# “Rainfall prediction for sustainable economic growth”


| | |
|---|---|
| **AUTHORS** | Retius Chifurira https://orcid.org/0000-0001-7889-3417<br>Delson Chikobvu https://orcid.org/0000-0003-3979-658X<br>Dorah Dubihlela |



| | |
|---|---|
| **ARTICLE INFO** | Retius Chifurira, Delson Chikobvu and Dorah Dubihlela (2016). Rainfall prediction for sustainable economic growth. *Environmental Economics*, *7*(4-1), 120-129. doi:10.21511/ee.07(4-1).2016.04 |
| **DOI** | http://dx.doi.org/10.21511/ee.07(4-1).2016.04 |
| **RELEASED ON** | Wednesday, 21 December 2016 |
| **JOURNAL** | "Environmental Economics" |
| **FOUNDER** | LLC “Consulting Publishing Company “Business Perspectives” |

Retius Chifurira (South Africa), Delson Chikobvu (South Africa), Dorah Dubihlela (South Africa)

# Rainfall prediction for sustainable economic growth


## Abstract

Agriculture is the backbone of Zimbabwe's economy with the majority of Zimbabweans being rural people who derive their livelihood from agriculture and other agro-based economic activities. Zimbabwe's agriculture depends on the erratic rainfall which threatens food, water and energy access, as well as vital livelihood systems which could severely undermine efforts to drive sustainable economic growth. For Zimbabwe, delivering a sustainable economic growth is intrinsically linked to improved climate modelling. Climate research plays a pivotal role in building Zimbabwe's resilience to climate change and keeping the country on track, as it charts its path towards sustainable economic growth. This paper presents a simple tool to predict summer rainfall using standardized Darwin sea level pressure (SDSLP) anomalies and southern oscillation index (SOI) that are used as part of an early drought warning system. Results show that SDSLP anomalies and SOI for the month of April of the same year, i.e., seven months before onset of summer rainfall (December to February total rainfall) are a simple indicator of amount of summer rainfall in Zimbabwe. The low root mean square error (RMSE) and root mean absolute error (RMAE) values of the proposed model, make SDSLP anomalies for April and SOI for the same month an additional input candidates for regional rainfall prediction schemes. The results of the proposed model will benefit in the prediction of oncoming summer rainfall and will influence policy making in agriculture, environment planning, food redistribution and drought prediction for sustainable economic development.


**Keywords:** sustainable economic growth, standardized Darwin sea level pressure anomalies, southern oscillation index, summer rainfall prediction, Zimbabwe.
**JEL Classification:** Q16, Q25, Q54, Q55, Q58.

## Introduction

The poor economic performance of Southern Africa continues to receive considerable amount of attention in the economic literature (Chilonda and Minde, 2007). For Zimbabwe, agriculture is the backbone of the economy. It provides employment and income for 60 to 70 percent of the population, supplies 60 percent of the raw materials required by the industrial sector and contributes 40 percent of the total export earnings. Despite agriculture offering high level employment opportunities, it only contributes at least 20 percent to the annual Gross Domestic Product (GDP) of the country depending on the rainfall patterns (Government of Zimbabwe, 2001; Jury, 1996). The contribution of the sector to the economy has not been fully realized. Concerns of economic growth, environmental issues and sustainable development is a relatively recent event which has captured the attention of researchers, aid agencies and development and environmental planners. This is because sustainable development may equate to sustainable economic growth (Lele, 1991). Economic sustainability seeks to avoid extreme future imbalances in production (Harris et al., 2003). With long run economic sustainability, welfare is maximized over time. The interest rests in the need for development programs rests on the need for eradication of poverty and food insecurity in most economies (Perman et al., 2003).

Ample theories have been put forward to explain the relatively poor economic growth of the Zimbabwe and other sub-Saharan countries (Jury, 1996; Manatsa et al., 2008; Collier and Gunning, 1999). In essence, the theories can be categorized into those arising from political and those due to exogenous factors. Political explanations usually refer to poor and inconsistent policies that are argued to have impacted negatively to economic growth in Zimbabwe and other sub-Saharan African countries (FAO, 2001; Mangoyana; Meda, 2001; UNECA, 2000). These include poor fiscal and trade policies, lack of good governance, corruption and illfunctional financial and labor markets. Exogenous explanations include external aid allocation (Burnside and Dollar, 1997) and lack of diversification of exports tropical climates including lack of early drought warning tools (Sachs and Warner, 1997). Given the importance of agriculture to a developing country such as Zimbabwe and the dependence of the sector to rainfall, as suggested by Manatsa et al., 2008 and Mangonyana and Meda, 2001, its decline and lack of prediction tools may have severe consequences for sustainable growth. Additionally, this decline and lack of prediction tools poses detrimental impact on energy supply in Zimbabwe due to its heavy reliant on hydro-power for electricity generation (Kaunda et al., 2012).

Rainfall patterns is affected by various natural phenomena. Firstly, large scale climatic variation that occur from one year to year (Panu and Sharma, 2002).

---


Retius Chifurira, School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, South Africa.
Delson Chikobvu, Department of Mathematical Statistics and Actuarial Science, University of the Free State, South Africa.
Dorah Dubihlela, Unit of Applied Economics and Real Estate, Cape Peninsula University of Technology, South Africa.

This is the Southern Oscillation climatic condition, which manifests itself in the differential oceanic temperature phenomenon across the tropical Pacific Ocean. The Southern Oscillation Index (SOI), as defined, is the difference between seasonally normalized sea level pressures of Darwin (in Australia) and Tahiti (in the Mid Pacific). Secondly, Darwin Sea Level Pressures (Darwin SLP) have been found to influence seasonal rainfall patterns in Zimbabwe (Manatsa et al., 2008).

Evidence of relationships among meteorological variables is well documented (Webster, 1981; Rocha, 1992; Ropelewski and Halpert, 1987). Most researches on rainfall patterns for Zimbabwe have focused on correlations between phases of SOI and rainfall (Matarira and Unganai, 1994; Torrance, 1990; Waylen and Henworth, 1995; Richard et al., 2000). Makarau and Jury (1997) used a host of meteorological variables to predict summer rainfall in Zimbabwe. Ismail (1986) proposed an empirical rule from which the mean seasonal rainfall over Zimbabwe can be predicted three months before the start of the rainy season and ten months before its end using SOI. The author concluded that SOI has an influence on the seasonal rainfall over Zimbabwe. Manatsa et al. (2008) used correlation analysis to identify the lag periods for which SOI and Darwin pressure anomalies are significantly correlated with the Zimbabwean Summer Precipitation Index. The authors conclude that progressive lagged four months averaged Darwin sea level pressure anomalies are correlated with the Zimbabwean precipitation index. Our work advances the work done by Manatsa et al. (2008) by trying to find the SDSLP anomalies for a particular month at a particular lag which correlates with summer rainfall for Zimbabwe.

In this paper, we aim to develop a simple early warning rainfall predictive model using climatic determinants such as SOI value and SDSLP anomalies for Zimbabwe, at a longer lead time before the onset of the rainfall season. In this paper, we focus on summer rainfall totals, i.e., monthly rainfall totals for the months December to February. The summer rainfall patterns is crucial for agriculture, water management, hydro-power electricity generation and infrastructure design, since the country usually receives the highest amount of rainfall during these months. Any meaningful planning requires information based on the rainfall patterns of the crucial months of the rainfall season. We are not aware of any literature relating to modelling summer rainfall using SDSLP anomalies and SOI for Zimbabwe.

The remainder of the paper is organized as follows. In section 1, we discuss the importance of rainfall for Zimbabwe's economic performance. Section 2 describes the data and data sources. In section 3, the description of the methodology used to analyze the data set is discussed. Section 4 and final section discuss the main empirical results and conclusions, respectively.

## 1. Rainfall and economic growth in Zimbabwe

Rainfall could potentially have a wide range of economic implications in the developing world (Barrios et al., 2010). There is an inherent interdependence between the amount of rainfall and the economic performance. The 5th Assessment Report (2007) of the United Nations' Intergovernmental Panel on Climate Change (IPCC) acknowledges the threat of climate change to Africa's recent economic gains. Increasing temperatures, rising sea levels and erratic rainfall put strain on climate-sensitive sectors such as agriculture, water management and other vital livelihoods systems. The increasing understanding of climate system, its system variability, and the variability in terms of specific regional rainfall patterns provides fundamental opportunity for a successful paradigm for economic growth. Erratic rainfall patterns in Zimbabwe seem to have catastrophic consequences on peoples' livelihoods and the smooth functioning of the economy. We briefly discuss the main channel through which rainfall is likely to have affected Zimbabwe's economic growth. Agriculture has traditionally had a higher share in Gross Domestic Product (GDP) in Zimbabwe than in any other Southern African country. The table below shows contribution to GDP by three major sectors in Zimbabwe for the period 2004 to 2015.

Table 1. Percentage contribution to GDP by sector in Zimbabwe in 2014

| Year | Agricultural sector % | Industrial sector % | Service sector % |
|---|---|---|---|
| 2004 | 19.58 | 26.42 | 54 |
| 2005 | 18.58 | 28.68 | 52.74 |
| 2006 | 20.28 | 32.33 | 47.39 |
| 2007 | 21.6 | 33.07 | 45.33 |
| 2008 | 19.4 | 31.09 | 49.51 |
| 2009 | 15.07 | 29.64 | 55.28 |
| 2010 | 14.51 | 30.82 | 54.64 |
| 2011 | 13.21 | 32.69 | 54.11 |
| 2012 | 13.15 | 31.6 | 55.24 |
| 2013 | 12 | 31.1 | 56.9 |
| 2014 | 14.01 | 29.41 | 55.59 |
| 2015 | 12.53 | 28.47 | 58.99 |

Source: Statistica (2015).

According to the table above, the agricultural sector's contribution to GDP was lowest in 2013 where the contribution was only 12% and highest in

2007 with 21.6% contribution. This contribution is, however, dependant on the climatic factors like the amount of precipitation received, whether or not that precipitation was received in time by farmers and availability of precautionary measures in cases of erratic rainy seasons. Considering the importance of agriculture, Zimbabwe has been devastated by severe droughts in recent years (United Nations, 2016). This has impacted negatively of the performance of the country to meet its potential growth since agricultural sector is susceptible to shortages in rainfall.

## 2. The data

To analyze summer rainfall for Zimbabwe, we use monthly mean annual rainfall for the period 1901 to 2014. The rainfall data set was obtained from Department of Meteorological Services of Zimbabwe. Rainfall which is critical for crop farming, are the rains for the months December to February. Any rainfall amount below the normal rainfall for the region will adversely affect crop farming especially maize. The summer rainfall is the mean rainfall for the months December to February of the following year for 16 weather stations in the study area. This is the period when region receives high rainfall crucial for agriculture hydro-power electricity generation and water management. The dataset is divided into in-sample data set (1901 to 2009) and out-of-sample data set (2010 to 2014). The out-of-sample dataset is used to check the forecasting power of the proposed model. Table 2 shows the descriptive statistics of in-sample summer rainfall for the study area.

Table 2. Descriptive statistics and normality test (*p*-value in brackets) of the summer rainfall data for the period 1901 to 2009 (in-sample)

| N | Min | Max | Mean | S. dev. | Skewness | Kurtosis | Jarque-Bera statistics |
|---|---|---|---|---|---|---|---|
| 108 | 233.22 | 810.70 | 497.66 | 123.49 | 0.29 | -0.29 | 1.80 (0.41) |

The coefficient of skewness is 0.29 which small and the kurtosis is less than 3. This indicates that the summer rainfall data are approximately normally distributed. This confirmed by the Jarque-Bera test. The *p*-value of the Jarque-Bera statistic is 0.41 < 0.05, thus, we fail to reject the null hypothesis of normality of the summer rainfall at 5% level of significance. The highest rainfall was obtained in 1923, while the lowest rainfall was in 1992 (the worst drought in the given history of the country). Figure 1 shows the time series plot of the summer rainfall for the period 1901 to 2009.

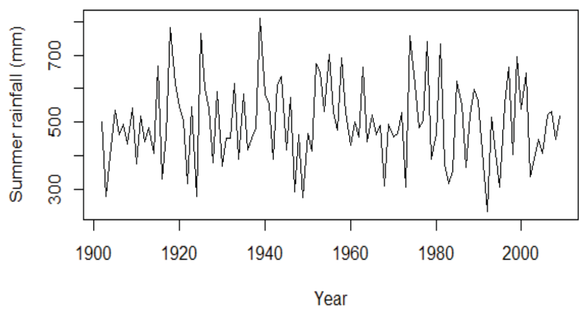


**Fig. 1. Time series plot of summer rainfall for the period 1901 to 2009 (in-sample)**

From Figure 1, it seems reasonable to assume that the pattern variation has stayed stationary over the observation period. However, we test for stationarity of the rainfall data using the Augment Dickey-Fuller (ADF) test, Phillips-Perron (PP) test and the Kwiatkoski-Phillips-Schmidit-Shin (KPSS) test. The null hypothesis for the ADF and PP tests is summer rainfall data is non-stationary, while the null hypothesis for the KPSS test is data is stationary. Table 2 shows the results of testing for stationarity of summer rainfall data. We observe that the data are stationary, as confirmed by the tests of stationarity presented in Table 3.

Table 3. Stationarity tests statistics for summer rainfall data (*p*-value in brackets)

| ADF statistic | PP statistic | KPSS statistic |
|---|---|---|
| -5.358 (<0.01) | -10.915 (<0.01) | 0.067 (0.10) |

Monthly SDSLP anomalies and SOI values are used in this study. Sea Level Pressure is the atmospheric pressure at mean sea level either directly measured by stations at sea level or empirically determined when the station is not at sea level (Mason, 1997). The monthly SDSLP anomalies and SOI values are obtained from NOAA, National Weather Service Climate Prediction Centre website. The SOI is calculated from the monthly or seasonal fluctuations in the air pressure difference of the area between Tahiti (in the mid-Pacific) and Darwin (in Australia).

## 3. Research methodology

In this paper, dependent rainfall variable is expressed in terms of independent explanatory variables, SDSLP anomalies and SOI. Multiple linear regressions can be used to model a relationship between the dependent variable and the explanatory variables. It allows investigating the effect of changes in the various factors on the dependent variable. If the observations are measured over time, the model becomes a time series regression model. The resulting statistical relationship can be used to predict values of rainfall. To ascertain the predictive power of the model, all assumptions of multiple linear regressions must be met.

**3.1. Multiple regression model.** Probabilistic models that include more than one independent variables are called multiple regression. The model can be written as:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_p x_{p,t} + N_t, \quad (1)$$

where $y_t$ is the $t^{th}$ observation of the dependent variable, $x_{i,t}$ for $i = 1,2,\dots,p$ is the corresponding observation of the explanatory variable whose predictive influence is of interest. Parameters $\beta_i$ are unknown and the probabilistic component of the model $N_t$ is the unknown error term. The value of the coefficient $\beta_i$ determines the contribution of the independent variable $x_{i,t}$ given that the other independent variables are held constant. Using classical estimation techniques estimates for the unknown parameters are obtained. If the estimated values for $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ are given by $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$, then, the dependent variable is estimated as:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t} + \hat{\beta}_2 x_{2,t} + \dots + \hat{\beta}_p x_{p,t} \quad (2)$$

and the estimate $\hat{N}_t$ for the error term $N_t$ is determined as the difference between the observed and the predicted dependent variable; $\hat{N}_t = y_t - \hat{y}_t$. In the theoretical model, several assumptions are made about the explanatory variables and the error term.

Firstly, there must be insignificant correlation between the explanatory variables. When the explanatory variables are correlated, multicollinearity problem arises. The estimated parameters will be unstable and unreliable if highly correlated variables are used in the model as explanatory variables. In the study at hand, the predictive power of SOI and SDSLP anomalies at a maximum lag is important. SOI values are calculated using SDSLP values and, thus, high correlation is anticipated. Principal component analysis is used to produce orthogonal explanatory variables.

**3.2. Principal component analysis.** Principal Component Analysis (PCA) also known as empirical orthogonal function has been used in many different disciplines including finance, agriculture, biology, chemistry, climatology, demography, ecology, psychology and meteorology. PCA is a technique used to combine highly correlated factors into principal components that are much less highly correlated with each other. This improves the efficiency of the model.

In this study, the predictive power of SDSLP anomalies ($I_1$) and SOI values ($I_2$) is explored. Two new, uncorrelated factors, $I_1^*$ and $I_2^*$, can be constructed as follows:

Let $I_1^* = I_1$.

Then, we carry out a linear regression analysis to determine the parameters $\gamma_1$ and $\gamma_2$ in the equation:

$$I_2 = \gamma_1 + \gamma_2 I_1^* + \varepsilon_1 . \quad (3)$$

$\gamma_1$ and $\gamma_2$ are the intercept and slope parameters of the regression model, respectively, and $\varepsilon_1$ is the 'error' term, which by definition is independent of $I_1^* = I_1$.

We, then, set:

$$I_2^* = \varepsilon_1 = I_2 - (\gamma_1 + \gamma_2 I_1^*) . \quad (4)$$

By construction, $I_2^*$ is uncorrelated with SDSLP anomalies ($I_1$), since $I_2^* = \varepsilon_1$, the residual term in the equation. Changes in $I_2^*$ are interpreted as the change in the observed values of SOI ($I_2$) that cannot be explained by the observed change in SDSLP ($I_1$). $I_2^*$ in the rainfall model (1) explains the component of rainfall that cannot be explained by the SDSLP anomalies.

The other assumptions of the rainfall model (1) are that there is no serial correlation and heteroscedasticity of error terms. These assumptions are likely to be violated in regression models with time series data. Autocorrelation (the error terms being correlated among themselves through time) leads to regression coefficients which are unbiased, inefficient and the standard errors are probably wrong making $t$ tests and $F$ tests unreliable. In a regression with auto-correlated errors, the errors will probably contain information that is not captured by the explanatory variables. The Durbin-Watson test is used to assess whether the residuals are significantly correlated. A Durbin-Watson statistic of 2 indicates absence of autocorrelation. The autocorrelation function and partial autocorrelation function can also be used to detect autocorrelation among the residuals.

**3.3. Weighted regression model.** The multiple least squares criterion weighs each observation equally in determining the estimates of the parameters. The procedure treats all of the data equally, giving less precise measured points more influence than they should have and gives highly precise points too little influence. The weighted least squares weighs some observations more heavily than others, giving each data point its proper amount of influence over the parameter estimates, and this maximizes the efficiency of parameter estimation. Weighted least square reflects the behavior of the random errors in the model.

The model $y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + N_t$

Let $Y = [y_1, y_2, \ldots . y_t]'$,

$\beta = [\beta_0, \beta_1, \beta_2]'$,

$X = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} \\ \ldots . & \ldots . & \ldots . \\ 1 & x_{1,t} & x_{2,t} \end{bmatrix}$ and $N = [N_1, N_2 \ldots N_t]'$,

then, the same model equation 1 can be written as

$$Y = X\beta + N, \quad (5)$$

parameter estimates using ordinary least squares can be found as

$$\hat{\beta} = (X^{/}X)^{-1}(X^{/}Y) = [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2]'. \quad (6)$$

To find the weighted least squares parameters of the weighted model, we minimize the weighted sum of squared errors (WSSE).

$$WSSE = \sum_{t=1}^{n} w_t (y_t - \hat{y}_t)^2 = \sum_{t=1}^{n} w_t (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1,t} - \hat{\beta}_2 x_{2,t})^2, \quad (7)$$

where $w_t > 0$ is the weight assigned to the $t^{th}$ observation. The weight $w_t$ can be the reciprocal of the variance of that observation's error term, $\sigma_t^2$, i.e.

$$w_t = \frac{1}{\sigma_t^2}. \quad (8)$$

Observations with larger error variances will receive less weight (and, hence, have less influence on the analysis) than observations with smaller error variances. The estimates are:

$$\hat{\beta} = (X^{/}WX)^{-1}(X^{/}WY), \quad (9)$$

where $W = [w_1, w_2, \ldots w_t]$ is the weight vector.

The biggest disadvantage of weighted least squares is the fact that the theory behind this method is based on the assumption that the weights are known exactly. This is almost never the case in real applications where, instead, estimated weights are used (Carrol and Ruppert, 1988).

**3.4. Assessing model performance.** To evaluate the performance of the considered models, we apply the measures of average error, namely, mean absolute error (MAE) and root mean square error (RMSE). These measures are based on statistical summaries of $N_t$ $(t = 1, 2, \ldots \ldots, T)$. The average model-estimation error can be written generically as:

$$\bar{N}_t = [\sum_{t=1}^{T} \upsilon_t |N_t|^{\gamma} \backslash \sum_{t=1}^{T} \upsilon_t]^{1 \backslash \gamma} \;, \quad (10)$$

where $\gamma \geq 1$ and $\upsilon_t$ is a scaling assigned to each $|N_t|^{\gamma}$ according to its hypothesized influence on the total error (Willmott and Matsuura, 2005). For the calculation of RMSE, $\gamma = 2$ and $\upsilon_t = 1$. RMSE is measured in the same unit as the forecast and is given by:

$$RMSE = [T^{-1} \sum_{t=1}^{T} |N_t|^2]^{1 \backslash 2}. \quad (11)$$

The MAE is also measured in the same unit as the forecast, but gives less weight to large forecast errors than the RMSE. To obtain the MAE, we set $\gamma = 1$ and $\upsilon_t = 1$ and is given by:

$$RMSE = [T^{-1} \sum_{t=1}^{T} |N_t|]. \quad (12)$$

According to Willmott and Matsuura (2005) and Trück and Liang (2012), MAE is the most natural measure of average error magnitude, and that it is an unambiguous measure of average error magnitude. The MAE and RSME values can range from 0 to infinity, and smaller values indicates a better model.

**3.5. Model selection criteria.** The scope of model selection is to identify the model that is better suited to predict summer rainfall using SDSLP anomalies and a component of SOI not explained by SDSLP anomalies. In this paper, we consider the Akaike information criterion (AIC). The Akaike information criterion uses the Kullback-Leibler's information as the discrepancy measure between the true model $f(x)$ and the approximating model $M_j = \hat{y}_t(x, \hat{\beta}_i)$. The Kullback-Leibler information between the two models is defined as

$$I\left(M_j, f(x)\right) = \int \ln\left(\frac{f(x)}{\hat{y}_t(x,\hat{\beta}_i)}\right) f(x)dx, \quad (13)$$

where $I\left(M_j, f(x)\right)$ denotes the information lost when $M_j$ is used to approximate $f(x)$ (Laio et al., 2009). Thus, $I\left(M_j, f(x)\right)$ is regarded as the distance from $M_j$ to $f(x)$. A good approximating model is the one that minimises the information lost, i.e., minimizing $I\left(M_j, f(x)\right)$ over $M_j$. The AIC for the $j^{\text{th}}$ model is given by

$$\text{AIC}_j = -2\ln\left(L_j(\hat{\beta}_i)\right) + 2p_j, \quad (14)$$

where $L_j = \Pi_{t=1}^{n} \hat{y}_t(x_t, \hat{\beta}_i)$ is the likelihood function corresponding to the maximum likelihood estimator of the parameter $\beta_i$ and $p_j$ is the number of estimated parameters of the $j^{\text{th}}$ model. The model with the minimum AIC value is, then, selected to be the best predictive model.

## 4. Results

In this section, correlations between summer rainfall for Mashonaland region and SDSLP anomalies are discussed. The results from simple and weighted regression models are also presented.

**4.1. Comparing lead times of SDSLP anomalies.** Correlation analysis between the SDSLP anomalies and the summer rainfall is used to identify the month of the previous year whose SDSLP anomalies is significantly correlated to the summer rainfall (December to February total rainfall). Table 4 reports the correlation analysis between the climatic variables.

Table 4. Correlations analysis between summer rainfall and climatic variables

| Month | SDSLP anomalies | SOI |
|---|---|---|
| Jan | -0.001 | 0.058 |
| Feb | -0.036 | 0.101 |
| Mar | -0.159 | 0.067 |
| Apr | -0.276 | 0.193 |
| May | -0.093 | 0.215 |
| June | -0.137 | 0.116 |
| July | -0.159 | 0.207 |
| Aug | -0.263 | 0.344 |
| Sept | -0.176 | 0.329 |
| Oct | -0.168 | 0.274 |
| Nov | -0.288 | 0.263 |
| Dec | 0.208 | 0.221 |

The focus of this study is to determine a particular month's SDSLP anomaly which has a high correlation with summer rainfall at a lead time of more than six months. The highest correlation is –0.246 (SDSLP anomaly for April of the previous year). At a lead time of more than a year, the correlations between the SDSLP anomalies and summer rainfall are insignificant. In this paper, we propose the summer rainfall model using SDSLP anomalies for April of the previous year as the explanatory variable.

At a longer lead time, SOI value for April is positively correlated to summer rainfall. We also observed that SOI value for April is highly correlated with SDSLP anomalies for April (–0.804), thus, it is used as an explanatory variable in a regression model to construct orthogonal explanatory variables $I_1^*$ and $I_2^*$. SDSLP anomalies for April ($I_1^*$) is used as the dependent variable. $I_2^*$ is the residual series obtained from the equation (3) which is uncorrelated to SDSLP for April ($I_1^*$).

**4.2. Summer rainfall predictive model.** Table 4 shows the results of the multiple regression approach to predict Zimbabwe's summer rainfall using the SDSLP anomalies and the principal component of SOI of April which is not explained by SDSLP anomalies. The multiple regression model is:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t-1} + \hat{\beta}_2 x_{2,t-1}, \tag{15}$$

where $\hat{y}_t$ is the predicted summer rainfall, $x_{1,t-1}$ is SDSLP anomalies for April of the previous year and $x_{2,t-1}$ is $I_2^*$ the component of SOI value for April of which is not explained by the corresponding SDSLP anomalies.

Table 5. Parameter estimates for regression model (standard errors in brackets)

| Predictor variable | Parameter estimate |
|---|---|
| $SDSLP_{Apr}$<br>$I_2^*$ | $\hat{\beta}_0$ = 487.608 (12.229)<br>$\hat{\beta}_1$ = - 13.164 (13.107)<br>$\hat{\beta}_2$ = - 1.008 (1.612) |

Note: all parameters are significant at 5% level. Adjusted $R^2 = 0.06$. Durbin-Watson statistic = 2.226.

From Table 5, the model estimates are all significant at 5% level of significance. The model explains only 6% of variations in the summer rainfall. The Durbin-Watson statistic of 2.226 indicates that the model does not violate the assumption of serial correlation of the residuals. However, the model can be improved by using different weights in the regression model.

We now use different weighting scales to improve the forecasting power of the model. Table 5 shows the weighted linear regression models results for rainfall $\hat{y}_t^*$:

$$\hat{y}_t^* = \hat{\beta}_0^* + \hat{\beta}_1^* x_{1,t-1} + \hat{\beta}_2^* x_{2,t-1}, \tag{16}$$

where $x_{1,t-1}$ is SDSLP anomalies for April of the previous year and $x_{2,t-1}$ is $I_2^*$ the component of SOI which is not explained by SDSLP anomalies for the same month. Various weights are considered in arriving at estimates using weighted regression.

Table 6. Parameter estimates for weighted regression model (standard errors in brackets)

| Weights | Parameter estimate $\hat{\beta}_0$ $\hat{\beta}_1$ $\hat{\beta}_2$ | Adjusted $R^2$ | AIC | MAE | RMSE |
|---|---|---|---|---|---|
| $\frac{1}{I_2^*}$ | 575.442 (34.038)-<br>92.272 (16.014)<br>-11.486 (3.000) | 0.425 | 11.991 | 127.504 | 154.7 |
| $I_2^*$ | 549.730 (30.032)<br>-76.364 (16.460)<br>-8.446 (3.250) | 0.295 | 12.067 | 146.778 | 183.0 |
| $\frac{1}{(I_2^*)^3}$ | 432.849 (8.918)<br>-186.638 (14.959)<br>-5.502 (0.533) | 0.590 | 15.371 | 147.411 | 183.0 |

We use the MAE and RMSE measures to assess the forecasting performance of the models and the AIC to select the best fitting model. The weights are assumed to be proportional to inverse standard deviation. The weight $\frac{1}{\sigma_t^2}$ , of $I_2^*$ used as the weighting model is selected, since it has the least AIC, MAE and RMSE values . The model performs better than the other models in forecasting summer rainfall, since it has the least MAE and RMSE values. The model is significant at 5% significance level.

The multiple adjusted $R^2 = 0.425$, which indicates that the model explains 42.5% of the variations in summer rainfall from just two variables. ENSO explains only approximately 30% of the rainfall variability, which means that other factors should also be taken into account when predicting rainfall. The ACF and PACF correlogram (Appendix) shows that the residuals are approximately independent. This is confirmed by the Durbin-Watson statistic value equal to 1.8, indicating that the residuals are approximately independent. The model does not violate the assumption of homoscedasticity. This is confirmed by the White test, which produce an F-statistic of 12.759 ($p$-value = 0.00 < 0.05). It is important to check the in-sample forecasting power of the model. Figure 2 shows the observed summer rainfall against the predicted rainfall from the selected model.

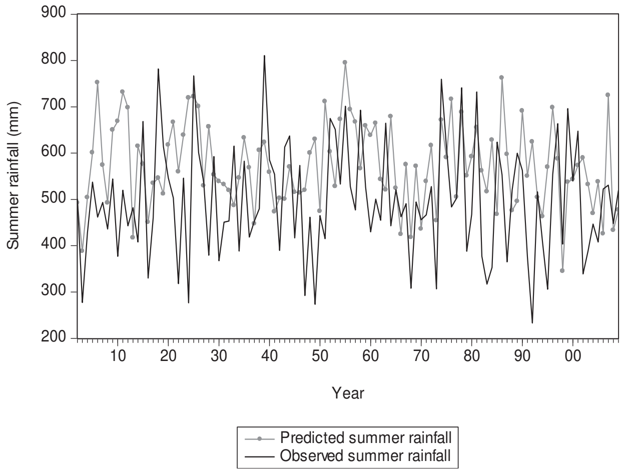


**Fig. 2. Summer rainfall versus predicted rainfall**

From Figure 2, the model seems to be able to predict in-sample summer rainfall. The model seems to shows some little variability between forecasts and actual rainfall. The model can be used to predict one year ahead summer rainfall for Zimbabwe. The out of sample forecasts for the years 2010 to 2014 are shown in Table 7. The out of sample forecasts seem to be reasonable with insignificant differences between the actual summer rainfall and the predicted values and are significant at 5% level. The MAE value of 111.22 and RMSE value of 141.65 are not so big considering that only two climatic explanatory variables have been used in the model. The model seems to under-forecast the summer rainfall for the years 2011 and 2014 while over-forecast for the year 2012. However, during the entire out-of-sample period, the model is forecasting the actually summer rainfall by 3.11%. This suggests that the proposed model is reasonable and can be used for predicting summer rainfall for Zimbabwe and other southern African countries.

Table 7. Out-of-sample forecasts

| Year | Summer rainfall (mm) | Predicted summer rainfall (mm) | Performance of model (%) |
|---|---|---|---|
| 2010 | 532.99 | 573.68 | -7.63 |
| 2011 | 548.31 | 683.24 | -24.61 |
| 2012 | 604.42 | 342.74 | 43.29 |
| 2013 | 524.94 | 515.17 | 1.86 |
| 2014 | 426.09 | 535.12 | -25.35 |
| Average | | | -3.11 |

Note: - indicates that the observed summer rainfall is less than the rainfall predicted by the proposed model.

## Conclusions

We developed a simple summer rainfall predicting tool for Zimbabwe using SDSLP anomalies and SOI. The simple model can be used as part of a drought early warning system. This paper's main finding is that summer rainfall for Zimbabwe correlates with SDSLP anomalies for the month of April. We employed principal component analysis to construct orthogonal factors (non-collinear variables), since SDSLP anomalies and SOI are correlated. The combination of regression and time series analysis offers a powerful tool for predicting summer rainfall using SDSLP anomalies and SOI values of a particular month with a lag of at least six months. Using SDSLP anomalies and the component of SOI for April which is not explained by SDSLP anomalies, the summer rainfall for the year ahead can be predicted. Developing a simple model for summer rainfall prediction helps in reducing the adverse effects on the productivity of different crops, as highlighted by IPCC (2007). Since agriculture is linked to the other sectors of the economy, variations in rainfall will affect farming activities and the rest of the economy through inter sectorial relations.

## Implications of the study

A predictive model of summer rainfall seven months before the onset of summer rainfall will help in choosing of planting drought resistant crops in time. This will have important effects on agriculture supply and demand for input factors such as fertilizers. This, in turn, provides important forecast for drought in time so that precautionary measures can be taken in advance. These measures may involve an adjustment in the national budget expenditure in order to carter for the forthcoming drought, deciding on the type of seeds to plant and other possible measures of saving water as a scarce resource. The following measures may be taken based on the research results:

- Adaptive measures, thus, create standby measures to deal with climate related disasters.
- Shifting agricultural activities to be in line with the amount of anticipated rain.

Regarding the data, it is clear that the explanatory variables incorporated in the model are limited. It would be interesting to include other climatic determinants such as Sea Surface Temperatures at Darwin and wind speed. However, the use of weighted regression gives an acceptable fit in the absence of these other factors. Extending the model with more factors may give a better understanding of the rainfall patterns in Zimbabwe. This is an area for further research.

## References


1. Barrios, S., Bertinelli, L. and Strobl, E. (2010). Trends in Rainfall and Economic Growth in Africa: A neglected cause of the African growth tragedy, *The Review of Economics and Statistics*, 92(2), pp. 350-366.
2. Burnside, C. and Dollar, D. (1997). *Aid, Policies and Growth, World Bank Research*, Working Papers on International Economics, Trade and Capital Flows, No. 1777.
3. Carroll, R.J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*. New York: Chapman and Hall.
4. Chilonda, P. and Minde, I. (2007). *Agricultural growth trends in Southern Africa*, Regional Strategic Analysis and Knowledge Support System for Southern Africa, South Africa.
5. Collier, P. and Gunning, J.W. (1999). Why has Africa grown slowly? *Journal of Economics Perspectives*, 13, pp. 3-22.
6. Common, M. and Perrings, C. (1992). Towards an ecological economics of sustainability, *Ecological economics*, 6(1), pp. 7-34.
7. Emas, R. (2015). The Concept of Sustainable Development: Definition and Defining Principles, *Florida International University*.
8. FAO. (2001). *Global Watch Report*. United Nations, Rome.
9. Government of Zimbabwe. (2001). *The agricultural sector of Zimbabwe*, Statistical Bulletin, Zimbabwe.
10. Harris, J.M., Roach, B. and Codur, A.M. (2000). *The economics of global climate change*, Medford: Tufts University.
11. Ismail, S.A. (1986). Long-Range seasonal rainfall forecast for Zimbabwe and its relation with $El - Niño/$ Southern Oscillation (ENSO), *Theory and Applied Climatology Journal*, 38, pp. 93-102.
12. Jabareen, Y. (2008). A new conceptual framework for sustainable development, *Environmental Development Sustainability*, 10, pp. 179-192.
13. Jury, M.R. (1996). Regional tele-connection pattern associated with summer rainfall over South Africa, Namibia and Zimbabwe, *International Journal of Climatology*, 16, pp. 135-153.
14. Kaunda, C.S., Kimambo, C.Z. and Nielson, T.K. (2012). Potential of Small-Scale Hydropower for Electricity Generation in Sub-Saharan Africa, *International Scholarly Research Network*, Volume 2012, pp. 1-15.
15. Laio, F., Di Baldassarre, G. and Montanari, A. (2012). Model selection techniques for the frequency analysis of hydrological extremes, *Water Resources Research*, 45 (W07416), pp. 1-11.
16. Lele, S.M. (1991). Sustainable development: A critical Review, *World Development*, 19(6), pp. 607-621.
17. Maiyaki, A.A. (2010). Zimbabwe's Agricultural Industry, *African Journal of Business Management*, 4(19), pp. 4159-4166.
18. Makarau, A. and Jury, M.R. (1997). Predictability of Zimbabwe Summer Rainfall, *International Journal of Climatology*, 17, pp. 1421-1432.
19. Manatsa, D. Chingombe, W. Matsikwa, H. and Matarira, C.H. (2008). The superior influence of the Darwin Sea Level Pressure anomalies over ENSO as a single drought predictor in Southern Africa, *Theory and Applied Climatology Journal*, 92, pp. 1-14.
20. Mangoyana, S. and Meda, T. (2001). *Food Security and Sustainability in Zimbabwe*. University of Zimbabwe Publications, Harare.
21. Mason, S.J. (1997). Recent changes in El Nino-Southern Oscillation events and their implications for southern African climate, *Transaction Royal Society South Africa*, 52, pp. 377- 403.
22. Matarira, C.H. and Unganai, L.S. (1994). *A rainfall prediction scheme for southern Africa based on the southern oscillation*, Technical report, SADC – REWU, Harare.
23. Panu, U.S. and Sharma, T.C. (2002). Challenges in drought research: some perspectives and future directions, *Hydrological Science Journal*, 47(S).
24. Perman, R., Ma, Y., McGilvray, J. and Common, M. (2003). *Natural Resource and Environmental Economics* 3rd edition, New York: Pearson Addison Wesley.
25. Peterson, T.C., Vose, R.S., Schmoyer, R. and Razuvaev, V. (1998). Global Historical Climatology Network quality control of monthly temperature data, *International Journal of Climatology*, 18, pp. 1169-1179.
26. Richard, Y., Trzaska, S., Roucou, P. and Rouault, M. (2000). Modification of the southern African rainfall variability / ENSO relationship since the 1960s, *Climatology Dynamics*, 16, pp. 883-895.
27. Rocha, A.M.C. (1992). The Influence of Global Sea Surface Temperatures on Southern African Summer Climate, PhD thesis, University of Melbourne.

28. Ropelewki, C.F. and Halpert, M.S. (1987). Global and regional scale precipitation patterns associated with the ElNiño/Southern Oscillation, *Mon Weather Review*, 115, pp. 1606-1625.
29. Sachs, J. and Warner, A. (1997). Sources of Slow Growth in African Economies, *Journal of African Economics*, 6(3), pp. 335-376.
30. Torrance, J.D. (1990). The Southern Oscillation and the rainy season in Zimbabwe, *Zimbabwe Science News*, 24, pp. 4-6.
31. Trück, S. and Liang, K. (2012). Modelling and forecasting volatility in the gold market, *International Journal of Banking and Finance*, 9(1), pp. 48-80.
32. United Nations. (2016). *Drought caused by El Nino threatening Southern Africa*, Sydney: United Nations.
33. United Nations. (2015). *Integrating the three dimensions of sustainable development*, Sydney: United Nations.
34. Waylen, P. and Henworth, P. (1996). A note on the timing of precipitation variability in Zimbabwe as related to the Southern Oscillation, *International Journal of Climatology*, 16, pp. 1137-1148.
35. Webster, P.J. (1981). Mechanism in determining the atmospheric response to sea level temperature anomalies, *Journal of Atmospheric Science*, 38, pp. 554-571.
36. Willmott, C.J. and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, *Climate Research*, 30, pp. 79-82.
37. http://www.cpc.ncep.noaa.gov/data/indices/da

## Appendix

Table 1. ACF and PACF correlogram of square residuals for the weighted regression model

<table>
<tr><td colspan="3">Included observations: 103</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Autocorrelation</td><td>Partial correlation</td><td></td><td>ACF</td><td>PACF</td><td>Q-Stat</td><td>Prob</td></tr>
<tr><td>.|** |</td><td>.|** |</td><td>1</td><td>0.230</td><td>0.230</td><td>5.6017</td><td>0.018</td></tr>
<tr><td>.|. |</td><td>*|. |</td><td>2</td><td>-0.031</td><td>-0.088</td><td>5.7036</td><td>0.058</td></tr>
<tr><td>.|* |</td><td>.|* |</td><td>3</td><td>0.089</td><td>0.124</td><td>6.5582</td><td>0.087</td></tr>
<tr><td>.|. |</td><td>*|. |</td><td>4</td><td>-0.054</td><td>-0.118</td><td>6.8717</td><td>0.143</td></tr>
<tr><td>*|. |</td><td>.|. |</td><td>5</td><td>-0.082</td><td>-0.025</td><td>7.6094</td><td>0.179</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>6</td><td>-0.004</td><td>0.000</td><td>7.6114</td><td>0.268</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>7</td><td>0.004</td><td>0.013</td><td>7.6132</td><td>0.368</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>8</td><td>-0.002</td><td>0.002</td><td>7.6137</td><td>0.472</td></tr>
<tr><td>.|* |</td><td>.|* |</td><td>9</td><td>0.089</td><td>0.089</td><td>8.5296</td><td>0.482</td></tr>
<tr><td>.|* |</td><td>.|* |</td><td>10</td><td>0.119</td><td>0.075</td><td>10.176</td><td>0.425</td></tr>
<tr><td>.|. |</td><td>*|. |</td><td>11</td><td>-0.033</td><td>-0.074</td><td>10.305</td><td>0.503</td></tr>
<tr><td>*|. |</td><td>*|. |</td><td>12</td><td>-0.104</td><td>-0.088</td><td>11.581</td><td>0.480</td></tr>
<tr><td>*|. |</td><td>.|. |</td><td>13</td><td>-0.085</td><td>-0.064</td><td>12.459</td><td>0.490</td></tr>
<tr><td>*|. |</td><td>.|. |</td><td>14</td><td>-0.078</td><td>-0.024</td><td>13.202</td><td>0.511</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>15</td><td>-0.023</td><td>0.020</td><td>13.269</td><td>0.582</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>16</td><td>-0.046</td><td>-0.061</td><td>13.529</td><td>0.634</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>17</td><td>-0.027</td><td>-0.009</td><td>13.624</td><td>0.694</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>18</td><td>-0.029</td><td>-0.054</td><td>13.734</td><td>0.746</td></tr>
<tr><td>*|. |</td><td>*|. |</td><td>19</td><td>-0.069</td><td>-0.070</td><td>14.352</td><td>0.763</td></tr>
<tr><td>*|. |</td><td>*|. |</td><td>20</td><td>-0.098</td><td>-0.088</td><td>15.612</td><td>0.740</td></tr>
<tr><td>*|. |</td><td>.|. |</td><td>21</td><td>-0.074</td><td>-0.021</td><td>16.341</td><td>0.750</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>22</td><td>-0.054</td><td>-0.012</td><td>16.731</td><td>0.778</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>23</td><td>-0.021</td><td>0.009</td><td>16.793</td><td>0.819</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>24</td><td>0.018</td><td>0.003</td><td>16.838</td><td>0.855</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>25</td><td>0.008</td><td>-0.022</td><td>16.847</td><td>0.887</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>26</td><td>0.027</td><td>0.024</td><td>16.952</td><td>0.911</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>27</td><td>-0.018</td><td>-0.050</td><td>17.001</td><td>0.931</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>28</td><td>-0.036</td><td>-0.014</td><td>17.187</td><td>0.945</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>29</td><td>-0.040</td><td>-0.028</td><td>17.420</td><td>0.955</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>30</td><td>-0.038</td><td>-0.012</td><td>17.636</td><td>0.964</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>31</td><td>0.010</td><td>0.013</td><td>17.650</td><td>0.974</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>32</td><td>0.070</td><td>0.039</td><td>18.402</td><td>0.974</td></tr>
<tr><td>.|. |</td><td>*|. |</td><td>33</td><td>-0.033</td><td>-0.104</td><td>18.573</td><td>0.980</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>34</td><td>-0.026</td><td>-0.025</td><td>18.679</td><td>0.985</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>35</td><td>0.006</td><td>-0.032</td><td>18.683</td><td>0.989</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>36</td><td>-0.033</td><td>-0.029</td><td>18.861</td><td>0.992</td></tr>
</table>