distribution.

13. This framework is useful for determining the probability of drought occurrences and in assessing the conditional distribution of severity given duration exceeding various durations and conditional distribution of duration given severity exceeding various severity values.

7   CONCLUSIONS AND WAY FORWARD

▪ Our study shows that techniques within the realm of hydroinformatics can be useful in precipitation prediction and other key parameters, such as water levels in reservoirs. Further analysis is recommended to assess the usefulness and limitations of related techniques in machine learning, artificial intelligence, and statistics.

▪ Data on water levels (or capacity) in reservoirs along with consumption can be used to model the relationship between large-scale circulation patterns and water availability. Along the lines of precipitation prediction, models for water level prediction can be developed.

▪ Our study was based on only one meteorological variable at one station. This is highly insufficient for faithful modeling. We suggest that the analysis be extended to include other variables, such as temperature, pressure, evaporation, humidity, etc. It has to be spatially extended to cover the island. This will help in cross validating the results obtained.

▪ The study covered a short period (1961 to 2012). If longer time series are available, trends related to climate change may be determined.

▪ Droughts in Mauritius, at least in the public's perception, are identified when the water level in the major reservoirs fall below the long term mean. This is probably the simplest and easiest way to identify droughts, but it is not necessarily the most effective way. We propose an index that captures the multi-dimensional nature of drought: duration, severity, and inter-arrival time.

▪ Again, our analysis on drought duration and severity was limited to one station only. This study can be extended to cover the whole island and conditional drought duration and severity maps (similar to isohyet) can be developed for long-term water planning.

▪

8 REFERENCES

Barnston, A. G., M. Chelliah, and S. B. Goldenberg (1997), Documentation of a highly ENSO-related SST region in the equatorial Pacific: Research Note, *Atmosphere-Ocean*, *35*(3), 367 - 383.

Barnston, A. G., M. Chelliah, and S. B. Goldenberg (1997), Documentation of a highly ENSO-related SST region in the equatorial Pacific: Research Note, *Atmosphere-Ocean*, *35*(3), 367 - 383.

Booneeady, P. (2010), Trends in Climate Indices in the Republic of Mauritius: 1950-2008 *Rep.*, 18 pp, Mauritius Meteorological Services.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (2008),), *Time series analysis : forecasting and control,* , 4 ed., Wiley Hoboken, N.J.

Chowdhary, H., L. Escobar, and V. Singh (2011), Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data, *Hydrology Research*, *42*(2–3), 193–216.

Cordery, I., and S. L. Yao (1993), Extreme Hydrolosical Events: Precipitation, Floods and Droughts in *IAHS Symposium*, edited, Yokohama.

Deepchand, D., and C. P. Khedun (2005), Water Resources Development and Management - Present Status and Future Program, in *Development of a Mauritius Country Framework for Action within the Southern African Vision for Water, Life, and Environment in the 21st Century*, edited by E. Seenyen, C. Seenyen and C. P. Khedun, Pailles, Mauritius.

Govindaraju, R. S., and A. R. Rao (2000), *Artificial Neural Networks in Hydrology*, Kluwer Academic Publishers, Amsterdam.

Guttman, N. B. (1999), Accepting the Standardized Precipitation Index: A Calculation Algorithm, *Journal of the American Water Resources Association*, *35*(2), 311-322.

Hirsch, R. M., D. R. Helsel, T. A. Cohn, and E. J. Gilroy (1992), Statistical Analysis of Hydrologic Data, in *Handbook of Hydrology*, edited by D. R. Maidment, McGraw-Hill, Inc., New York.

Kumar, P., M. Folk, M. Markus, and J. C. Alameda (2006), *Hydroinformatics: Data Integrative Approaches in Computation, Analysis, and Modeling*, CRC Press, Boca Raton, FL.

Mackay, D. J. C. M. (2003), *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press.

Maier, H., and G. Dandy (2000), Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications, *Environmental Modelling & Software*, *15*, 101-124.

McKee, T. B., N. J. Doesken, and J. Kleist (1993), The Relationship of Drought Frequency and Duration to Time Scales, in *Eighth Conference on Applied Climatology*, edited, Anaheim, California.

Mitchell, T. M. (1997), *Machine Learning*, McGraw Hill.

Pittock, A. B. (1984), On the reality, stability and usefulness of southern hemisphere teleconnections, *Aust. Met. Mag.*, *32*, 75-82.

Shiau, J. (2006), Fitting Drought Duration and Severity with Two-Dimensional Copulas, *Water Resources Management*, *20*(5), 795-815.

Silverman, D., and J. A. Dracup (2000), Artificial Neural Networks and Long-Range Precipitation Prediction in California, *Journal of Applied Meteorology*, *39*(1), 57-66.

Sklar, A. (1959), Fonctions de repartition à n dimensions et leurs marges, *Publications de l'Institut de Statistique de l'Université de Paris*, *8*, 229-231.

Sun, H., and D. Furbish (1997), Annual precipitation and river discharges in Florida in response to El Niño and La Niña-Southern Oscillation of sea surface temperature, *Journal of Hydrology*, *199*(1-2), 74-87.

United Nations (2011), International Decade for Action Water for Life 2005-2015: Water scarcity, edited.