A technique for extracting potentially predictable patterns from climate data

C. S. Frederiksen A. P. Kariko* X. Zheng†

(Received 1 June 2001; revised 1 October 2002)

Abstract

We propose a computational technique which makes it possible to extract long-range potentially predictable patterns of interannual variability of meteorological seasonal mean fields. These patterns arise from slowly varying external forcing, such as sea surface temperatures, and slowly varying internal dynamics. The method provides a means of decomposing the covariance matrix of a seasonal mean field into covariance matrices for the potentially predictable and the chaotic, or weather-noise, components, separately. We illustrate the technique using Australian surface maximum temperatures during December-January-February (DJF) for the period 1958–1991. The dominant patterns, arising from the potentially predictable covariance matrix, are shown to

*Bureau of Meteorology Research Centre,
mailto:c.frederiksen@bom.gov.au, mailto:alex.kariko@bom.gov.au
†National Institute of Water and Atmospheric Research
[0]See http://anziamj.austms.org.au/V44/CTAC2001/Fred for this article,
 Published 1 April 2003. ISSN 1446-8735

be more closely related to slowly varying external forcing and slowly varying internal dynamics than those from a conventional analysis. The importance of tropical sea surface temperatures in forcing the potentially predictable patterns is also discussed.

Contents

1 Introduction

Over the last two and a half decades much effort has gone into determining the potential for long-range (in advance of three months) predictability of regional and global climate. Although the atmosphere is inherently chaotic, there is the potential for predicting the seasonal averaged atmospheric circulation beyond the deterministic predictability of synoptic scale weather (less than 15 days) because the low frequency state of the atmosphere is coupled to the slowly varying potentially predictable ocean circulation. Madden [5] was the first to introduce the idea of decomposing, conceptually, a

seasonal mean of a climate variable into a potentially predictable component and an unpredictable, or chaotic, component. The potentially predictable component is associated with slowly varying external forcings (such as, for example, sea surface temperatures (SSTs), sea ice, vegetation coverage and radiative forcing) and slowly varying internal dynamics (such, as for example, the quasi-biennial oscillation). This component was referred to as "potentially" predictable because the external forcing and the slowly varying (inter-annual/supra-annual) internal dynamics are themselves potentially predictable. Thus, for example, forecast schemes presently exist for predicting, with good skill, variations in tropical Pacific SSTs. A climate variable which depended on these SSTs would therefore also be potentially predictable. The chaotic component, commonly referred to as weather noise, is associated with day-to-day weather events, which are unpredictable beyond a couple of weeks. There are now a number of techniques available for estimating the relative importance of the potentially predictable and weather noise components of climate variability at any geographical location on interannual, or greater, time-scales (see, for example, [7] and references therein).

However, an even more important problem, in the analysis and prediction of climate, is to identify spatial patterns of inter-annual variability (covariability) associated with the potentially predictable component. Until now, empirical orthogonal function (EOF) analysis, using the variance/covariance matrix of the seasonal mean climate variable, has been a common tool used to study the dominant spatial patterns of variability in climate data. However, there is no guarantee that the dominant patterns derived by such techniques will be closely related to the slowly varying external forcings or internal dynamics. This is especially true in the extra-tropics where the weather noise component of atmospheric variability may be quite large. Ideally, the EOF analysis should be applied to the variance/covariance matrix of the potentially predictable component. Until

now this has not been possible.

Here, we propose a method for separating the covariance matrix of a climate variable into covariance matrices associated with the potentially predictable and weather noise components, thus allowing patterns of each component to be derived separately. The technique is applied to a study of inter-annual variability of Australian surface air maximum temperatures using high quality data from 78 stations of the Bureau of Meteorology, for the period 1958–1991. We also show how the potentially predictable patterns can be used to identify significant predictors of temperature variability associated with specific external forcings, such as, for example, tropical SSTs, which are perhaps one of the most important external forcing of climate at the inter-annual time-scale.

The plan of this paper is as follows. In Section 2, we discuss the methodology, the observed temperature data and SST data are described in Section 3, and the results are presented in Section 4. Our conclusions are summarised in Section 5.

2 Methodology

The method described herein relies on the availability of a daily time series of the meteorological variable (say x) over a number of decades. It is also assumed that the mean annual cycle has been removed from the daily time series. Conceptually, we consider a daily anomaly in a particular year, day and geographical location, as consisting of a potentially predictable component, which is related to the interannual and supra-annual variability of external forcings and internal dynamics and may therefore be regarded as a constant over the season, and an unpredictable chaotic component arising from day-to-day weather variability. Thus, a daily anomaly $x_{yt}(r)$ is

represented by a simple linear model as

$$x_{yt}(r) = \mu_y(r) + \varepsilon_{yt}(r)\,. \tag{1}$$

Here, $y = 1, \ldots, Y$ is the year; $t = 1, \ldots, T$ is the day in a season of length T days; $r = 1, \ldots, R$ denotes a location in a field with R locations; $\mu_y(r)$ represents the potentially predictable component; $\varepsilon_{yt}(r)$ represents daily weather noise, modelled here as the residual daily departure of $x_{yt}(r)$ from the seasonal value $\mu_y(r)$. The set $\{\varepsilon_{yt} \mid t = 1, \ldots, T\}$ is assumed to represent a stationary normal stochastic process in time with mean zero and to be statistically independent and identically distributed with respect to year y. Note, in this model SST-forced intra-seasonal variability is included in the weather noise component. It is further assumed that the sets $\{\mu_y\}$ and $\{\varepsilon_{yt}\}$ are statistically independent.

As in [7], it is convenient to introduce the convention of using a circle as a subscript whenever an average is done over that index. Thus, for example, from Eqn. (1), a seasonal mean anomaly derived from daily values is written as

$$x_{y\circ}(r) = \mu_y(r) + \varepsilon_{y\circ}(r)\,. \tag{2}$$

Also, the symbol $\hat{V}$ will be used to denote the estimated variance of a single variable or the covariance of two variables. The covariance of the seasonal mean anomaly can be estimated by the sample covariance,

$$\hat{V}(x_{y\circ}(r_1), x_{y\circ}(r_2)) = \frac{1}{Y-1} \sum_{y=1}^{Y} \left[x_{y\circ}(r_1) - x_{\circ\circ}(r_1)\right] \left[x_{y\circ}(r_2) - x_{\circ\circ}(r_2)\right]\,. \tag{3}$$

With daily data, an estimate of the covariance of the weather noise component can be derived from the sample covariance, and

using the definition of the seasonal mean, as,

$$\begin{aligned}
&\hat{V}\left(\varepsilon_{y\circ}(r_1), \varepsilon_{y\circ}(r_2)\right) \\
&\simeq \frac{1}{Y}\sum_{y=1}^{Y} \epsilon_{y\circ}(r_1)\epsilon_{y\circ}(r_2) \\
&= \frac{1}{YT^2}\sum_{y=1}^{Y}\left(\left[\sum_{t=1}^{T}\epsilon_{yt}(r_1)\right]\left[\sum_{t=1}^{T}\epsilon_{yt}(r_2)\right]\right) \\
&= \frac{1}{YT^2}\sum_{y=1}^{Y}\Re\left(\left[\sum_{t=1}^{T}\epsilon_{yt}(r_1)e^{it0}\right]\left[\sum_{t=1}^{T}\epsilon_{yt}(r_2)e^{-it0}\right]\right). \qquad (4)
\end{aligned}$$

This is just the cross periodogram between $\epsilon_{yt}(r_1)$ and $\epsilon_{yt}(r_2)$ at frequency zero. If we assume that the cross spectrum at zero frequency is smooth and T is sufficiently large, then the weather noise covariance can be approximated at the nearest frequency $2\pi/T$ [1]. Thus,

$$\begin{aligned}
&\hat{V}(\varepsilon_{y\circ}(r_1), \varepsilon_{y\circ}(r_2)) \\
&\simeq \frac{1}{YT^2}\sum_{y=1}^{Y}\Re\left(\left[\sum_{t=1}^{T}\epsilon_{yt}(r_1)e^{it2\pi/T}\right]\left[\sum_{t=1}^{T}\epsilon_{yt}(r_2)e^{-it2\pi/T}\right]\right). \qquad (5)
\end{aligned}$$

Of course, $\epsilon_{yt}(r)$ is not observable and so Eqn. (5) can not be used directly to estimate $\hat{V}(\varepsilon_{y\circ}(r_1), \varepsilon_{y\circ}(r_2))$. However, because the Fourier transform of any constant with respect to t at frequency $2\pi/T$ is identically zero, it follows, using Eqn. (1), that,

$$\begin{aligned}
&\hat{V}(\varepsilon_{y\circ}(r_1), \varepsilon_{y\circ}(r_2)) \\
&\simeq \frac{1}{YT^2}\sum_{y=1}^{Y}\Re\left(\left[\sum_{t=1}^{T}x_{yt}(r_1)e^{it2\pi/T}\right]\left[\sum_{t=1}^{T}x_{yt}(r_2)e^{-it2\pi/T}\right]\right), \qquad (6)
\end{aligned}$$

since $\mu_y(r_1)$ is a constant with respect to t. Because the sets $\{\mu_y\}$

and $\{\varepsilon_{yt}\}$ are assumed to be statistically independent, we can estimate $V(\mu_y(r_1), \mu_y(r_2))$ by

$$\hat{V}(\mu_y(r_1), \mu_y(r_2)) = \hat{V}(x_{y\circ}(r_1), x_{y\circ}(r_2) - \hat{V}(\varepsilon_{y\circ}(r_1), \varepsilon_{y\circ}(r_2)) \,. \quad (7)$$

An empirical orthogonal function (EOF) analysis [2], or eigenanalysis, can then be applied to each of the spatial covariance matrices given by Eqns. (3), (6) and (7) to determine dominant patterns of interannual variability (the eigenvectors of the covariance matrix), which we shall refer to as the *total*-EOF*s*, the *weather*-EOF*s* and the *predictable*-EOF*s*, respectively. The corresponding eigenvalues then give the amount of variance, in each component, explained by each pattern. The corresponding principal component (PC), or daily time series, $\{p_{yt} \,|, y = 1, \ldots, Y, t = 1, \ldots, T\}$, associated with an EOF $\{e(r) \mid r = 1, \ldots, R\}$ is then defined as,

$$p_{yt} = \sum_{r=1}^{R} e(r) x_{yt}(r) = \sum_{r=1}^{R} e(r)\mu_y(r) + \sum_{r=1}^{R} e(r)\varepsilon_{yt}(r) \equiv \widetilde{\mu}_y + \widetilde{\varepsilon}_{yt} \,, \quad (8)$$

where $\widetilde{\mu}_y$ and $\widetilde{\varepsilon}_{yt}$ represent the potentially predictable and weather noise components, respectively, of the PC time series. The potential predictability of the PC is then defined as $\hat{V}(\widetilde{\mu}_y, \widetilde{\mu}_y)/\hat{V}(p_{y\circ}, p_{y\circ})$, that is, as the fraction of the interannual variance in the PC series which is due to the potentially predictable component. This ratio can be determined from Eqns. (3), (6) and (7), with x_{yt} replaced by p_{yt} and $r_1 = r_2 = r$. Here, $p_{y\circ}$ represents the interannual PC time series of the corresponding EOF pattern. For convenience, we shall use the notation *total*-PC, *weather*-PC and *predictable*-PC to differentiate between the three types of possible PCs.

Rotation [2] is often applied to increase the localization of the EOF patterns and is often thought to lead to more physically realistic patterns. In the example below we have chosen to use varimax

rotation to achieve this, with only significant EOFs (determined from a scree diagram) used in the rotation. In the case of the potentially predictable EOFs, we will only rotate those EOFs that, in addition, have significant potential predictability at a 5% level, which for the length of our time series, means a potential predictability $> 35\%$.

3 Data

To illustrate the usefulness of the technique, we have applied the methodology to a dataset containing daily maximum surface temperatures for DJF from 78 high quality Australian stations for the period 1958–1991. This dataset is a subset of data, gathered and quality tested by the Australian Bureau of Meteorology, and described in [4] and [6]. The location and distribution of these stations is shown in Figure 1.

The most important external forcing of climate variability on the inter-annual time-scale is SST forcing, and particularly tropical SST forcing. An understanding of this type of climate forcing forms the basis of many seasonal forecast schemes. Typically, such schemes are based on relationships between the total-PCs of a climate variable and those of the SST. However, as we show below, the total-EOFs may, in many cases, reflect both the predictable and unpredictable weather components in combination. As a consequence, the total-PCs may have much lower potential predictability than the corresponding predictable-PCs. Because the weather component is inherently unpredictable, one should aim to forecast the predictable component. Thus, an understanding of the relationship between this component and SSTs is essential.

To try to quantify this relationship, we have applied a multilinear regression analysis between seasonal means of the predictable-

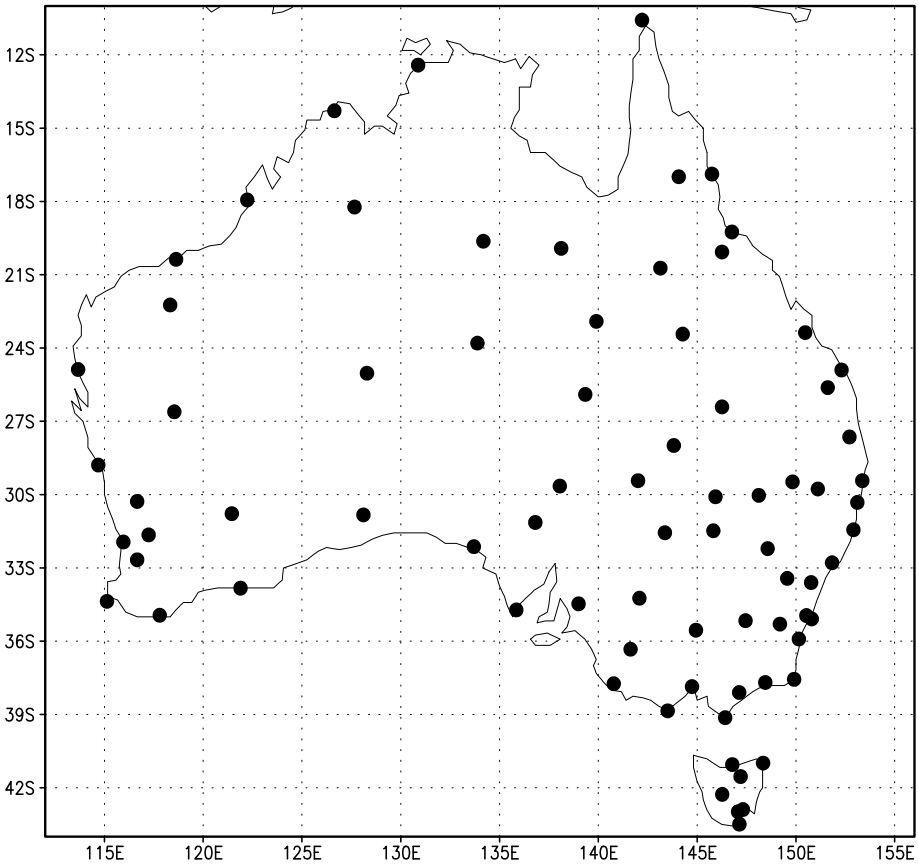

FIGURE 1: Location and distribution of Australian high quality surface air temperature stations.

PCs and PCs of the seasonal means of tropical SSTs. The SSTs used are taken from the Hadley Centre (UK Meteorological Office) global ice and sea surface temperature dataset (HADISST1.1) sub-sampled onto a $4° \times 4°$ latitude/longitude grid. We have used a standard principal component analysis [2] to generate dominant EOFs of these SSTs in three regions, covering the tropical Pacific Ocean (pac), the Atlantic Ocean (atl) and the Indian Ocean/Indonesian archipelago (ind).

Because we are interested in both simultaneous (0-lag) and one season lagged (1-lag) relationships with SSTs, the analysis has been conducted on both DJF and September-October-November (SON) SSTs. For each season three dominant SST EOFs have been derived for the Pacific and Indian regions, and two dominant EOFs for the Atlantic region. The analysis is little changed by including more SST EOFs. The corresponding SST PCs will play the role, in the regression analysis, of predictors for the predictable-PCs. The SST EOFs are shown in Figures 2 and 3.

4 Results

The first four dominant rotated EOFs (REOFs) of the total field, the predictable and weather components of Australian maximum surface temperatures are shown in Figures 4, 5 and 6, respectively, with the potential predictability shown in brackets. The contours in this diagram represent surface temperature anomalies. The total-REOFs explain 27%, 20%, 16% and 7%, respectively, of the variance in the total field; the predictable-REOFs explain 24%, 19%, 10% and 6%, respectively, in the predictable component; the weather-REOFs explain 23%, 19%, 13% and 8%, respectively, in the weather noise component. The first 9 unrotated EOFs, explaining 89% and 79% of

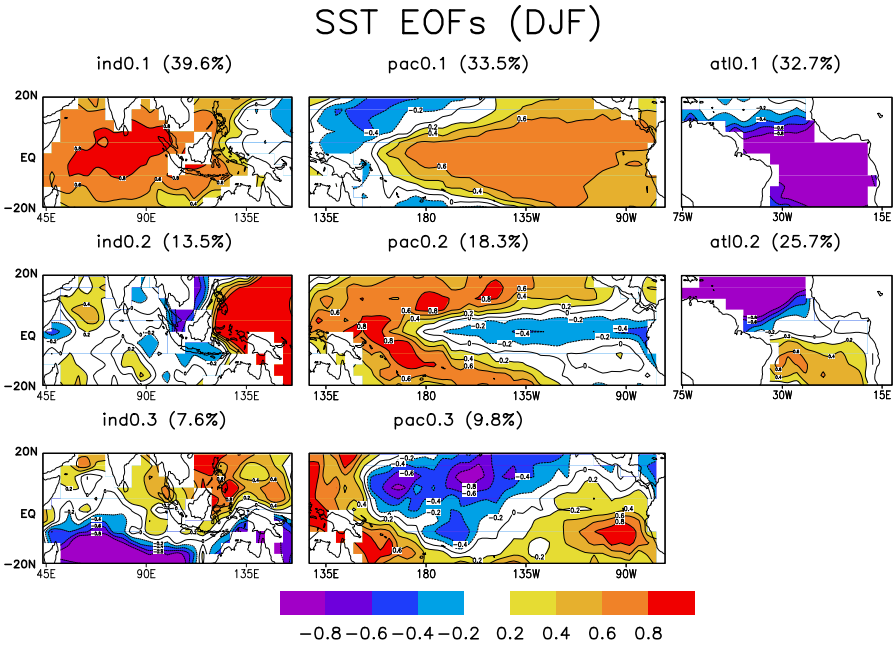

FIGURE 2: Dominant EOFs of tropical SSTs in the Indian (ind), Pacific (pac) and Atlantic (atl) Oceans during DJF and SON. The first index specifies the lag length (0-lag or 1-lag) and the second index the EOF number (ordered by decreasing explained variance). Contour interval 0.2.

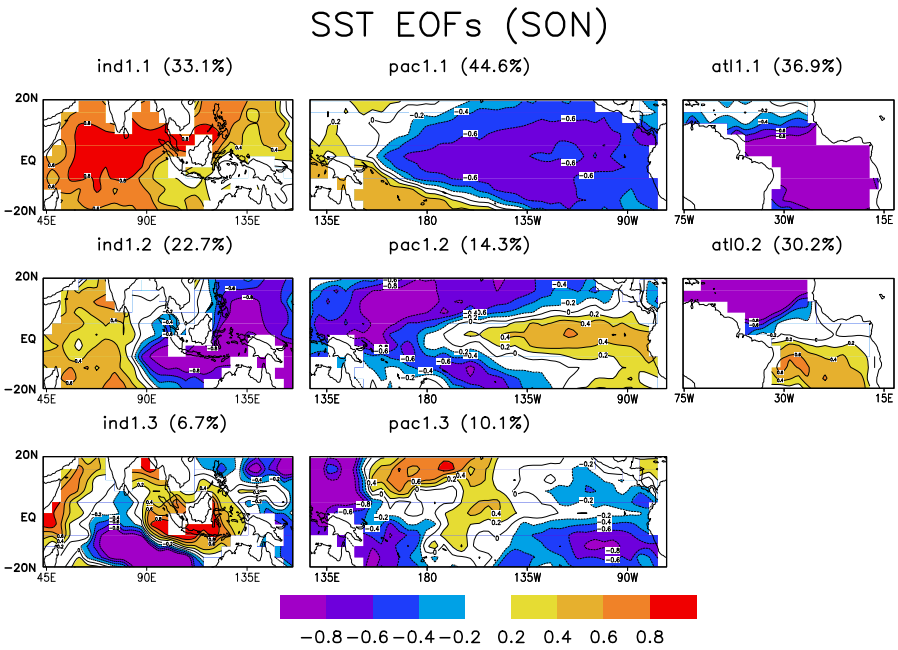

FIGURE 3: As in Figure 2, but for SON.

the total and weather variance, respectively, were used to derive the total-REOFs and weather-REOFs. Unrotated predictable-EOFs 1, 2, 3, 5, 6, 8, 9, 11 and 12, which explain 79% of the variance in the predictable component, all have significant potential predictability and were used to derive the predictable-REOFs.

Comparing the total-REOFs and the predictable-REOFs, there are similarities in the spatial patterns with total-REOFs 1, 2, 3, and 4 having pattern (or anomaly) correlations of -0.91, 0.98, -0.73 and 0.88 with predictable-REOFs 2, 1, 3 and 4, respectively. However, in each case the corresponding predictable-REOF has greater potential predictability. This is especially the case with total-EOF 3 and 4. In fact, total-EOF 4 has only a potential predictability of 29%, which is not statistically significant at the 5% level.

The total-REOFs appear in most cases to be a mixture of the predictable and weather REOFs. Thus, for example, total-REOF 1, as well as having a high pattern correlation with predictable-REOF 2, has pattern correlation of 0.94 with weather-REOF 2. Similarly, total-REOF 2 and 3 have pattern correlations of 0.71 and 0.65 with weather-REOFs 3 and 1, respectively. Total-REOF 1 is a particularly well-known mode of variability and has been associated with the El Nino-Southern Oscillation (ENSO) phenomenon [3, e.g.] which is associated with pacific SST variability similar to that depicted in SST EOF 1 in both DJF and SON (see Figures 2 and 3). However, our analysis here suggests that it also represents substantial weather noise and hence a reduced potential predictability compared with the corresponding predictable-REOF 2. There are also subtle differences in the structure of the two patterns, with the predictable pattern having maximum weighting centred nearer the east coast and a secondary maximum over western Australia. In contrast, total-REOF 1 has only one maximum centred over eastern Australia but more westward, much like the weather pattern (weather-REOF 2).

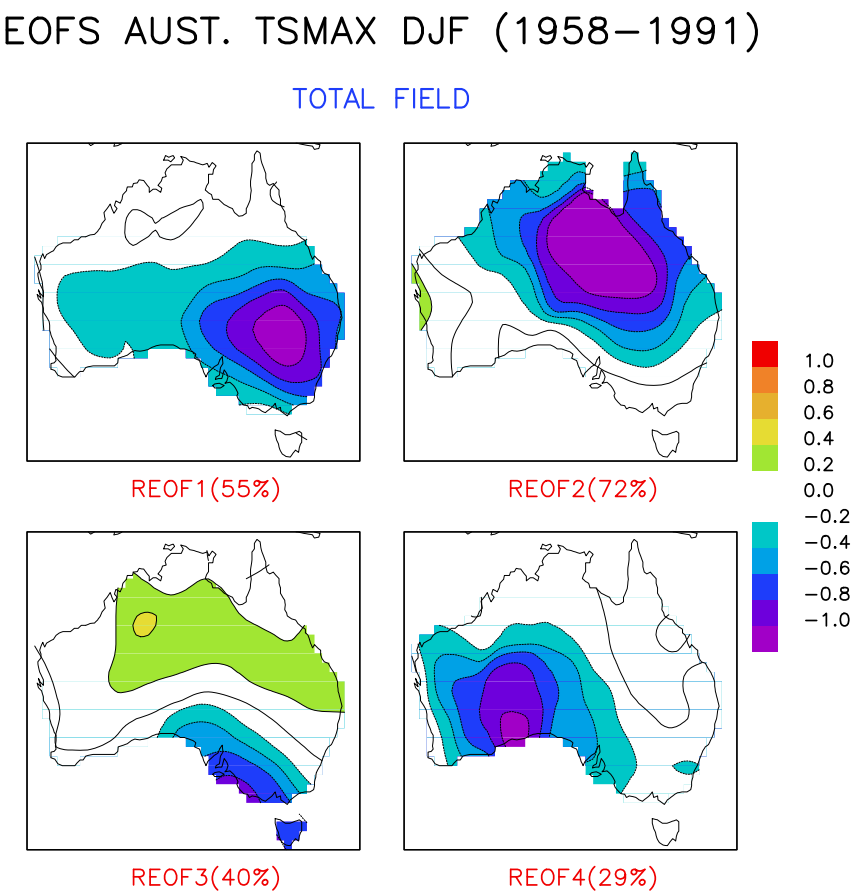

FIGURE 4: The first four dominant rotated EOFs of the total component, potentially predictable and weather components of Australian maximum surface air temperature. The potential predictability appears in brackets.

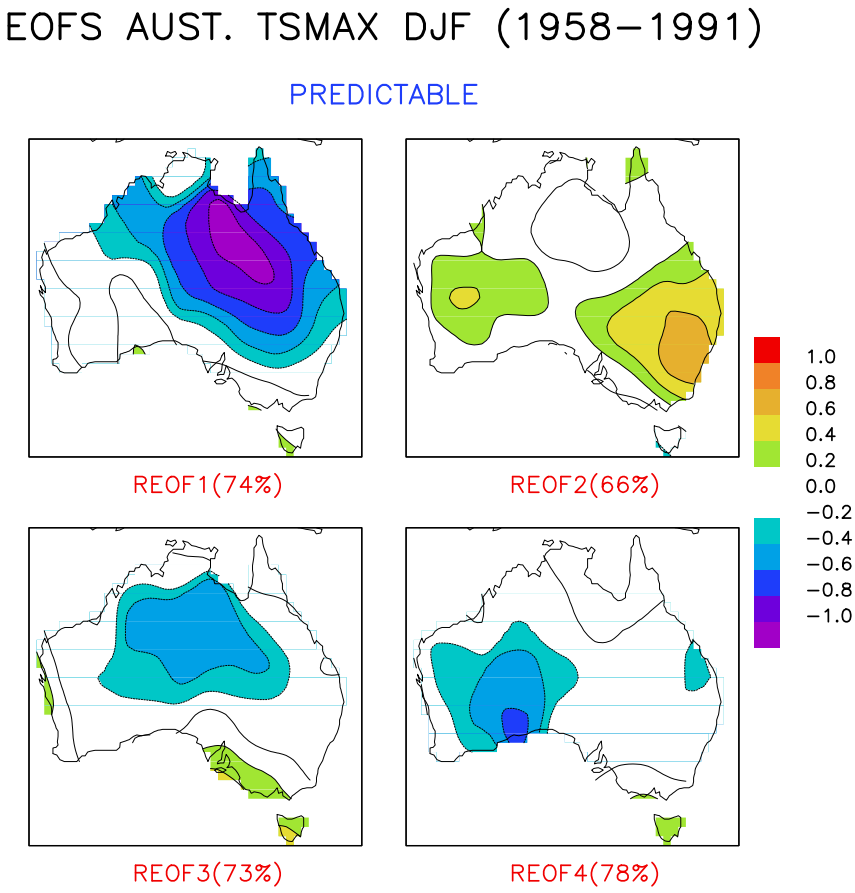

FIGURE 5: As in Figure 4, but for the predictable component.

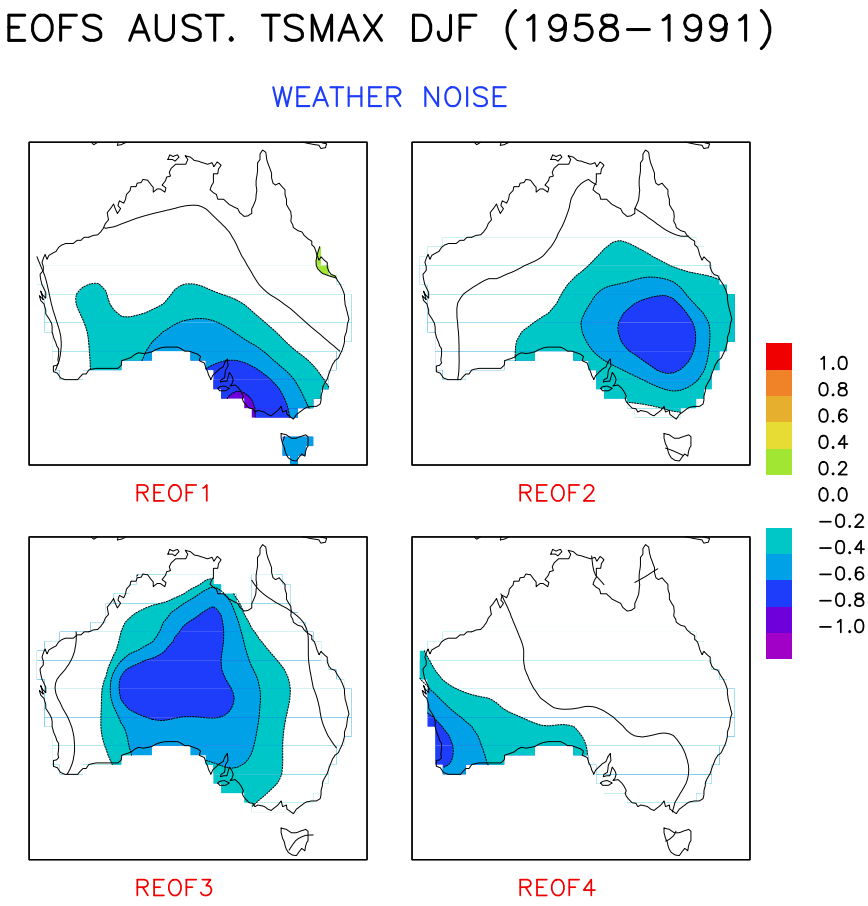

FIGURE 6: As in Figure 4, but for the weather noise component.

This example illustrates the need to separate the weather variability from the total field when studying the predictability of climate variability.

Finally, we shall consider the role of tropical SSTs in forcing the predictable patterns. To this end, as discussed in the last section, we have used multi-linear regression analysis between the predictable-PCs and SST PCs. Before the regression analysis is carried out, the linear trend should be removed from the PC time series. A detrended potential predictability can then be defined as (original potential predictable variability-trend variability)/(original total variability).

It is also convenient to introduce the concepts of SST forced variability and predictability. For our purposes, here, SST forced variability of a predictable-PC is estimated by regressing the PC with the 0-season lag and 1-season lag SST PCs, and SST forced predictability by regressing the PC with only the 1-season lag SST. Both are expressed as a percentage of the variance in the predictable-PC. Table 1 summarizes the results of the regression analysis.

Predictable-PC 1, which represents maximum temperature variations over northeastern Australia, has no statistically significant SST predictors, even at the 10% level. Thus, its predictability comes from other sources including slowly varying internal dynamics. Predictable-PC 2, has significant predictors in 1-season lag Pacific mode 1 (pac1.1) and Indian mode 1 (ind1.1), and quite high SST forced variability and predictability of 39% (see Table 1). Also, of the 54% detrended potential predictability, about 72% (39/54) is due to tropical SST-forcing. Pacific mode 1 is the well-known ENSO mode of Pacific SST interannual variability, and hence predictable-PC 2 is also strongly related to ENSO. In addition, the phase of predictable-PC 2 in DJF is predictable from tropical SSTs in the preceding season SON.

Potentially predictable patterns	SST PC predictors	% SST-forced variability/ predictability	% potential predictability (detrended)
PC1			69
0-1 lag			
1 lag			
PC2			54
0-1 lag	*-pac1.1 -ind1.1*	39	
1 lag	*-pac1.1 -ind1.1*	39	
PC3			73
0-1 lag	*-ind0.1 +ind1.3 +ind1.1*	39	
1 lag	*+pac1.1 +ind1.1 +ind1.3* +ind1.2	26	
PC4			74
0-1 lag	*+pac1.2* -ind0.3	29	
1 lag	*pac1.2*	19	

TABLE 1: SST predictors of the potentially predictable patterns of inter- annual variability. Predictors significant at 1% are indicated in *italics*; predictors significant at 5% in normal type.

Predictable-PC 3, which involves an out of phase relationship between maximum temperature variability over central and southeast Australia, has SST forced variability (39%) which is dominated by Indian Ocean interannual variability in 0-season and 1-season lag mode 1 and 1-season lag mode 3 (see Table 1). SST forced predictability is a little lower at 26%, and also involves a significant predictor in the 1-season lag ENSO mode. Again, the phase of this mode is predictable from a knowledge of SON tropical SSTs. Of the 73% potential predictability, about 53% is related to tropical SST forcing. This is less than for PC 2 and reflects a larger role of low frequency internal dynamics.

Predictable-PC 4, involves maximum temperature variations over the southwest of Australia, and has as its most significant predictor Pacific SST mode 2 at 1-season lag. This SST mode (see Figure 3) represents a transition between the warm and cold phases of ENSO. There is a secondary predictor (at 5% significance level) in 0-season Indian mode 3, which represents a dipole in SST between the central Indian Ocean and Indonesian archipelago. SST forced variability (Table 1) is still quite large at 29% (39% of the potential predictability) and there is some predictability (19%) of the phase from SON Pacific SSTs.

5 Conclusions

In this paper, we have examined a technique for decomposing the spatial covariance matrix of a seasonal mean climate variable into a potentially predictable matrix and unpredictable weather noise component. We have applied this technique to an Australian maximum surface temperature time series and have shown that the method enables a clearer identification of the potentially predictable char-

acteristics of a climate variable at inter-annual, or greater, time-scales. Also, the method provides for a deeper understanding of the differing roles of the predictable and weather noise components at these time-scales.

References

[1] P. J. Brockwell and R. A. Davis, 1987: Time Series Theory and Methods. *Springer-Verlag,* 577pp. C165

[2] I. T. Jolliffe, 1986: Principal Component Analysis. *Springer Verlag,* 271pp. C166, C169

[3] D. A. Jones, 1999: Characteristics of Australian Land Surface Temperature Variability. *Theor. Appl. Climatol.*, 63, 11–31. C172

[4] D. A. Jones and B. C. Trewin, 2000: The spatial structure of monthly temperature anomalies over Australia. *Aust. Met. Mag.*, 49, 261–276. C167

[5] R. A. Madden, 1976: Estimates of the natural variability of time averaged sea level pressure. *Mon. Wea. Rev.*, 104, 942–952. C161

[6] B. C. Trewin, 1999: The development of a high-quality daily temperature data set for Australia, and implications for the observed frequency of extreme temperatures. *Proc. Sixth Nat. AMOS Conf.,* Canberra ACT, Aust., 87pp. C167

[7] X. Zheng and C. S. Frederiksen, 1999: Validating Interannual Variability in an ensemble of AGCM Simulations. *J. Climate,* 12, 2386–2396. C162, C164