Nonlinear Statistical Methods for Climate Forecasting

Edward P. Campbell

CSIRO Mathematical and Information Sciences

Bryson C. Bates and Stephen P. Charles

CSIRO Land and Water

Report of Second Research Phase

for the

Indian Ocean Climate Initiative

CONTENTS

***** - Contains material of a technical statistical nature- please see *Suggested Reading* at the end of the summary.

LIST OF TABLES

LIST OF FIGURES

Summary

We summarise in this report the Indian Ocean Climate Initiative's (IOCI) work to date to develop a statistical procedure for modelling nonlinear climate phenomena. This procedure is motivated by the physical idea of a switching mechanism that causes different climate states to prevail for some period of time. There is a range of evidence to support these ideas, much of which was discussed in the Phase I report, and will be further discussed here. A simple example is the well-known phenomenon in the tropics where enhanced convection can occur when sea surface temperatures are greater than about 29°C, so different rainfall regimes apply either side of this threshold temperature.

A key feature of the procedure that we have developed is that it is not based on the typically strong statistical assumptions needed to perform a text book analysis. The statistical assumptions required are instead implied by the physical paradigm. We call this procedure a Bayesian threshold model, for reasons that will become clear below.

At the conclusion of this phase of IOCI we have developed and tested the Bayesian threshold model using monthly rainfall data at Rottnest Island and Manjimup. This represents the diversity of Southwest WA to some extent; Rottnest Island has been included in particular because there seems to be little scope for land-use change to have impacted rainfall patterns at this location. This has required some statistical research to underpin the Bayesian threshold model, and the detailed development is contained in a manuscript that has been submitted for publication. The manuscript is attached in Appendix C to this report. The method is relatively easy to use and can identify important predictors and the key lags at which they act to influence a climate variable, such as rainfall. It is possible to examine the impact of different switching variables and identify the likely number of thresholds in the switching variable, although this is the subject of ongoing work as it is currently somewhat ad hoc.

Whilst the main purpose of the case studies to date has been to support the development of the Bayesian threshold model, some interesting physical links have been observed. We cite them as evidence at this stage that the model is behaving as we would expect, rather than providing great insights. Searching for leading rainfall indicators is necessarily collaborative in nature, not simply a statistical modelling exercise, but can now incorporate a new nonlinear statistical tool.

Clear indications of switching behaviour have been found in the rainfall time series. For Rottnest Island no connection with monthly rainfall and the El Niño-Southern Oscillation (ENSO), as measured by the Southern Oscillation Index (SOI), has been found. A seemingly strong link has been found at Manjimup however. We have also examined the use of sea surface temperature gradient in the mid-Indian Ocean as a switching variable. There is evidence that it plays a role in switching rainfall regime at both sites considered, but the relationship appears to be especially strong at Rottnest Island, providing a leading indicator of winter rainfall.

If we conceive of a 'true' switching mechanism, it seems unlikely that this will be a very simple process. Physical intuition suggests that a *combination* of patterns in the Indian, Southern and Pacific Oceans is more likely to cause a switch in rainfall regime. This is because the climate system is driven by interaction between oceans as well as the oceans

and the atmosphere. Preliminary work suggests that the Bayesian switching model can readily be adapted to a more general framework that will facilitate the search for such climate interactions. We intend to pursue this as time permits.

Our key priority for the last phase of IOCI is to apply the Bayesian switching model, and our statistical expertise more generally, to a range of case studies developed in collaboration with the contributing partners.

Key Points:

- We have developed a physically motivated statistical model ('Bayesian switching model') for modelling nonlinear climate processes.
  - Changes between climate regimes are triggered by a switching variable, and alternative switching variables can be compared.
- The Bayesian switching model can identify good predictors and the lags at which they influence climate variables, such as rainfall.
- We have reached the point where a nonlinear time series approach can be applied to practical problems.
- There is some evidence that SOI and mid-Indian Ocean SST gradients play a role in switching between rainfall regimes. This is cited at this stage as evidence that the new nonlinear approach is producing sensible results, rather than new insights *per se*.
- Interactions between climate processes are likely to influence rainfall in Southwest WA. Some reasonably straightforward extensions to the Bayesian switching model will facilitate the search for subtler climate teleconnections arising from such interactions.
- The focus of future work will be the development of case studies with IOCI's contributing partners.

Suggested Reading

The work summarised in this report represents an overview of an statistical research effort. For most readers there is more technical detail than is necessary to understand the methods used and the progress made. The technical details are important for completeness however. We suggest two paths through this report:

For readers interested in the statistical research issues: Read in the order presented, but a first reading of Appendix C is appropriate after §2.

For readers not interested in the statistical research issues: §2.5 contains an overview of the technical material and so is optional but is of general interest; §3 is optional as it contain some substantial technical material; §4.2 and Appendix C should be skipped completely.

1 Introduction

Many advances in climate forecasting have been brought about by the statistical analysis of available data. These advances occur when the analysis of climate data poses questions that encourage us to new physical understanding; when we can achieve this then there is a solid foundation on which to build better climate forecasting systems. Given the complex, inter-related nature of the climate system this can be a very difficult process in practice.

The bulk of statistical climatology work reported in the literature (see: *Campbell et al., IOCI Phase I Report*) uses conventional statistical methods to explore for physical relationships. Such techniques often make strong assumptions about the nature of the physical systems being studied. For example, it is typically assumed that linear relationships exist between variables of interest and that the physical processes do not change their behaviour with time. The gap between what the statistical methods can deliver and the true nature of the physical systems being studied must be bridged by a good deal of intuition. This is unfortunate because it is very likely that many important physical questions are never posed because the statistical "searchlight" is inadequate.

IOCI is quite unique in that a statistical research capability is woven into the initiative. The objective of our work is to examine the nature of climate processes and to develop statistical methods appropriate for the analysis of data arising from such processes. Based on our work in Phase I it became clear that there is a need to develop methods that can model nonlinear climate phenomena.

We have undertaken to develop a statistical methodology that will also provide uncertainty measures for forecasts. That is, rather than just giving a rainfall estimate we will provide a probability distribution for a rainfall forecast. We have been using monthly rainfall to develop our methods because it is such a difficult quantity to model, but do not limit our scope to rainfall. It is our intention to identify relevant applications of the methods developed in partnership with IOCI's contributing partners.

In this report we document the work to date in the statistical method development. In section 2 we describe the physical rationale for the methods being developed; section 3 describes case studies of monthly rainfall at Rottnest Island and Manjimup that have been used to test the methods as they're developed. A discussion of our results to date is given in section 4 with some conclusions in section 5.

2 Year 3 Development Path

2.1. Probability Distributions for Forecasts

There are two key features of a climate forecast that a decision-maker in climate-impacted sector must balance in reaching their decision. First, the climate pattern forecasted and, second, the uncertainty associated with the forecast. Different decisions are required for different levels of uncertainty. For example, if a forecast is known to be highly accurate and very clearly forecasts boom conditions for wheat farmers say, then a sensible decision might be to expand wheat production. In a situation of greater uncertainty, with the same forecast climate pattern, to manage risk in a sensible fashion it is advisable to maintain a more balanced crop portfolio.

To provide a complete statement of uncertainty we need to integrate (assimilate) information from a variety of sources. These encompass uncertainties in available climate data, the forecast system used and the availability of expert knowledge. Such expert knowledge would typically include both a meteorological and a decision-maker’s perspective. A final integrated statement of uncertainty would be a probability distribution for a climate output- such as next year’s wheat production, to continue the above example.

The Bayesian statistical framework is ideal for integrating uncertainty information, so we have chosen to develop our methods within the Bayesian framework. The Bayesian approach begins with a statement of knowledge prior to the collection of data (“prior knowledge”). This information is expressed as a probability distribution, allowing us to specify quantities such as “most likely value,” “average value” etc. Our uncertainty can then be expressed via the spread of prior knowledge. The prior knowledge is then combined with the data via a mathematical rule known as Bayes’ Theorem to form an integrated expression of uncertainty posterior to data collection (“posterior knowledge”). This is illustrated heuristically in Figure 1. We see that the data have greatly reduced uncertainty, as the posterior is much more concentrated on a particular value than the prior.

In addition to providing a powerful scientific framework for drawing inferences from data there are also a number of technical advantages. In particular, in climate prediction we are most concerned with finding good predictors and the time lags at which these predictors influence climate. In comparison with more conventional statistical methods the Bayesian framework offers much more flexibility in identifying good predictors with fewer technical mathematical concerns.

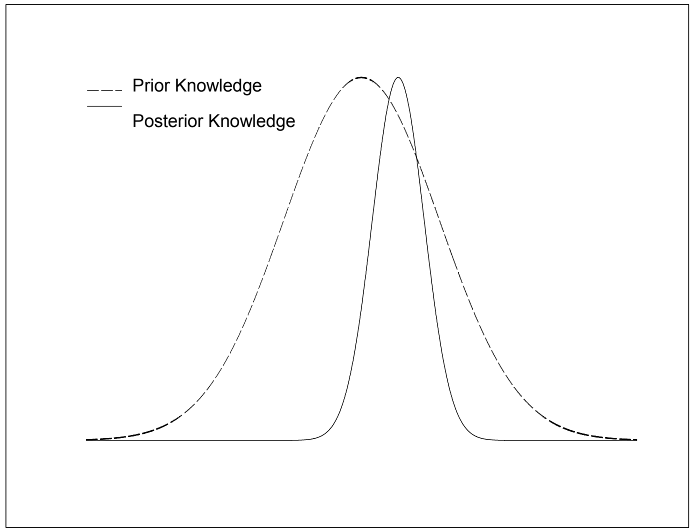


Figure 1 The Bayesian process of integrating data and expert opinion.

2.2. Climate Switching Models

Much of the statistical analysis undertaken in climate research uses so-called linear statistical techniques to identify climate processes. There is however significant evidence that the climate system can behave in strikingly nonlinear ways. For example, *Graham and Barnett* [1987] show that in the tropics enhanced convection occurs at Sea-Surface Temperatures (SSTs) greater than about 29ºC. This implies that different rainfall forecasting systems apply depending on whether SST is above or below the threshold temperature of 29ºC. In general there may well be a delay between the threshold SST being reached and the resulting switch in rainfall regime. In each rainfall regime we assume that different linear climate processes apply. This is depicted in Figure 2 with a linear approximation superimposed for reference.

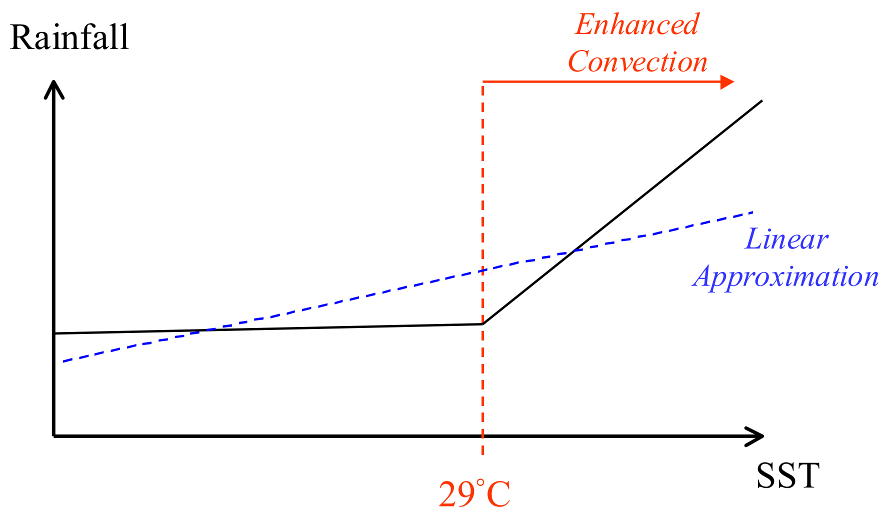


Figure 2 Nonlinear relationship between Rainfall and SST.

A similar result was found by *Hsieh et al.* [1999] in the context of predicting Canadian prairie wheat yield from Pacific Ocean SSTs. If a linear system applies then SST anomalies in low and high yield years should be of similar magnitude, but with opposite signs. Instead they found major asymmetries in SST anomalies and that only low yields are predictable from SST anomalies. If this information is ignored then a potentially misleading forecast system will result; at best it will have little skill.

Palmer [1999] examined climate prediction from a nonlinear perspective, providing theoretical justification for a climate system that resides in equilibrium states for periods of time, subject to occasional rapid switching between states. Palmer found some evidence in available data to support this view. The threshold SST idea described above is consistent with this view, with SST providing the switch between states.

There is therefore a strong argument for developing statistical models incorporating the concept of threshold behaviour. Such models will allow a more physically motivated analysis of available data than has hitherto been the case. The key research activity of the CSIRO Mathematical & Information Sciences (CMIS) group has been to develop a statistical method for identifying good predictors in a Bayesian nonlinear framework. The results of the work to date are described in section 3 where the monthly rainfall case studies are described. The theoretical work underpinning the case studies is described in the manuscript appended to this report.

2.3. Incorporating Ocean-Atmosphere Interactions

Rainfall arises from an interaction between the oceans and the atmosphere. This means that information on the atmosphere or oceans alone my not be sufficient to forecast climate; it may be necessary to have knowledge of both. In particular there may be combinations of conditions in the oceans and atmosphere that provide a leading indicator of enhanced rainfall or drought. It may also be the case that *combinations* of past ocean conditions are more important than individual SST values.

The interaction concept is a very powerful one. An example of a statistically significant interaction effect is shown in Figure 3. In this case we require SST to be high both 3 and 6 months prior to winter to experience high winter rainfall. Whilst SST 6 months ago is a leading indicator of rainfall it must be sustained until 3 months before winter to produce high winter rainfall.

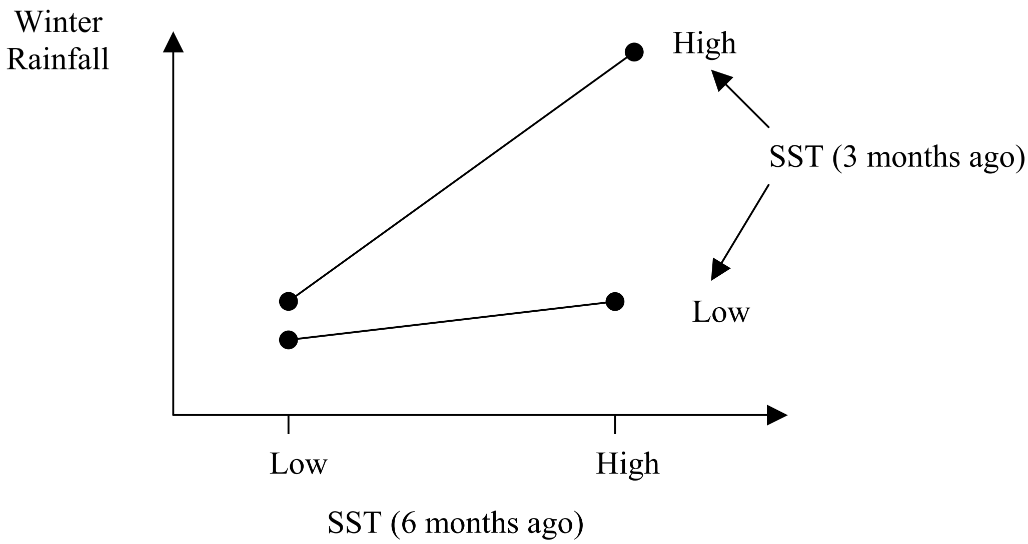


Figure 3 Example of an interaction effect in SST influencing rainfall.

Interaction effects of this type can be very difficult to detect in climate data. An approach to doing this is to include a product term "SST (-6 months) × SST (-3 months)" in the models we fit to the data. An approach to this problem is currently under development for application in the final stage of IOCI.

In the context of IOCI's work, Indian, Pacific and Southern ocean conditions as well as atmospheric conditions are of great relevance. Australia's island continent status is uniquely complex in this regard, so interactions between 3 oceans and the atmosphere are clearly relevant to climate prediction.

2.4. Modelling Approach

It is common practice to model deviations ("anomalies") from long term climate trends in preference to modelling the raw data. The main purpose for this is so that predictors beyond simple climatological averages can be sought. From a nonlinear perspective the calculation of anomalies carries the risk that important information might be lost. For example, it could well be that different climate regimes have different seasonal patterns. It is also the case that the method used to remove long term trends can introduce features of its own.

We have chosen not to model anomalies. Instead we seek to capture the seasonal pattern of, e.g. rainfall, by incorporating rainfall as the leading term in the model. Subsequent terms in the model, such as SST, will therefore only be included if they explain rainfall variation additional to the rainfall history itself. The modelling approach may be summarised as:

$$\text{Rain}\left(\text{today}\right) = \text{Rain}\left(\text{history}\right) + \text{SST}\left(\text{additional to rain history}\right).$$

This applies to each regime identified, so different terms can be selected in each regime.

2.5. Development of Methodology

Within a nonlinear, climate-switching framework we have identified a need to provide probability distributions for forecasts and to identify the time lags at which important predictors influence climate. These requirements present a challenging statistical problem. An approach developed by CSIRO Mathematical and Information Sciences is described in detail in the manuscript in Appendix C. This manuscript is still in peer review at the time of writing.

For the interested reader, the approach uses an extended Markov chain Monte Carlo (MCMC) approach known as Reversible Jump MCMC (RJMCMC). In MCMC we set-up a carefully defined random walk over the parameter space of a particular model in order to summarise the posterior distribution of the model's parameters. RJMCMC extends the random walk to range over a collection of models. In this way we can conduct model selection and parameter estimation simultaneously.

3 Case Studies

3.1. Description of Data

Monthly rainfall data (mm) for Rottnest Island and Manjimup were selected from Bureau of Meteorology's high quality data set. Manjimup has been selected as representative of an inland site in the Southwest, whilst Rottnest Island has been selected since it is free of any concerns regarding land-use change. For Manjimup we used monthly rainfall data from 1950 to 1993 for model fitting. In the case of Rottnest Island, rainfall data were only available until 1992 at the time of this study. Data from 1950 were used so that credible use of sea surface temperature could be made.

Some pre-processing of the rainfall data has been undertaken. First, rainfall data are typically highly skewed which can cause large rainfall events to have undue influence of the model-fitting procedure. In particular, linear time series models will tend to have an artificially high order in such circumstances. This could bias the comparison of linear with nonlinear models. To avoid this we have first transformed the rainfall data to be more nearly symmetric (using a Box-Cox transformation). As an aid to numerical stability all of the time series used were scaled to have mean 0 and standard deviation 1.

One of the switching variables we have used is the SST gradient at Point 27, 3 (Bureau of Meteorology naming convention) in the mid-Indian Ocean. This measures the north-south difference in SST at this point. This has been found by to have substantial correlation with Manjimup rainfall [Lynda Chambers, *pers. comm.*]. The Bureau of Meteorology supplied the SST gradient data used in our case study, as were the SOI data.

3.2. Manjimup Monthly Rainfall

3.2.1 Predicting Rainfall From The Rainfall History Only

As a first step we fit a threshold model to the monthly rainfall data alone, so we assume that a threshold exists in the rainfall time series rather than in SST. This is a reasonable starting point for investigating the performance of the threshold approach by seeking evidence of nonlinearity in the rainfall time series.

To fit a threshold model we must determine the number of thresholds that are present. In Figure 4 we present a diagnostic[1] for choosing the appropriate number of thresholds. On the vertical axis we plot a measure of the support[2] in the data for a threshold parameter taking a value on the horizontal axis. We expect to see a cloud of points with a noisy spike at a value where there is support for a threshold. In this case there is a distinct spike at a value of about –1.1. This suggests that there is evidence for a low rainfall regime and a normal to high rainfall regime. In the analysis to follow we assume the presence of 1 threshold.

[1] In practice a number of these plots are produced at different scales. The final figure is focused on the region in which the threshold appears to be located.

[2] The likelihood of the time series plotted as a function of the threshold parameter.

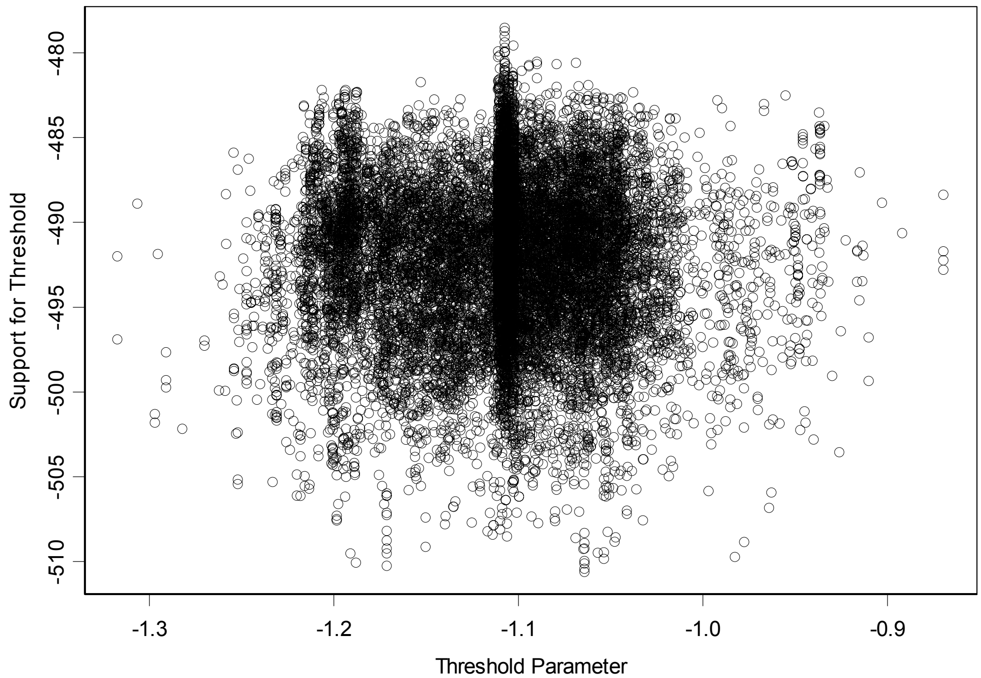


Figure 4 Diagnostic plot for choosing the number of thresholds and their approximate locations.

We can calculate some simple summaries of the resulting model fit. To get some insight into how well the procedure is performing we can use the resulting parameter estimates to calculate 1-month-ahead predictions for each observed monthly rainfall amount,[3] and the result is shown in Figure 5. We see that the correlation between the predicted and observed values is 0.782. This is a reasonable performance given that we have only used the rainfall history. In general the approach under-predicts the largest observed events. This is not unexpected, and it is likely that a predictor such as sea surface temperature will be required to predict such events.

[3] We must exclude a number of values at the beginning of the time series because the model uses past values of rainfall to forecast the future. This means that we cannot predict the first few values of the time series.

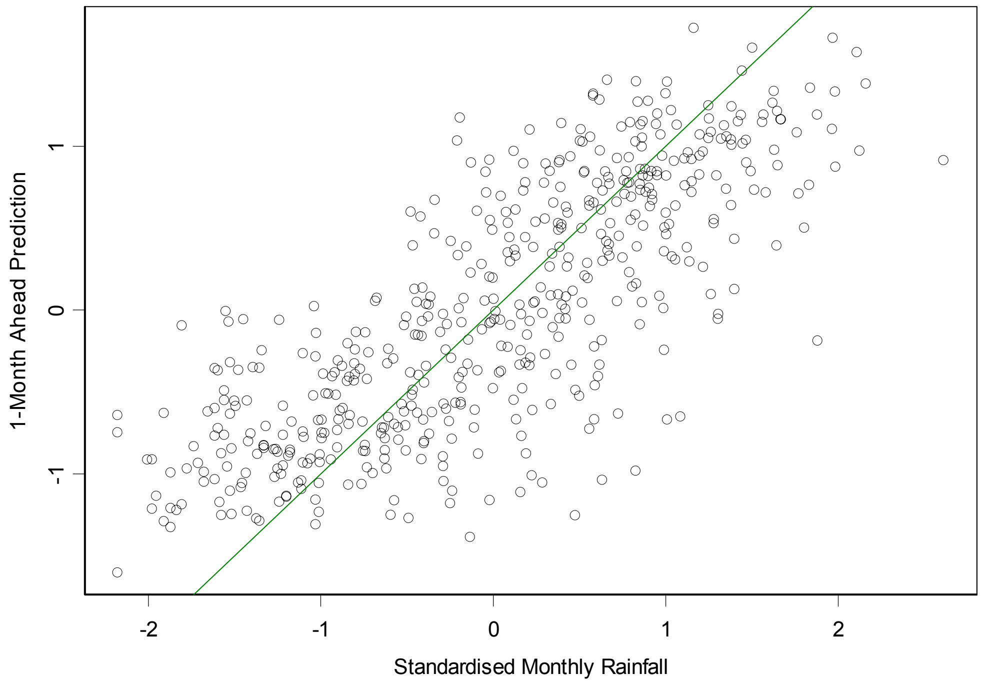


Figure 5 1-month-ahead predictions of monthly rainfall at Manjimup. The correlation between predicted and observed rainfall is 0.782.

An important feature of the method we are developing is that it identifies the important time lags essentially automatically. The parameter estimates are shown in Table 1. At this stage the method retains all lags up to the longest, so we present the statistically significant terms only; the estimated model orders correspond to the model having the highest probability. The credibility intervals quoted for individual parameters incorporate model uncertainty[4] by including competing models in the calculation, not just the model having highest probability. Notice first that the high and low rainfall regimes are quite different, having time lags of 6 and 16 months respectively. The high rainfall regime also has a relatively complicated structure. This is an important point: a linear fit may be able to produce a similar global fit, but is likely to perform poorly within a particular rainfall regime. The variance parameter is more clearly defined in regime 2 because there are more observations than in regime 1, which is quite typical of threshold models.

[4] Since we don't know the correct predictors to use we need to account for this important source of uncertainty.

	Parameter	Estimate	95% Credibility Interval
	Threshold	-1.11	-1.21, -1.01
Regime 1: 'Low' Rainfall	Order	6	4, 7
	Intercept	-0.261	-1.13, 0.639
	Lag-4	-0.332	-0.649, -0.0332
	Lag-6	-0.186	-0.510, 0.000
	Variance	0.54228	0.38445, 0.74765
Regime 2: 'Normal/High' Rainfall	Order	16	14, 17
	Intercept	-0.0188	-0.0843, 0.051
	Lag-1	0.232	0.126, 0.348
	Lag-4	-0.139	-0.231, -0.0457
	Lag-5	-0.0924	-0.180, -0.00445
	Lag-7	-0.105	-0.198, -0.00515
	Lag-11	0.179	0.0797, 0.273
	Lag-13	0.242	0.143, 0.340
	Lag-14	-0.0902	-0.190, 0.000
	Lag-16	-0.0756	-0.192, 0.000
	Variance	0.38454	0.33614, 0.44096

Table 1 Model-averaged parameter estimates with 95% credibility intervals. Only statistically significant parameters are shown, correct to 3sf except for variance parameters which are shown correct to 5sf.

For comparison with the threshold model we have also fitted a conventional linear time series model. The 1-month-ahead predictions are shown in Figure 6, with a correlation between predicted and observed values of 0.785 with an order 20 model. Notice that this does not include an intercept term, which we always include in the nonlinear model because such parameters can help to model highly volatile time series. Overall the number of parameters in this case is essentially the same, with the same global performance. Interestingly the order of the linear model is higher than for either of the rainfall regimes identified by the threshold model. This is in keeping with the concepts illustrated in Figure 2: in order to achieve the same overall quality of fit a higher order model is required. In reality there seems to be evidence for different rainfall regimes having different physical characteristics.

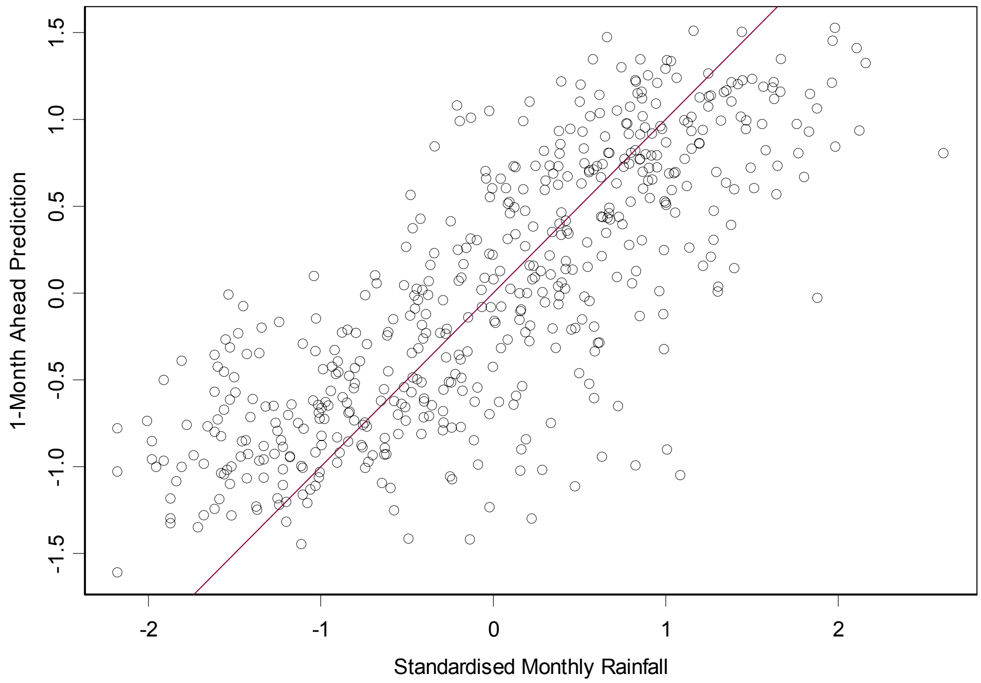


Figure 6 Results from fitting a conventional linear autoregressive model (order 20); the correlation between 1-month-ahead predictions and standardised monthly rainfall is 0.785.

3.2.2 Predicting Rainfall Using The Southern Oscillation Index (SOI) As A Switch

Nonlinear modelling of the rainfall time series in isolation has provided evidence for climate switching. However, it seems unlikely that rainfall itself is the cause of the switching behaviour. One possibility is the El Niño-Southern Oscillation (ENSO), as measured by the Southern Oscillation Index (SOI). However, it is likely that the impact of SOI on southwest WA rainfall will be delayed. We have used a delay of 1 month in the first instance to examine this issue.

A stable threshold seemed to be present at around +1.5, so we examined the threshold diagnostic plot in this vicinity (Figure 7). There is clear support for a threshold at approximately +1.7, with a rapid drop in support below +1.6 and above +1.8.

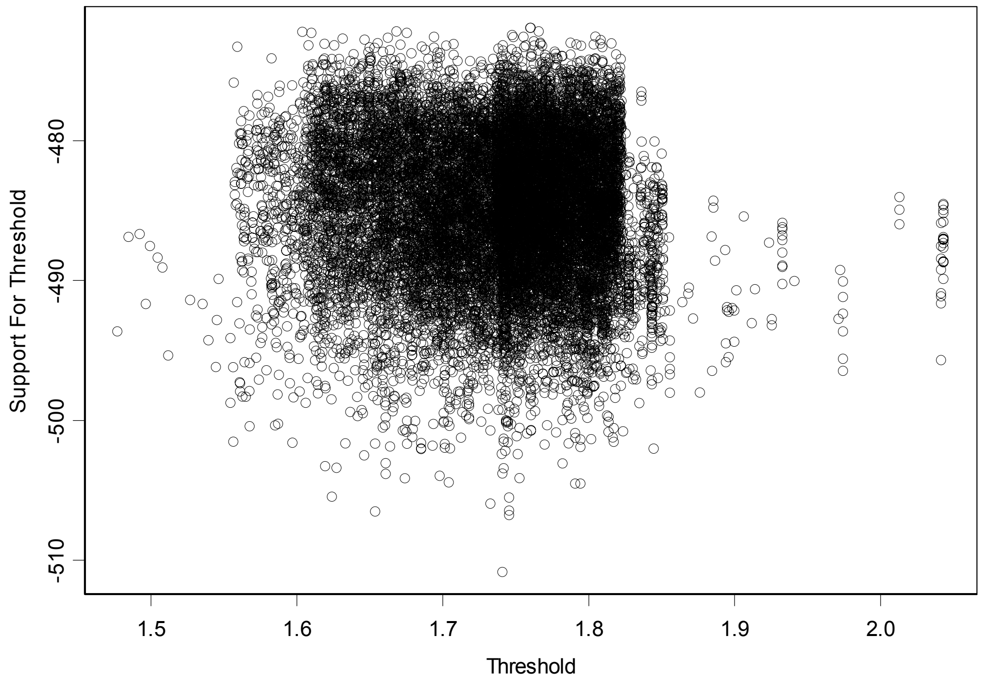


Figure 7 Diagnostic plot for choosing the number of thresholds and their approximate locations for Manjimup rainfall using SOI as the switching variable.

The 1-month ahead predictions obtained are shown in Figure 8; we see that there is a correlation of 0.788 between the observed and predicted values. The corresponding parameter values are shown in Table 2. The structure of the fitted model is remarkably similar to the rainfall-only model fitted above, particularly the 'Normal/High' rainfall regime. It is therefore very tempting to suggest that SOI plays a physical role in rainfall at Manjimup.

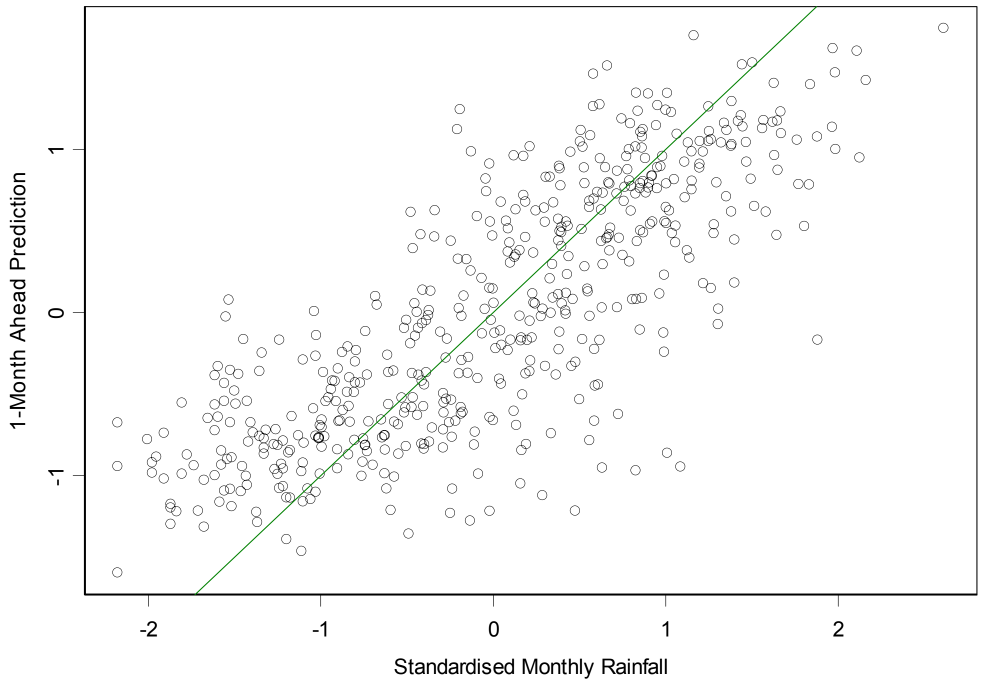


Figure 8 1-month-ahead predictions of monthly rainfall at Manjimup, using SOI as the switching variable. The correlation between predicted and observed rainfall is 0.788.

	Parameter	Estimate	95% Credibility Interval
	Threshold	1.74	1.61, 1.82
Regime 1: 'Low' SOI	Order	16	14, 17
	Intercept	-0.00965	-0.0639, 0.0454
	Lag-1	0.203	0.114, 0.295
	Lag-4	-0.163	-0.254, -0.0731
	Lag-7	-0.0918	-0.183, -0.00469
	Lag-11	0.150	0.0566, 0.240
	Lag-13	0.234	0.146, 0.326
	Lag-14	-0.0943	-0.181, -0.000874
	Lag-16	-0.114	-0.210, 0.000
	Variance	0.39108	0.34085, 0.44835
Regime 2: 'High' SOI	Order	5	1, 7
	Intercept	0.158	-0.360, 0.587
	Lag-4	0.505	0.000, 1.24
	Lag-5	-0.364	-1.00, 0.000
	Variance	0.65143	0.25265, 1.5537

Table 2 Model-averaged parameter estimates with 95% credibility intervals for Manjimup rainfall using SOI as the switching variable. Only statistically significant parameters are shown, correct to 3sf except for variance parameters which are shown correct to 5sf.

In this case it seems that low rainfall in Manjimup is associated with large positive values of SOI, which is an indicator of El Niño events (Figure 9), although there is a clear exception to this rule. The influence of ENSO on Northern and Eastern Australia is well known, but is a less recognised influence on the climate of Western Australia. On the evidence of this analysis it seems that quite extreme El Niño events can influence rainfall in Manjimup- the threshold for SOI being at +1.74 standard deviations to cause a low rainfall regime to be initiated. This does mean however that relatively few observations are available to characterise this regime.

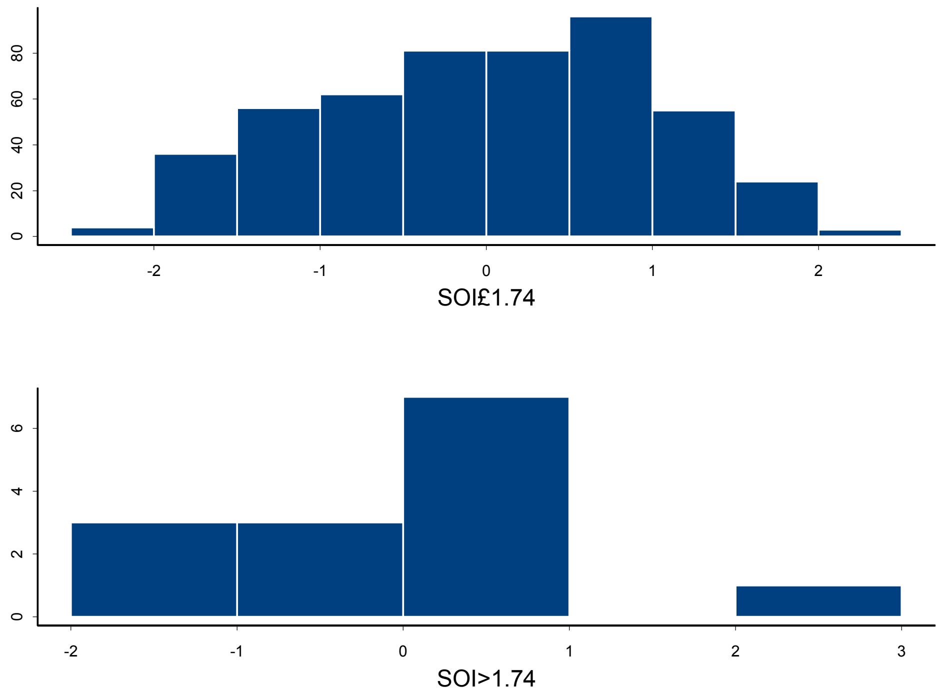


Figure 9 Distribution of Manjimup rainfall for the two SOI regimes.

The dates and rainfall amounts corresponding to the 'High SOI' regime are shown in Table 3, with runs of consecutive months highlighted. The exceptionally high rainfall event in the high SOI regime corresponds to August 1955. There is a run of high SOI values in April-May 1971 and August-October 1975. In general the events are scattered through the calendar year with all seasons represented; the months not present are March, June, July and November. It is interesting to note that the core winter rainfall months of June and July are not present, although there are of course relatively few observations in the 'High SOI' regime.

In the analysis so far a delay of 1 month in the influence of the switching variable has been assumed. With a large scale effect such as ENSO it is worth looking for a longer delay. Analysis of a delay of 3 months gave very similar results to those presented above, although the SOI threshold was found to be somewhat lower at +1.57. The threshold framework is not ideally suited to searching for delayed threshold effects. However, the extension noted in section 2.3 for incorporating interactions is much more suited to such a search. This issue will be explored further.

In experiments using SOI as a predictor in the model it was found that parameter estimates were less stable than when SOI was used only as a switching variable. Whilst there was some evidence that SOI helps to explain the historical data, it does not seem to provide any additional predictive capability, except as a switching variable.

Year	Month	Standardised Rainfall
1955	August	2.61340
1970	December	-1.90830
1971	**April**	**-0.50198**
1971	**May**	**0.85507**
1973	December	-1.57610
1974	February	-0.57156
1974	April	0.34738
1975	**August**	**0.39667**
1975	**September**	**0.54747**
1975	**October**	**-0.28775**
1976	January	0.22630
1988	October	0.15450
1988	December	-1.26000
1989	May	0.38692

Table 3 Dates and rainfall amounts corresponding to the 'High SOI' regime.

3.2.3 Predicting Rainfall Using SST Gradient As A Switch

The threshold diagnostic plot is shown in Figure 10. There is a clear clustering around –1.5, although there seems to be some uncertainty in the location of this threshold given the smear of points towards and beyond –2.0. This suggests that the threshold is not completely stable. Indeed there were signs of some instability during the course of the subsequent analysis, which requires further investigation.

1-month ahead predictions are shown in Figure 11 below. The correlation between observed and predicted rainfall is 0.789, which is comparable to the other models fitted to Manjimup rainfall. Note that the low-SST gradient regime is of zero order; that is, rainfall in this regime is just a random scatter with no correlation through time. The predicted rainfall therefore does not vary in time and is the estimated intercept in regime 1 (-0.809). The full set of parameter estimates used to produce Figure 11 are shown in Table 4.

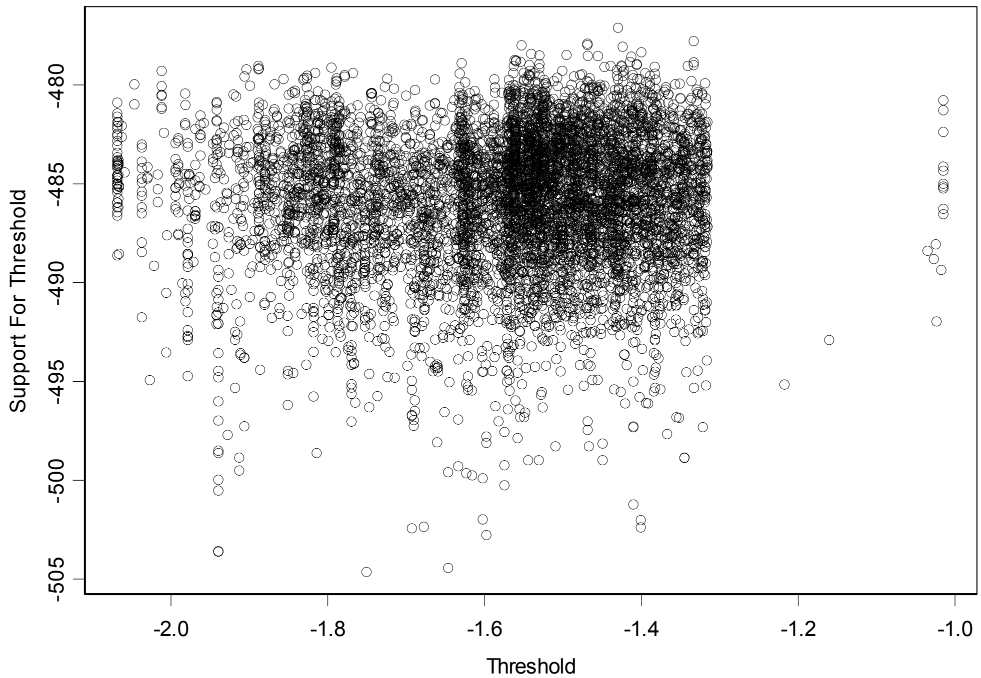


Figure 10 Diagnostic plot for choosing the number of thresholds and their approximate locations for Manjimup rainfall, using SST gradient as the threshold.

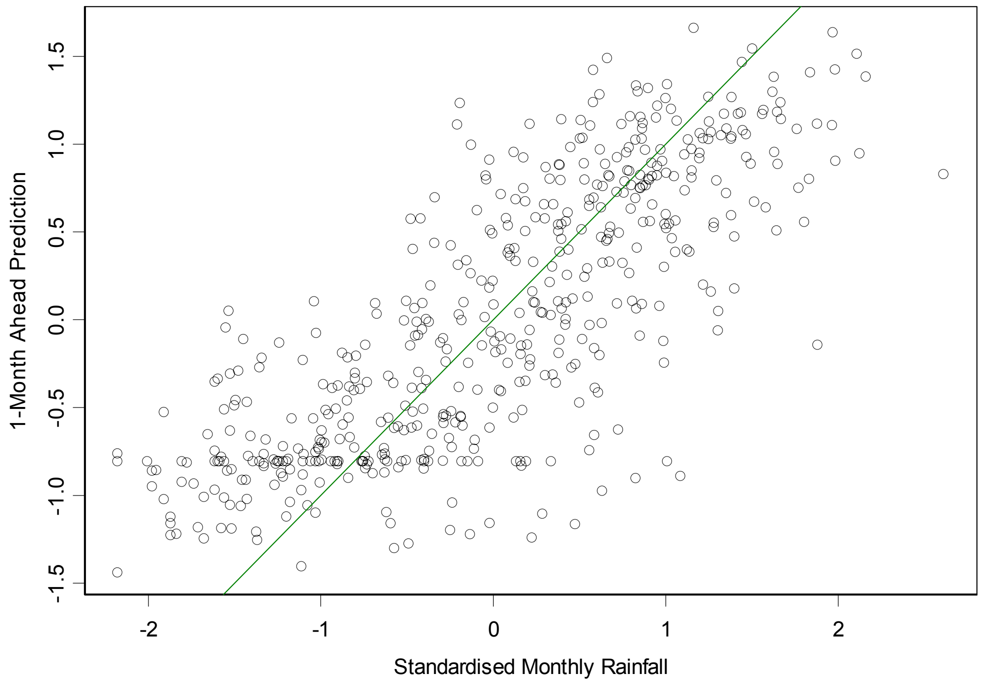


Figure 11 1-month-ahead predictions of monthly rainfall at Manjimup using SST gradient as the switching variable. The correlation between predicted and observed rainfall is 0.789.

	Parameter	Estimate	95% Credibility Interval
	Threshold	-1.59	-2.02, -1.15
Regime 1: Low SST gradient	Order	0	0, 1
	Intercept	-0.809	-1.09, -0.286
	Variance	0.57279	0.35474, 1.0464
Regime 2: High SST gradient	Order	16	13, 18
	Intercept	0.00941	-0.0522, 0.0708
	Lag-1	0.200	0.105, 0.295
	Lag-4	-0.153	-0.241, -0.0578
	Lag-5	-0.0998	-0.192, -0.00922
	Lag-7	-0.103	-0.190, -0.013
	Lag-13	0.226	0.135, 0.313
	Lag-16	-0.0916	-0.193, 0.000
	Variance	0.38890	0.33503, 0.44810

Table 4 Model-averaged parameter estimates with 95% credibility intervals. Only statistically significant parameters are shown, correct to 3sf except for variance parameters which are shown correct to 5sf.

The rainfall distributions in each of the SST gradient regimes are shown in Figure 12. The low SST gradient regime is clearly associated with low rainfall. Considering the parameter estimates in Table 4 once again, there are clear similarities with the previous fitted models in the structure of the normal to high rainfall regime in particular. This would suggest that there exists a reasonable basis for considering SST gradient as a switching mechanism.

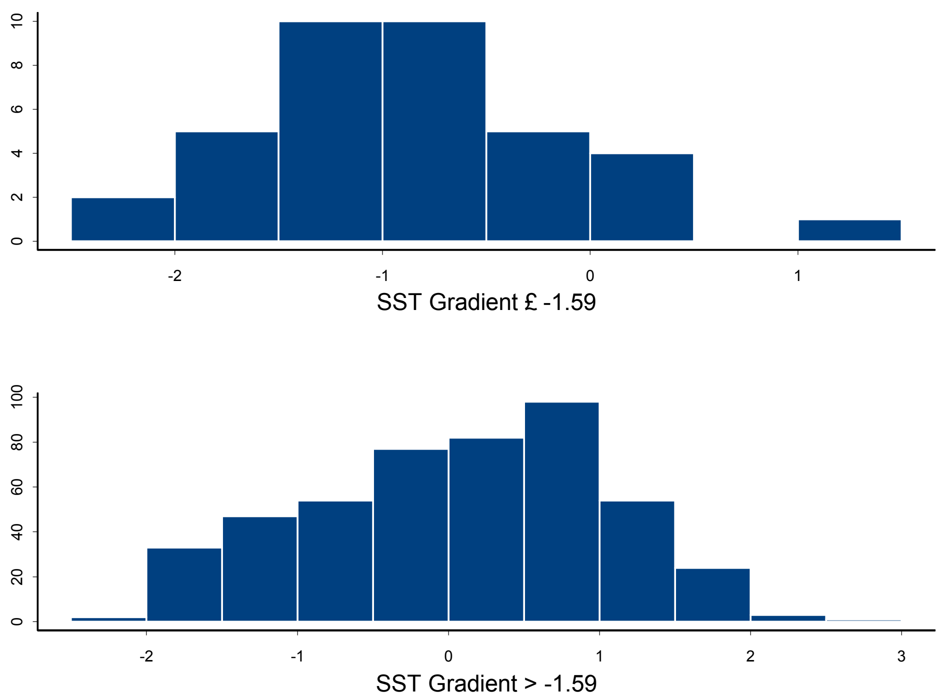


Figure 12 Monthly rainfall distributions at Manjimup for each SST gradient regime. (Note: different plotting scales for each regime)

The distribution of the two regimes is shown in **Table 5**. It is clear from this that the low SST gradient regime is a summer rainfall feature and does not provide an indicator of winter rainfall. A more detailed search of the high rainfall regime may detect a further threshold.

Month	Number of Rainfall Months In Regime:	
	SST Gradient ≤ -1.59	SST Gradient > -1.59
January	2	40
February	16	26
March	18	24
April	1	41
May	0	43
June	0	43
July	0	43
August	0	43
September	0	43
October	0	43
November	0	43
December	0	43
	37	475

Table 5 Number and distribution of months across SST gradient regimes for Manjimup monthly rainfall.

3.3. Rottnest Island Monthly Rainfall

3.3.1 Predicting Rainfall From The Rainfall History Only

The threshold diagnostic plot is shown in Figure 13. Once again there appears to be evidence for a threshold in the vicinity of –1.1, although on this occasion there appears to be a discontinuity present. This suggests that there is evidence for complicated behaviour in this region.

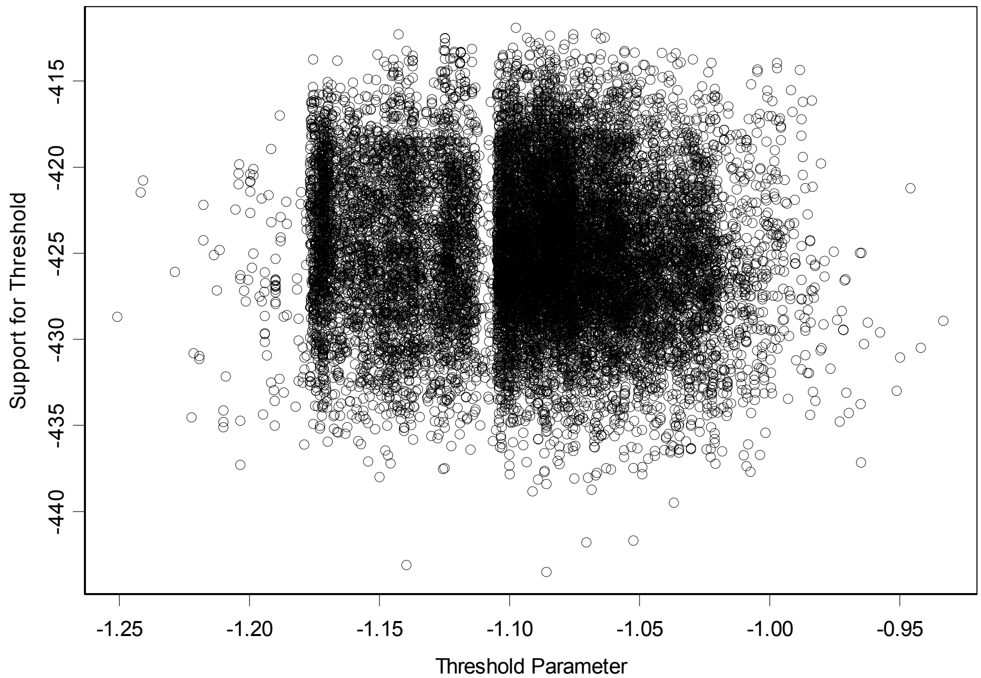


Figure 13 Diagnostic plot for choosing the number of thresholds and their approximate locations for Rottnest Island rainfall.

The 1-month ahead predictions are plotted against their observed values in Figure 14; the correlation of 0.821 is somewhat higher than for Manjimup. The vertically aligned points in the bottom left of the plot correspond to dry months in the observed record. The time series model at present does not explicitly account for dry periods. We see again that the largest observed events are consistently under-predicted.

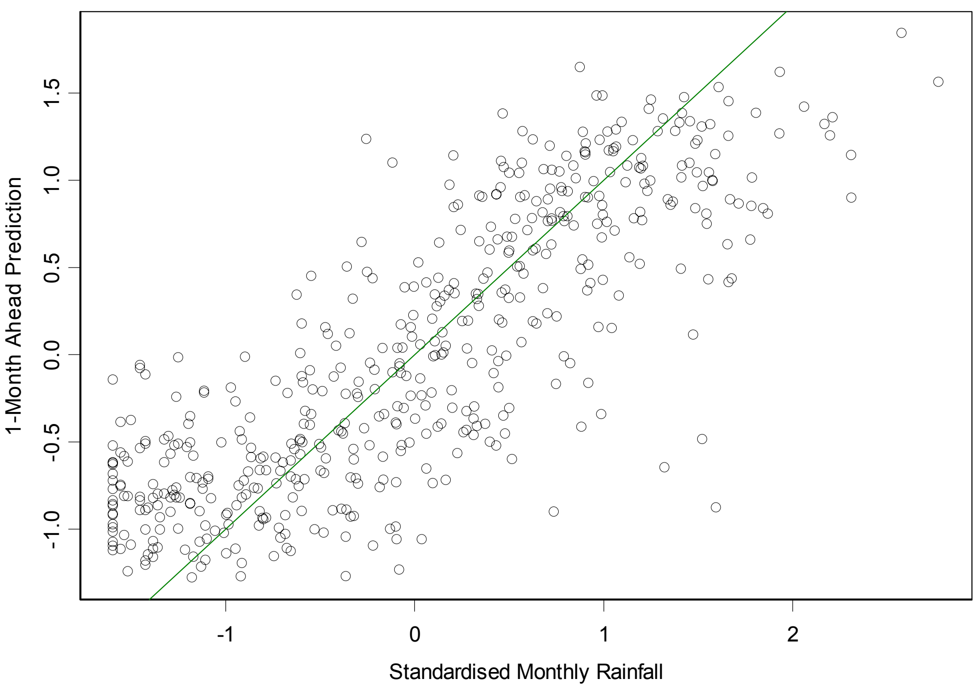


Figure 14 1-month-ahead predictions of monthly rainfall at Rottnest Island. The correlation between predicted and observed rainfall is 0.821.

The parameter estimates used to derive the predictions are shown in Table 6. The 1-month-ahead predictions for the corresponding linear model are shown in Figure 15, which is an order 19 model. The threshold model suggests that there are two rainfall regimes distinguished by a threshold at about –1.1. There is a ‘Low’ regime of order 4, which is essentially a contrast between rainfall 2 and 4 months previously. The ‘Normal/High’ regime has a much longer time dependency of 13 months, with more structure than for the ‘Low’ regime.

	Parameter	Estimate	95% Credibility Interval
	Threshold	-1.09	-1.17, -1.02
Regime 1: 'Low' Rainfall	Order	4	4, 6
	Intercept	-0.427	-1.19, 0.335
	Lag-2	0.321	0.0776, 0.571
	Lag-4	-0.433	-0.629, -0.208
	Variance	0.55373	0.41108, 0.74165
Regime 2: 'Normal/High' Rainfall	Order	13	13, 18
	Intercept	-0.00376	-0.0648, 0.0568
	Lag-1	0.144	0.0465, 0.250
	Lag-7	-0.101	-0.188, -0.0140
	Lag-11	0.262	0.172, 0.355
	Lag-12	0.126	0.0175, 0.230
	Lag-13	0.277	0.189, 0.366
	Variance	0.28853	0.24976, 0.33405

Table 6 Model-averaged parameter estimates with 95% credibility intervals. Only statistically significant parameters are shown, correct to 3sf except for variance parameters which are shown correct to 5sf.

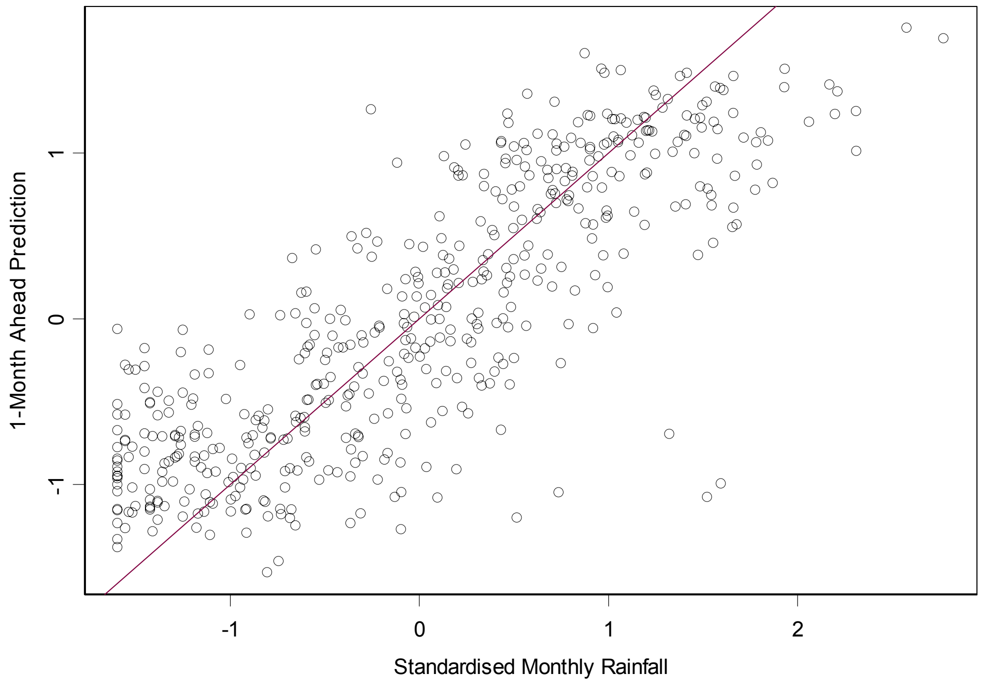


Figure 15 Results from fitting a conventional linear autoregressive model (order 19); the correlation between 1-month-ahead predictions and standardised monthly rainfall is 0.825.

3.3.2 Predicting Rainfall Using The Southern Oscillation Index (SOI) As A Switch

No evidence was found that SOI provides any additional contribution to the rainfall history. When SOI was included as a predictor, using rainfall as the switching variable, the estimates obtained were unstable. When SOI was used as the switching variable, but not as a predictor, no stable threshold was found.

This is quite a stark comparison with Manjimup where SOI did seem to influence monthly rainfall. The case for an SOI influence on Rottnest Island monthly rainfall seems to be weak on the basis of this analysis.

3.3.3 Predicting Rainfall Using SST Gradient As A Switch

We first investigate using SST gradient as the switching variable, but without including SST gradient as a predictor in the model. The threshold diagnostic plot is shown in Figure 16, and there is clear evidence for a threshold at approximately –0.85. The 1-month ahead predictions for rainfall are shown in Figure 17, and we see that the correlation between predicted and observed rainfall is 0.829. The parameter estimates used to generate these predictions are shown in Table 1. There are some strong similarities in the structure of the SST gradient regimes here with the rainfall regimes found in section 3.3.1. They are not as closely matched as the SOI results for Manjimup however.

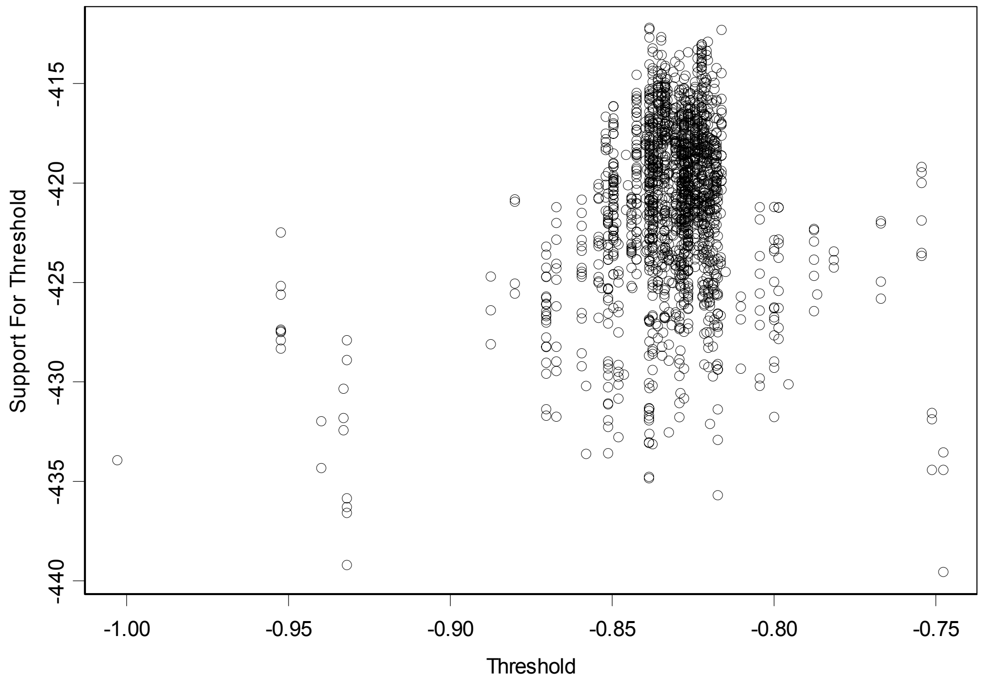


Figure 16 Diagnostic plot for choosing the number of thresholds and their approximate locations for Rottnest Island rainfall, using SST gradient as the threshold.

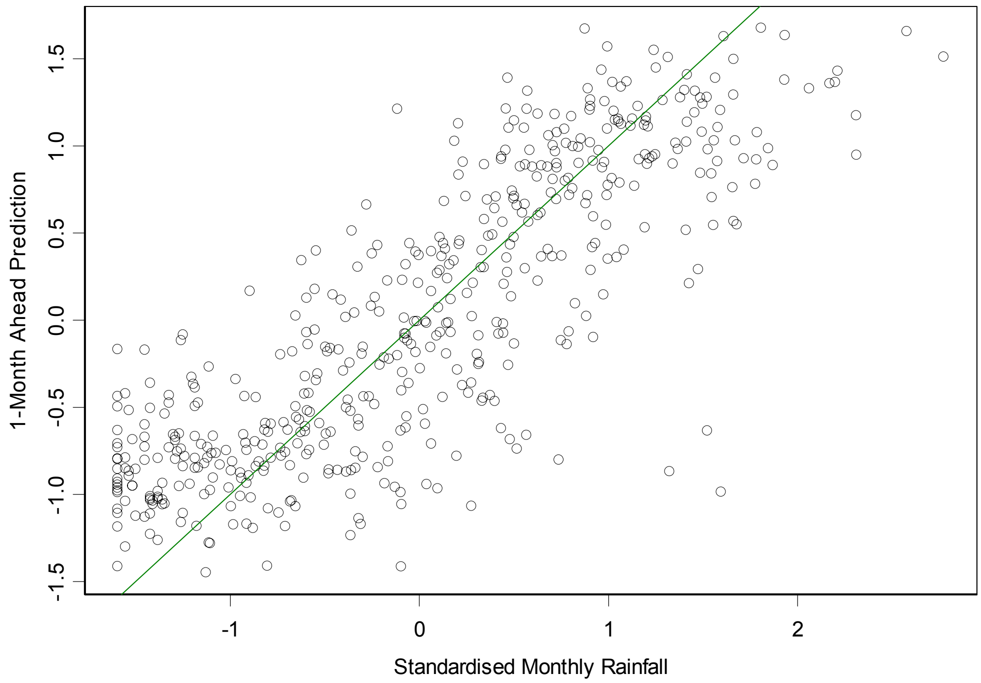
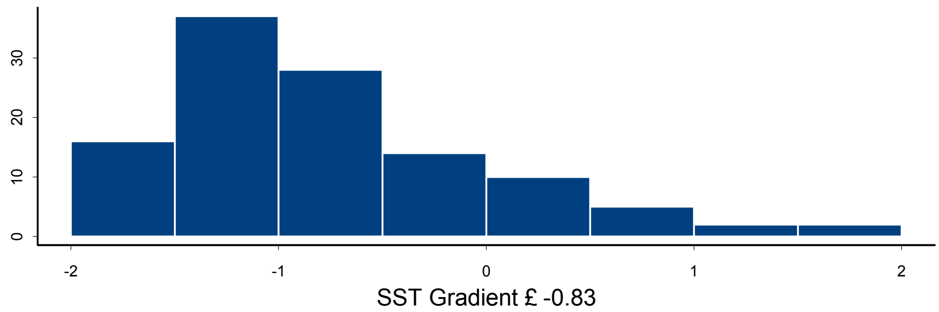


Figure 17 1-month-ahead predictions of monthly rainfall at Rottnest Island using SST gradient as the switching variable. The correlation between predicted and observed rainfall is 0.829.

	Parameter	Estimate	95% Credibility Interval
	Threshold	-0.831	-0.851, -0.805
Regime 1: Low SST gradient	Order	5	5, 5
	Intercept	-0.510	-0.804, -0.248
	Lag-2	0.295	0.0878, 0.478
	Lag-4	-0.378	-0.664, -0.141
	Lag-5	-0.313	-0.572, -0.0569
	Variance	0.51144	0.38359, 0.67864
Regime 2: High SST gradient	Order	13	13, 13
	Intercept	0.0555	-0.00467, 0.113
	Lag-1	0.215	0.119, 0.307
	Lag-4	-0.0874	-0.167, -0.00691
	Lag-7	-0.0950	-0.181, -0.0112
	Lag-11	0.265	0.172, 0.359
	Lag-13	0.242	0.150, 0.329
	Variance	0.27942	0.23820, 0.32471

Table 7 Model-averaged parameter estimates with 95% credibility intervals. Only statistically significant parameters are shown, correct to 3sf except for variance parameters which are shown correct to 5sf.

It is of some interest to compare the rainfall patterns defined by the two SST gradient regimes identified, and these are shown in Figure 18. We see that the low SST gradient regime is associated predominantly with below average rainfall. The rainfall months associated with the regimes are shown in Table 8, and it seems clear that the high SST gradient is associated with winter rainfall in particular. There is therefore some potential here to develop a predictive model for monthly rainfall.



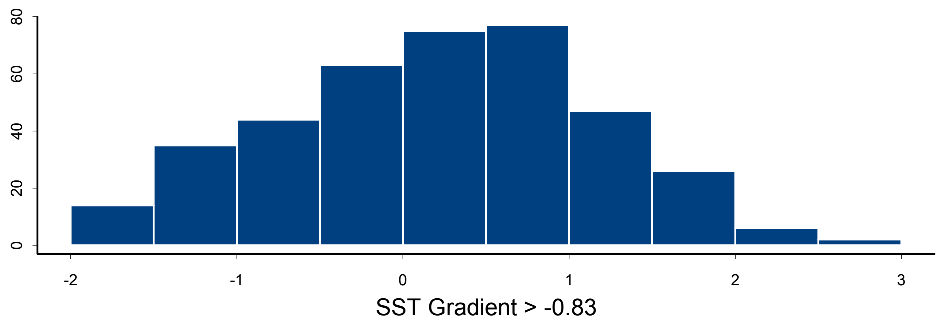


Figure 18 Monthly rainfall distributions at Rottnest Island for each SST gradient regime. (Note: Different plotting scales for each regime)

Month	Number of Rainfall Months In Regime: SST Gradient ≤ -0.83	SST Gradient > -0.83
January	21	20
February	37	5
March	38	4
April	14	28
May	1	41
June	1	41
July	1	41
August	1	41
September	0	42
October	0	42
November	0	42
December	0	42
	114	389

Table 8 Number and distribution of months across SST gradient regimes for Rottnest Island monthly rainfall.

4 Discussion

4.1. Physical Interpretations

The case studies presented in this report were primarily used to develop and test the nonlinear statistical methodology. For the purposes of discussion we set this aside for the moment. We note that the new method has produced sensible results, which is the outcome we were seeking at this stage of its development.

Some interesting differences have emerged between the two study sites of Rottnest Island and Manjimup. In the case of monthly rainfall at Rottnest Island we have found no significant link to the Southern Oscillation. There does however appear to be a link in the case of Manjimup. This is supported by a comparison of the results where SOI and rainfall are used as the switching variable. The regimes found using rainfall as the switching variable were almost identical to those found when using SOI as the switching variable. This suggests that SOI can be used to explain the switching behaviour of the monthly rainfall series at Manjimup.

Whilst a rainfall teleconnection with the Southern Oscillation is not widely recognised, there is some existing evidence. For example, *Crowder* [1995] pp240-41 notes severe rainfall deficits in April 1982 through February 1983 in the far Southwest. This was a very severe El Niño event. This period was not detected in our Manjimup case study, but this may be because the high-SOI regime was not particularly well defined.

It was found that SOI did not work well as a predictor in the statistical modelling, only as a switching variable. Thus SOI can in principle be used to forecast switches in rainfall regime, but not as a simple predictor of future rainfall values.

There is some evidence that SST gradient in the Indian Ocean influences rainfall at both Rottnest Island and Manjimup. This is physically reasonable since SST gradient is measuring a capacity for winds to be generated in the mid-Indian Ocean. It is not difficult to see how a change in circulation patterns beyond a critical value could cause a switch in rainfall regime.

The nature of circulation patterns is that they act on very large scales. It may therefore be more realistic to develop a switching variable that is a *combination* of broad scale circulation patterns. It would be of particular relevance to incorporate data from the Southern (if feasible) and Pacific Oceans if this were to be done.

Key Points:

- Different physical mechanisms appear to influence rainfall at Rottnest Island and Manjimup.
- There is some evidence that the Southern Oscillation plays a role in causing switches between rainfall regimes.
- Sea surface temperature gradient in the mid-Indian Ocean seems to influence switching of rainfall regime at Rottnest Island. It seems to be a leading indicator of winter rainfall, and there seems to be the basis for a winter rainfall prediction scheme. It also seems to play a role at Manjimup, but is less well defined than for Rottnest Island.
- There is a case for developing switching variables that are combinations of variables representing circulation patterns in the Indian, Southern (if feasible) and Pacific Oceans. The extension to the Bayesian threshold method discussed below will be of some use.

4.2. Statistical Issues

The statistical methodology has been found to work well on a practical level. It is reasonably straightforward to simultaneously identify important lags and estimate the corresponding parameters. The identification of the number of thresholds and the delay is somewhat ad hoc however. In standard practice penalised likelihood methods would be used, although not particularly satisfactorily. However, a more promising approach is to use spline methods to estimate the generating mechanism of the time series. In this richer setting the thresholds become knot points, and the choice of knot points is a somewhat easier problem to solve. The delay is expressed through the lags associated with the knot points, and is essentially automatic. The methodology developed so far seems to be reasonably straightforward to adapt to this more general approach.

During the final phase of IOCI we will produce probability distributions for forecasts, which will incorporate a complete statement of uncertainty. These probability distributions can also be used to validate the nonlinear modelling.

An issue that has yet to be resolved in an entirely satisfactory way is the choice of prior distributions for the model orders in each regime. The results reported here use Poisson distributions, but some of our results suggest that better mixing could be obtained by placing a hyper-prior on the Poisson mean- a gamma distribution would result in a negative binomial prior overall. We will investigate this point in the on-going work.

Key Points:

- The statistical methodology we have developed is working well.
- Some physically interesting results have already been obtained.
- Probability forecasts will be produced in the next phase.
- The choice of prior distributions for model orders has not been fully resolved. An approach using hyper-priors on model order will be considered.
- A collaborative effort with the new nonlinear tool is required to extract maximum value from it.

5 Conclusions

The primary task for CSIRO Mathematical & Information Sciences (CMIS) during this phase of IOCI has been to develop the statistical methodology to the point where it can usefully be applied. We have reached this point, although there are still some potentially useful extensions that can be pursued, particularly in relation to modelling climate interactions. However, the focus of CMIS' work for the remainder of IOCI will be on the detailed development and analysis of case studies identified by the contributing partners.

The case studies described here have been presented in the spirit of testing whether sensible results are obtained using the methodology developed, rather than seriously seeking rainfall predictors. The task of seeking rainfall predictors is a collaborative exercise, which now has an additional nonlinear tool to make use of. In our case studies we have found some evidence that predictors such as the Southern Oscillation (SOI) and sea surface temperature (SST) gradient have some potential to provide a climate switching mechanism in the threshold model framework described by Figure 2. In the case of SOI we found a stark difference in that there appeared to be an influence on Manjimup rainfall but not on Rottnest Island rainfall. For Manjimup it seems that the Southern Oscillation has some impact in extreme cases, but by no means supplies a complete picture.

There is some evidence that SST gradient in the mid-Indian Ocean causes switching of rainfall regime at Rottnest Island in particular, and clearly seems to be linked to winter rainfall. There is a basis here for exploring predictive models for winter rainfall. There does also seem to be some influence on Manjimup rainfall. It could well be that an underlying climate switch should be formed from a *combination* of processes, rather than Indian and Pacific Ocean influences on their own. The methodology developed by CMIS could be adapted to aid a search for such combinations.

Key Points:

- We have developed a physically motivated statistical model ('Bayesian switching model') for modelling nonlinear climate processes.
    - Changes between climate regimes are triggered by a switching variable, and alternative switching variables can be compared.
- The Bayesian switching model can identify good predictors and the lags at which they influence climate variables, such as rainfall.
- We have reached the point where a nonlinear time series approach can be applied to practical problems.
- There is some evidence that SOI and mid-Indian Ocean SST gradients play a role in switching between rainfall regimes. This is cited at this stage as evidence that the new nonlinear approach is producing sensible results, rather than new insights *per se*.
- Interactions between climate processes are likely to influence rainfall in Southwest WA. Some reasonably straightforward extensions to the Bayesian switching model will facilitate the search for subtler climate teleconnections arising from such interactions.
- The focus of future work will be the development of case studies with IOCI's contributing partners.

References

Crowder, B., *The Wonders of the Weather,* Australian Government Publishing Service, Canberra, 1995.

Graham, N. E. and Barnett, T. P., Sea surface temperature, surface wind divergence, and convection over tropical oceans, *Science,* **238,** 657-659, 1987.

Hsieh, W. W., Tang, B. and Garnett, E. R., Teleconnections between Pacific sea surface temperatures and Canadian prairie wheat yield, *Agric. For. Meteor.,* **96,** 209-217, 1999.

Palmer, T. N., A nonlinear dynamical perspective on climate prediction, *J. Clim.,* **12,** 575-591, 1999.

Appendix A- Glossary

Cross-referenced terms and acronyms are shown in italics.

Anomaly	It is usual to express climate data as deviations from the long term average, and this deviation is known as an anomaly.
Bayesian	A statistical framework that expresses uncertainty using probability distributions. Bayesian statisticians explicitly combine data with subjective knowledge to learn about physical processes. This is accomplished using *Bayes' theorem*.
Bayes' Theorem	As implemented in scientific practice, this theorem essentially states that uncertainty conditional on available data and expert knowledge is proportional to the product of the uncertainty in the data and the uncertainty in expert knowledge.
Delay	In physical systems there may well be a time delay between cause and effect, and this is captured by a so-called delay parameter.
Interaction	In physical systems the effect of one variable may depend on the value of another. For example, a low pressure system will not bring rainfall when sea surface temperature is low. In this case sea surface temperature is said to interact with air pressure.

Knot Point	When using *splines* we divide up the domain of a function so that it can be approximated by a set of simple functions. The points at which the domain is divided are known as knot points.
Linear	A general term to describe relationships that can be represented as straight lines between two variables, or hyperplanes for many variables.
Markov chain Monte Carlo	A computationally intensive technique that uses simulation techniques to implement *Bayesian* statistical methods. This term is universally known by the acronym *MCMC*.
Nonlinear	A general term to describe relationships that cannot be described as straight lines or hyperplanes, as is the case for *linear* relationships.
Posterior Distribution	A probability distribution that integrates expert knowledge and available data, and is typically calculated using *Bayes' theorem.*
Reversible Jump MCMC	A methodology for choosing optimal statistical models in a *Bayesian* statistical framework, motivated by *MCMC* ideas.
Spline	A technique for approximating functions, typically accomplished by breaking the domain of the function into segments within each of which some simple function is

fitted to the data. The boundaries between domains are known as *knot points*.

Switching Variable — In *threshold models* of a physical system a key variable causes the system to switch behaviour. This key variable is known as a switching variable.

Time Series — A set of data recorded sequentially in time.

Appendix B- List of Acronyms

CLW	CSIRO Land and Water.
CMIS	CSIRO Mathematical and Information Sciences.
CMR	CSIRO Marine Research.
ENSO	El Niño-Southern Oscillation
IOCI	Indian Ocean Climate Initiative.
MCMC	Markov chain Monte Carlo.
SOI	Southern Oscillation Index
SST	Sea Surface Temperature.
SWA	Southwest Western Australia.

Appendix C- Submitted manuscript describing the statistical methods developed by IOCI for the case studies.

At the time of writing the manuscript is under peer review for possible publication in the *Journal of Time Series Analysis*. A copy of the current version is available from the author:

e-mail: eddy.campbell@csiro.au

Tel: (08) 9333 6203