Modelling the SOI volatility through the Wishart distribution

Andrehette Verster

Daan De Waal

Abstract

One of the main objectives of the Eskom Risk Laboratory in the Department of Mathematical Statistic and Actuarial Science at the University of the Free State is to model and predict the rainfall in Bloemfontein. It has been shown that the rainfall in Bloemfontein greatly influence the water that flows into the Gariep dam. The water that flows into the Gariep dam is used by ESKOM for power generation purposes and the modelling of water inflow into the dam is therefore also one of the main interests of the Risk Laboratory. The rainfall in Bloemfontein is again influenced greatly by the Southern Oscillation Index of October month. It has been shown that if the SOI of October is incorporated into a hierarchical model the rainfall can be predicted which means the inflow can be predicted. The problem however remains that the SOI is quite volatile and the variation of the SOI needs to be taken into account before one can model the SOI to incorporate it into a hierarchal model to predict rainfall. In this paper we focus primarily on a method to model the SOI volatility.

Keywords: southern oscillation index, volatility, Wishart distribution, matrix Beta Type II

1 Introduction

It was shown in the ESKOM progress report of the Eskom Risk Laboratory, Department of Mathematical Statistic and Actuarial Science, University of the Free State in 2010 (available on request) that the October Southern Oscillation Index (SOI) is an indicator of rainfall in Bloemfontein. The SOI is computed from fluctuations in the surface air pressure difference between Tahiti and Darwin, Australia, The correlation between the October SOI of the previous year for the period 1970 – 2010 and the total Bloemfontein rainfall of the following year for the period 1971-2011from January to March (the rain season) is 0.3078. Figure 1 shows the linear fit of the October SOI against the total January to March rainfall.

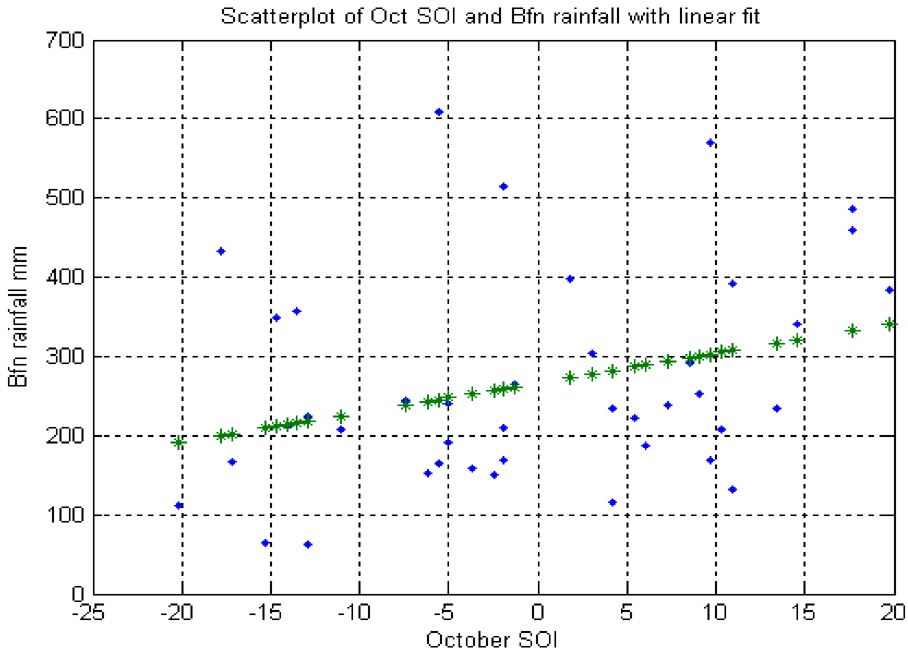

Figure 1: Scatter plot of the previous year's October SOI against the total January to March rainfall, of the following year, for Bloemfontein

In the ESKOM progress report of the Eskom Risk Laboratory, Department of Mathematical Statistic and Actuarial Science, University of the Free State in 2011 (available on request) Verster and De Waal showed that a hierarchical model, given below, can be used to predict the total rainfall in Bloemfontein by incorporating the October SOI of the previous year

$$f_{x,z}(x, z|\, k_x, \beta_x) = f_{x|z}(x|z,\sigma_x,k,\beta,\hat{\beta}_0,\hat{\beta}_1)f_z(z|\hat{\mu}_z,\hat{\sigma}_z). \tag{1}$$

X denotes the total rainfall in Bloemfontein from January to March, 1971 to 2011, and Z denotes the October SOI of the previous year, 1970 to 2010. The conditional density in Equation (1) given by

$$\mathrm{f}_{\mathrm{x|z}}(x) = \frac{2\Gamma\left(\frac{1}{2}+k\right)}{\beta\sqrt{\pi}\Gamma(k)}\left(x^{\frac{2}{\beta}}\right)^{\frac{1-\beta}{2}}(1+x^{\frac{2}{\beta}})^{-(\frac{1}{2}+k)},\ x>0 \tag{2}$$

denotes the Generalized t density with $x = |x - (\beta_0 + \beta_1 \mathrm{z})|/\sigma_x$ and $\mathrm{f_z}$ denotes a Normal density, $N(\mu_Z, {\sigma^2}_Z)$.

The posterior predictive density of the total Bloemfontein rainfall given an October SOI of, for example, 20 can be constructed as shown in Figure 2. The mean of the density is an estimate of the total Bloemfontein rainfall given an October SOI of 20. The estimated mean is this example was 429.31 mm.

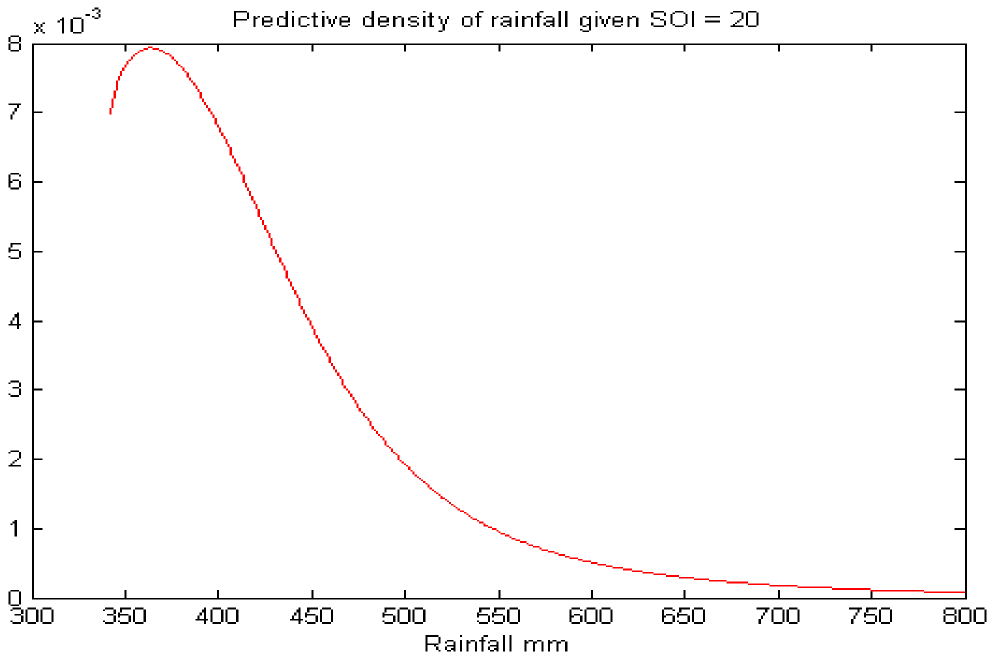

Figure 2: Predicted total January to March rainfall posterior density given an October SOI of 20

2 Predicting SOI variation

It is evident from the introduction that the SOI plays an important role in the prediction of rainfall in Bloemfontein. It is therefore important to model the SOI appropriately so that it can be incorporated into a hierarchical model to predict rainfall. To be able to model the SOI accurately the volatility and variation of the SOI need to the taken into consideration. This paper focuses mainly on the modelling of the SOI volatility that needs to be taken into consideration before any predictions can be made on the SOI. The following figure, taken from the internet, gives an idea of the variation one can expect in the sea-surface temperature.

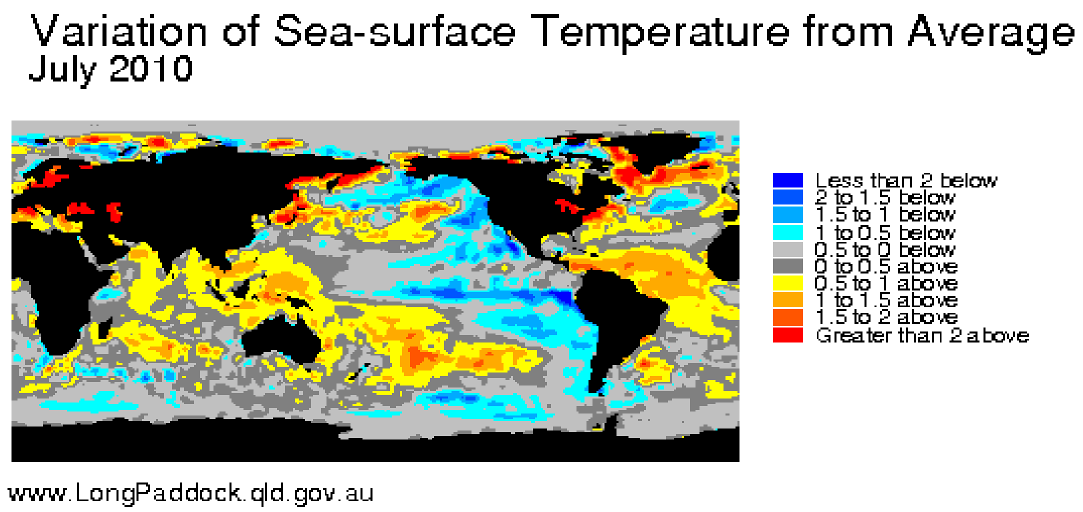

Figure 3: The variation of Sea-surface temperature for July 2010

Consider the Southern Oscillation Index (SOI) for October to February, say $\boldsymbol{X}(1{,}5)$. We assume that $\boldsymbol{X}$ is distributed multivariate normal, $N(\underline{0}, \Sigma)$. A large SOI data set is available at http://www.bom.gov.au/climate/current/soihtm1.shtml . Figure 4 shows the histograms of the SOI for the five months from 1876 to 2009. From the histograms of the SOI values for each month, the assumption of normality seems acceptable.

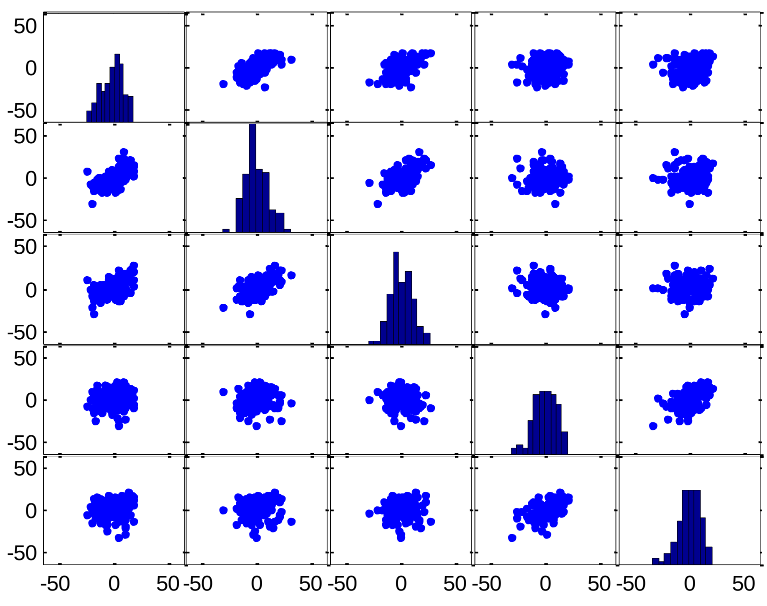

Figure 4: Histograms of the SOI values obtained from the SOI dataset for each of the five months

We now consider modelling the volatility of the SOI through a Wishart distribution. The Wishart distribution arises as the distribution of the sample covariance matrix for a sample from a Multivariate Normal. The Wishart distribution is defined as follows: If the columns of **X**(p,n) is distributed independently multivariate Normal(0, Σ), then the joint density of the $\frac{1}{2}p(p+1)$ elements of the symmetric positive definite random matrix $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{X}'$ is distributed Wishart with n degrees of freedom. The density function of $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{X}' \sim Wishart(n, \Sigma)$ is

$$f(A) = \frac{1}{\Gamma_p\left(\frac{n}{2}\right)}|2\Sigma|^{-\frac{n}{2}} etr\left(-\frac{1}{2}\Sigma^{-1}A\right)|A|^{\frac{1}{2}(n-p-1)},\ A > 0 \tag{3}$$

where $etr(.)$ denotes the exponent of the trace of the matrix and $\Gamma_p(\cdot)$ denotes the multivariate gamma function (Wishart, 1928 and Anderson, 1958).The mean of **A** is $n\Sigma$.

The data on the SOI for 1876-2009 is used to estimate Σ. Since it is a large dataset, 135 years, we assume that the estimate is the true Σ. Since we know that $\mathrm{E}(\boldsymbol{A}) = n\Sigma$, an estimate of Σ is $\hat{\Sigma} \approx \frac{\boldsymbol{A}}{n} = \frac{\boldsymbol{A}}{135}$. Σ was estimated as:

$$\hat{\Sigma} = \begin{bmatrix} 95.7963 & 64.2699 & 50.3629 & 54.2799 & 52.5683 \\ 64.2699 & 100.9013 & 56.3891 & 56.3479 & 59.2956 \\ 50.3629 & 56.3891 & 90.6804 & 57.4730 & 65.1136 \\ 54.2799 & 56.3479 & 57.4730 & 102.931 & 58.4602 \\ 52.5683 & 59.2956 & 65.1136 & 58.4602 & 99.6153 \end{bmatrix}.$$

Suppose we are interested in the distribution of the total variation over the 5 months, thus that of $tr\boldsymbol{A}$ if $n = 10$ years are considered where $\boldsymbol{A}$ is distributed Wishart(Σ, n). From the diagonal elements of Σ, we notice that the variances differ for the various months. January especially shows much larger variation than the other months. The exact distribution of $tr\boldsymbol{A}$ is given by the following equation

$$f(w) = \frac{1}{\Gamma(\frac{1}{2}np)}|2\Sigma|^{-\frac{1}{2}n} w^{\frac{1}{2}np-1} \sum_{k=0}^{\infty} \sum_K \frac{(\frac{1}{2}n)_K}{(\frac{1}{2}np)_k} \frac{w^k}{k!} C_K(-\tfrac{1}{2}\Sigma^{-1}), n > p,\ w > 0. \tag{4}$$

Equation (4) is calculated according the MATLAB algorithm mhg.m written by Koev and Edelman. The calculation of Equation (4) can be difficult and time consuming since it consists of hyper geometric functions and zonal polynomials. Equation (4) can easily be simulated by simulating a number of Wishart matrices and in each case calculating the trace of the Wishart matrix. The following figure shows a simulation of Equation (4) by simulating 500 Wishart $(\Sigma, 10)$ matrices and calculating their traces. The mean of the simulated total variation is $\overline{tr(A)} = 4890.1$ and the true mean is $ntr(\Sigma) = 4899.2$.

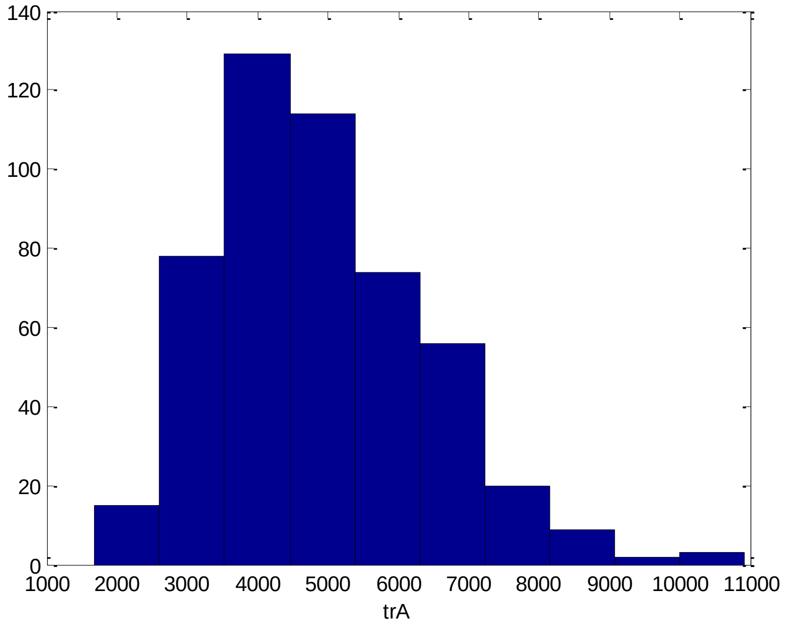

Figure 5: Histogram of the trace of 500 simulated Wishart $(\Sigma, 10)$ matrices.

3 Predicting a future Wishart matric A

In this section we discuss the prediction of future SOI volatility by predicting a future Wishart matric with a Bayesian approach. The Jeffreys prior on Σ^{-1} is $\pi(\Sigma^{-1}) \propto |\Sigma|^{\frac{1}{2}(p+1)}$ (Geisser, 1993, p. 192). The posterior distribution of Σ^{-1} is Wishart(nm, A^{-1}) given $A_1, \dots, A_m$ with density function

$$\pi(\Sigma^{-1}|A) \propto |\Sigma^{-1}|^{\frac{1}{2}(nm-p-1)} etr(-\tfrac{1}{2}\Sigma^{-1}A),\ \Sigma^{-1} > 0 \tag{5}$$

where $A = \sum_{i=1}^{m} A_i$. The posterior predictive density of a future A_{m+1} is then

$$pred(A_{m+1}|A) \propto \int_0^\infty f(A_{m+1}|A, \Sigma^{-1})\ \pi(\Sigma^{-1}|A)d\Sigma^{-1}$$

$$\propto |A_{m+1}|^{\frac{1}{2}(n-p-1)}|I + A^{-\frac{1}{2}}A_{m+1}A^{-\frac{1}{2}}|^{-\frac{1}{2}(mn+n)}\ . \qquad (6)$$

Let $V = A^{-\frac{1}{2}}A_{m+1}A^{-\frac{1}{2}}$. Equation (6) can then be written as a matrix Beta type II density given as follows:

$$g(V|A) = \frac{\Gamma_p(\frac{n+mn}{2})}{\Gamma_p(\frac{n}{2})\Gamma_p(\frac{mn}{2})}|V|^{\frac{1}{2}(n-p-1)}|I+V|^{-\frac{mn+n}{2}},\ V > 0. \qquad (7)$$

with parameters $\frac{n}{2}$ and $\frac{mn}{2}$ (de Waal, 1969).

3.1 Illustration

We start by constructing $m = 10$ covariance matrices $A_i, i = 1, \dots, 10$ from the October to February SOI over $n = 10$ years (1900 – 1999). A future matric A_{11} is predicted by $\frac{n}{nm-p-1}A$ as follows:

$$\hat{A}_{11} = \frac{10}{100-5-1}A = 10^3\begin{bmatrix} 1.0114 & 0.7028 & 0.6006 & 0.5373 & 0.5661 \\ 0.7028 & 1.0938 & 0.6103 & 0.5954 & 0.6185 \\ 0.6006 & 0.6103 & 0.9625 & 0.6068 & 0.6820 \\ 0.5373 & 0.5954 & 0.6068 & 1.0400 & 0.6228 \\ 0.5661 & 0.6185 & 0.6820 & 0.6228 & 1.0418 \end{bmatrix}, \text{ where}$$

$A = \sum_{i=1}^{10} A_i$. The true future A_{11} from the given data is

$$A_{11} = 10^3\begin{bmatrix} 0.8142 & 0.6821 & 0.6063 & 0.6100 & 0.7200 \\ 0.6821 & 1.2570 & 0.7505 & 0.5952 & 1.2421 \\ 0.6063 & 07505 & 0.9720 & 0.3599 & 1.0939 \\ 06100 & 0.5952 & 0.3599 & 0.7561 & 0.5156 \\ 0.7200 & 1.2421 & 1.0939 & 0.5156 & 2.0227 \end{bmatrix}.$$

From the two matrices one can see that the predicted February variation is much smaller than the true February variation.

A future Wishart matrix can also be predicted through simulation. Simulate a large number of $A_{m+1} \sim Wishart(\Sigma, n)$. For each simulated A_{m+1}, $V = A^{-\frac{1}{2}}A_{m+1}A^{-\frac{1}{2}}$ is obtained. The $V's$ obtained are substituted into the density function of the MBET2 [Eq. (7)]. 7000 of the $V's$ are then drawn from the MBET2 density and transformed to A_{m+1} as for example given below

$$\hat{A}_{m+1} = 10^3 \begin{bmatrix} 1.0598 & 0.7311 & 0.5738 & 0.6786 & 0.6091 \\ 0.7311 & 1.0873 & 0.6465 & 0.6458 & 0.6622 \\ 0.5738 & 0.6465 & 0.9670 & 0.6761 & 0.7022 \\ 0.6786 & 0.6458 & 0.7608 & 1.2456 & 0.6792 \\ 0.6091 & 0.6622 & 0.7022 & 0.6792 & 1.0928 \end{bmatrix}.$$

The results of the total variance (tr(A_{m+1})) and the average monthly variance $\left(\frac{tr(A_{m+1})}{5}\right)$ for the predicted future Wishart and the true future Wishart is given in the table below.

Table 1 Total variance and average monthly variance for the predicted future Wishart and the true future Wishart

	$\hat{A}_{11} = \frac{10}{100-5-1}A$	Predicted through Simulation	Observed (True)
Total variance (tr(A_{11}))	5149.4	5452.6	5822
Average monthly variance $\left(\frac{tr(A_{11})}{5}\right)$	1029.9	1090.5	1164.4

Once again the predicted variation is smaller than the true variation. To understand this outcome we look at the moving total variation average. The moving total variation average, $tr(A)/5$, for October to February for the years 1902 – 2011 is shown in Figure 6. From Figure 6 one can see that the variation has drastically increased over the last 40 years. This pattern is not clear from Figure 7 that shows the moving average on the SOI values for the 5 months separately during the same period.

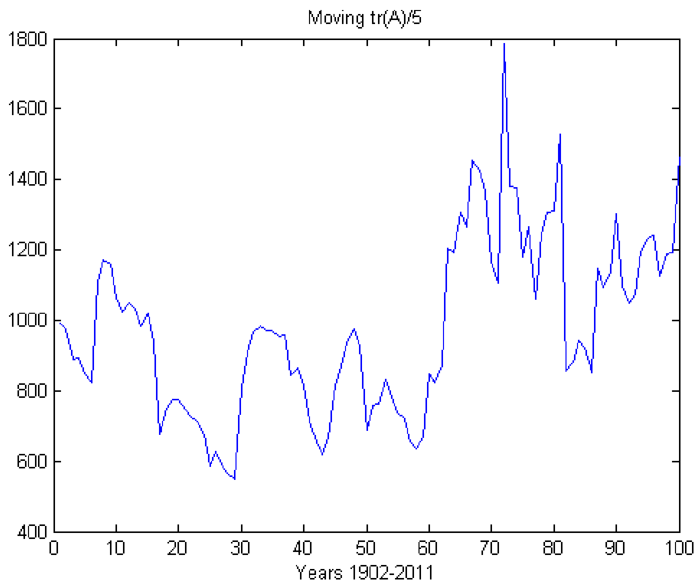

Figure 6: The moving total variation average, $tr(A)/5$, for October to February for the years 1902 – 2011

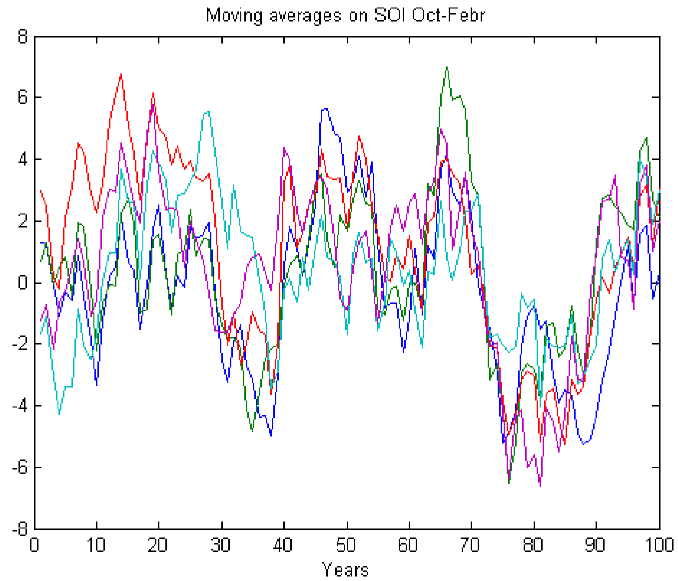

Figure 7: The moving average on the SOI values for the 5 months separately for October to February for the years 1902 – 2011

4 Conclusion

As mentioned in the introduction the SOI is a valid indicator of rainfall in Bloemfontein. The SOI can be incorporated into a hierarchical model that can be used to predict future rainfall. It is therefore important to model the SOI appropriately before rainfall can be predicted effectively. The SOI is predictable but the variation (volatility) of the SOI should be taken into consideration when modelling SOI. We have shown in this study that the volatility of the SOI can be modelled through the Wishart distribution.

5 References

Anderson TW (1958). An introduction to multivariate statistical analysis. Wiley.

Congdon P (2006). Bayesian Statistical Modelling, Second Edition. Wiley, England.

De Waal DJ (1969). Nie-sentrale Meerveranderlike Beta-verdeling. Ph.D proefskrif.

Geisser S (1993). Predictive inference, an introduction. Chapman & Hall.

Koev P and Edelman A (2006). Computation of hypergeometric functions of matrix argument in Matlab. *Mathematics of Computation*, 75, pp.833-846.

Wishart J (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* 20A, pp. 32-52.