A method for constructing skillful seasonal forecasts using slow modes of climate variability

C. S. Frederiksen[1]   X. Zheng[2]

Abstract

A methodology for constructing skillful statistical seasonal forecasts of climate fields is described and applied to predict the Southern Hemisphere summer mean sea level pressure anomalies for the period 1993–2004. The method employs a recently developed variance decomposition approach, which allows a separation of the predictable and unpredictable components of climate variation. The proposed forecast scheme is based on finding predictors for the amplitude time series of the dominant slow (or predictable) modes of interannual climate variation.

Contents

1 Introduction

The development of statistical seasonal climate forecast schemes is largely based on an understanding of lagged relationships between the climate variable being forecasted and some source of external forcing (for example, sea surface temperature (SST)) or slowly varying internal atmospheric variability (for example, the Southern Annular Mode) a month, or more, in advance. The seasonal mean of many climate variables can be thought of as consisting of two components [7, e.g.]; one is related to slowly varying boundary, or external, forcings on the climate system (for example, SST, sea-ice coverage and greenhouse gas concentration) and from slowly varying internal atmospheric variability; the other is related to climatic noise. For the purposes of long-range forecasting, the former is generally considered potentially predictable, in that the forcings are themselves potentially predictable. The latter is related to meteorological phenomena that vary significantly within the season (for example, storms and atmospheric blocking, or intraseasonal variability associated with the Madden-Julian Oscillation) and is essentially

unpredictable, or chaotic.

Zheng and Frederiksen [7, 1, 2] developed a methodology for estimating, from monthly mean data, spatial patterns of these two components, which they referred to as the slow and intraseasonal components. The method provides estimates of the covariance matrices, of the two components, for a subsequent principal component analysis [5]. Recently, Zheng and Frederiksen [8] showed how the spatial patterns of the slow components could be used to improve climate prediction. In particular, they were able to double the predictive skill of New Zealand rainfall forecasts using a statistical prediction scheme based on the prediction of the principal component time series of the slow components of rainfall variability. This article outlines a method for constructing skillful statistical seasonal forecasts of climate variables based on these ideas. We illustrate the efficacy of the method by using it to construct a forecast scheme for the prediction of the summer (December–February, SDJF) Southern Hemisphere (SH) mean sea level pressure (MSLP).

2   Methodology

2.1   Decomposition of covariability

Zheng and Frederiksen [7, 1, 2] proposed and detailed a methodology for extracting, from monthly mean climate data, spatial patterns of interannual variability in seasonal mean fields that is related to variability of slow and intraseasonal components. Here, we summarize the method. Let $x_{ym}(r)$ represent monthly anomalies of a climate variable in year y $(= 1, \ldots, Y)$, month m $(= 1, 2, 3$ corresponding to December, January, February, respectively) and at some location $r = 1, \ldots, R$. Then, following, for example, [7], we assume

$$x_{ym}(r) = \mu_y(r) + \epsilon_{ym}(r) \,, \tag{1}$$

where $\mu_y(r)$ represents the slow, or more potentially predictable component, and $\epsilon_{ym}(r)$ the intraseasonal, or essentially unpredictable. The vector $(\epsilon_{y1}(r), \epsilon_{y2}(r), \epsilon_{y3}(r))$ is assumed to comprise a stationary and independent random vector with respect to year. Equation (1) implies that month-to-month fluctuations, or intraseasonal variability, arise entirely from this component; for example, $x_{y1}(r) - x_{y2}(r) = \epsilon_{y1}(r) - \epsilon_{y2}(r)$. We use the convention that an average over any index will be represented by a circle. Thus, for example, $x_{y\circ}(r)$ is an average over all months in the season, and $x_{\circ\circ}(r)$ is an average over all months and years. The symbol V denotes the covariance of the climate field at two different locations, or the variance if at the same location.

Hence, the seasonal mean of x is

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

Here, $\epsilon_{y\circ}(r)$ represents that component of the seasonal mean that is associated with intraseasonal variability, and $\mu_y(r)$ with the interannual variability of external forcings and slowly varying (interannual/supra-annual) internal dynamics [7].

Zheng and Frederiksen [7] derived the following estimate of the covariance $V(\epsilon_{y\circ}(r_1), \epsilon_{y\circ}(r_2))$ for the intraseasonal component at locations r_1 and r_2, respectively,

$$V(\epsilon_{y\circ}(r_1), \epsilon_{y\circ}(r_2)) = \frac{\alpha(3 + 4\beta)}{9} \,, \tag{3}$$

where

$$\alpha = a + b \,, \tag{4}$$

$$\beta = \frac{a + 2b}{2(a + b)} \,, \tag{5}$$

$$a = \frac{1}{2}\left\{ \frac{1}{Y} \sum_{y=1}^{Y} [x_{y1}(r_1) - x_{y2}(r_1)] [x_{y1}(r_2) - x_{y2}(r_2)] \right.$$

$$+ \frac{1}{Y} \sum_{y=1}^{Y} \left[x_{y2}(r_1) - x_{y3}(r_1) \right] \left[x_{y2}(r_2) - x_{y3}(r_2) \right] \right\}, \tag{6}$$

$$b = \frac{1}{2} \left\{ \frac{1}{Y} \sum_{y=1}^{Y} \left[x_{y1}(r_1) - x_{y2}(r_1) \right] \left[x_{y2}(r_2) - x_{y3}(r_2) \right] \right.$$

$$+ \frac{1}{Y} \sum_{y=1}^{Y} \left[x_{y2}(r_1) - x_{y3}(r_1) \right] \left[x_{y1}(r_2) - x_{y2}(r_2) \right] \right\}. \tag{7}$$

In order to reduce the estimation error in equation (3), Zheng and Frederiksen [7] provide a detailed proof that the estimate for β (equation (5)) has to be constrained to lie in the interval $[0, 0.1]$.

The covariance $V(x_{y\circ}(r_1), x_{y\circ}(r_2))$ of the seasonal mean, which we refer to as the total covariance, is estimated simply by the sample covariance:

$$V(x_{y\circ}(r_1), x_{y\circ}(r_2)) = \frac{1}{Y-1} \sum_{y=1}^{Y} \left[x_{y\circ}(r_1) - x_{\circ\circ}(r_1) \right] \left[x_{y\circ}(r_2) - x_{\circ\circ}(r_2) \right]. \tag{8}$$

Thus, using equations (3) and (8), define the residual covariance

$$V(x_{y\circ}(r_1), x_{y\circ}(r_2)) - V(\epsilon_{y\circ}(r_1), \epsilon_{y\circ}(r_2))$$
$$= V(\mu_y(r_1), \mu_y(r_2)) + V(\mu_y(r_1), \epsilon_{y\circ}(r_2)) + V(\mu_y(r_2), \epsilon_{y\circ}(r_1)). \tag{9}$$

In the case where the intraseasonal and slow components are independent, the residual covariance reduces to the covariance of the slow components only. Even when this is not true, the residual covariance can be shown to be better related to the covariance between the slow components than is the total covariance because the weather noise component has been largely removed. Equations (3, 8, 9) underlie the construction of the corresponding covariance matrices for all pairs of grid points. The eigenvectors of these matrices, scaled to be orthonormal, and commonly referred to as empirical orthogonal functions (EOFs), provide the spatial patterns of interannual variation in each component. The eigenvalues indicate the explained variance associated with

the eigenvector [7, 2, 5, e.g.]. Associated with these patterns, one can derive time series showing how the sign of the patterns varies year by year. These time series are derived by projecting the original monthly data in each year onto each pattern [2, e.g.] and will be referred to as Principal Component (PC) time series.

2.2 Prediction

The essence of our prediction strategy is to focus on the slow, or potentially predictable, eigenvectors (or patterns) of interannual variations and try to derive predictors for the corresponding PC time series. Using these slow eigenvectors as a basis, and the PC time series, forecasts of variations in the climate field are constructed as follows.

Let column vector $\mathbf{x}_{ym} = (x_{ym}(1), \ldots, x_{ym}(R))$ denote the climate field anomaly in month m and in year y and $\mathbf{V}$ denote the matrix with columns that are the unit norm eigenvectors of the residual covariance matrix, equation (9), constructed using the data from the training period. Let the row vector $\mathbf{p}_y$ denote the projection of the climate field onto $\mathbf{V}$ in year y, that is,

$$\mathbf{p}_y \equiv \mathbf{x}_{y\circ}^T \mathbf{V}, \qquad (10)$$

where the row vector $\mathbf{p}_y$ consists of the amplitudes of the slow PC time series in year y and $\mathbf{x}_{y\circ}^T$ denotes the transpose. By construction, matrix $\mathbf{V}$ is orthonormal and therefore the climate field

$$\mathbf{x}_{y\circ}^T = \mathbf{p}_y \mathbf{V}^T. \qquad (11)$$

It follows that, if the slow PC time series can be predicted, a prediction of the climate anomaly field is constructed using equation (11). That is, a prediction of the amplitude of each, or a subset, of the slow PC time series is used to make a prediction, or forecast, of the climate anomaly field itself.

2.3 Predictive skill

We use two measures of skill to evaluate our forecasts. The most commonly used measure of skill, s, of a forecast, relative to a baseline forecast (usually climatology), is defined as the ratio of sums of squares over the verification period: that is,

$$s = \frac{\sum_{r=1}^{R} \sum_{y=1}^{Y_v} (b_y(r) - o_y(r))^2}{\sum_{r=1}^{R} \sum_{y=1}^{Y_v} (p_y(r) - o_y(r))^2} , \tag{12}$$

where Y_v is the total number of years in the verification period; and $o_y(r)$, $p_y(r)$ and $b_y(r)$ are the anomaly of the observed, predicted and baseline climatology forecasted seasonal mean climate field in year y and at location r, respectively. It follows from equation (12) that s is greater than or equal to zero and (a) identically 1 if the prediction is no better than climatology, (b) greater than 1 if the prediction is better than climatology and (c) less than 1 if worse than climatology. Also, the percentage of explained variance above climatology is defined as

$$V_{\text{expl}} = 100\Big(1 - \frac{1}{s}\Big). \tag{13}$$

Another important skill measure is the normalized spatio-temporal correlation (stc) which gives an indication of how well the climate field anomaly pattern is reproduced over the verification period:

$$\text{stc} = \frac{\sum_{y=1}^{Y_v} \sum_{r=1}^{R} (p_y(r) - \bar{p})(o_y(r) - \bar{o})/\nu(r)}{\sqrt{\sum_{y=1}^{Y_v} \sum_{r=1}^{R} (p_y(r) - \bar{p})^2/\nu(r) \sum_{y=1}^{Y_v} \sum_{r=1}^{R} (o_y(r) - \bar{o})^2/\nu(r)}} . \tag{14}$$

Here, $\bar{p}$ and $\bar{o}$ represent the predicted and observed anomaly averaged over all years, in the verification period, and over all grid points, and $\nu(r)$ is the variance of the observed seasonal mean at grid point r estimated for the training period.

3 Prediction of summer MSLP

This section illustrates the methodology by using it to construct a statistical prediction scheme for the summer SH MSLP field. The dataset used is the National Centers for Environmental Prediction (NCEP) and National Center for Atmospheric Research re-analyses (Kalnay et al. [3]) for 1948–2004. The data has been sub-sampled on a $5° \times 5°$ latitude/longitude grid. The data for the period December 1952 through February 1992 are used as the training period for the statistical scheme. The verification period is taken as 1993–2004 for the forecasts of the MSLP. To help identify important SST forcing (or predictors) of the MSLP patterns, we used the UK Meteorological Office Hadley Centre $1° \times 1°$ (latitude/longitude grid) HADISST1.1 dataset (Rayner et al. [4]).

Figure 1 shows the first four dominant slow summer EOFs of MSLP variability derived for the training period. They explain 46%, 8%, 7% and 6%, respectively, of the interannual variance and form the basis of our prediction scheme. By basing our prediction scheme on the slow EOFs, we effectively remove the influence of unpredictable modes of variability on the potential skill of our scheme.

The aim then is to estimate the slow PC time series in terms of a set of relevant predictors and to use this functional relationship to estimate the MSLP anomalies in the verification period. Here, we estimate these relationships using multivariate linear regression [5]. We look for possible predictors based on atmospheric circulation features and SSTs at one season and one month lead, that is, for September-October-November (SON) and November. Stepwise regression [5] selects those predictors that are most appropriate to train the regression coefficients. The correlation maps between the slow PCs and SON SST are shown in Figure 2. Similar maps (not shown) were constructed for November SSTs.

Slow EOF1 is a pattern of variation associated with the Southern Annular
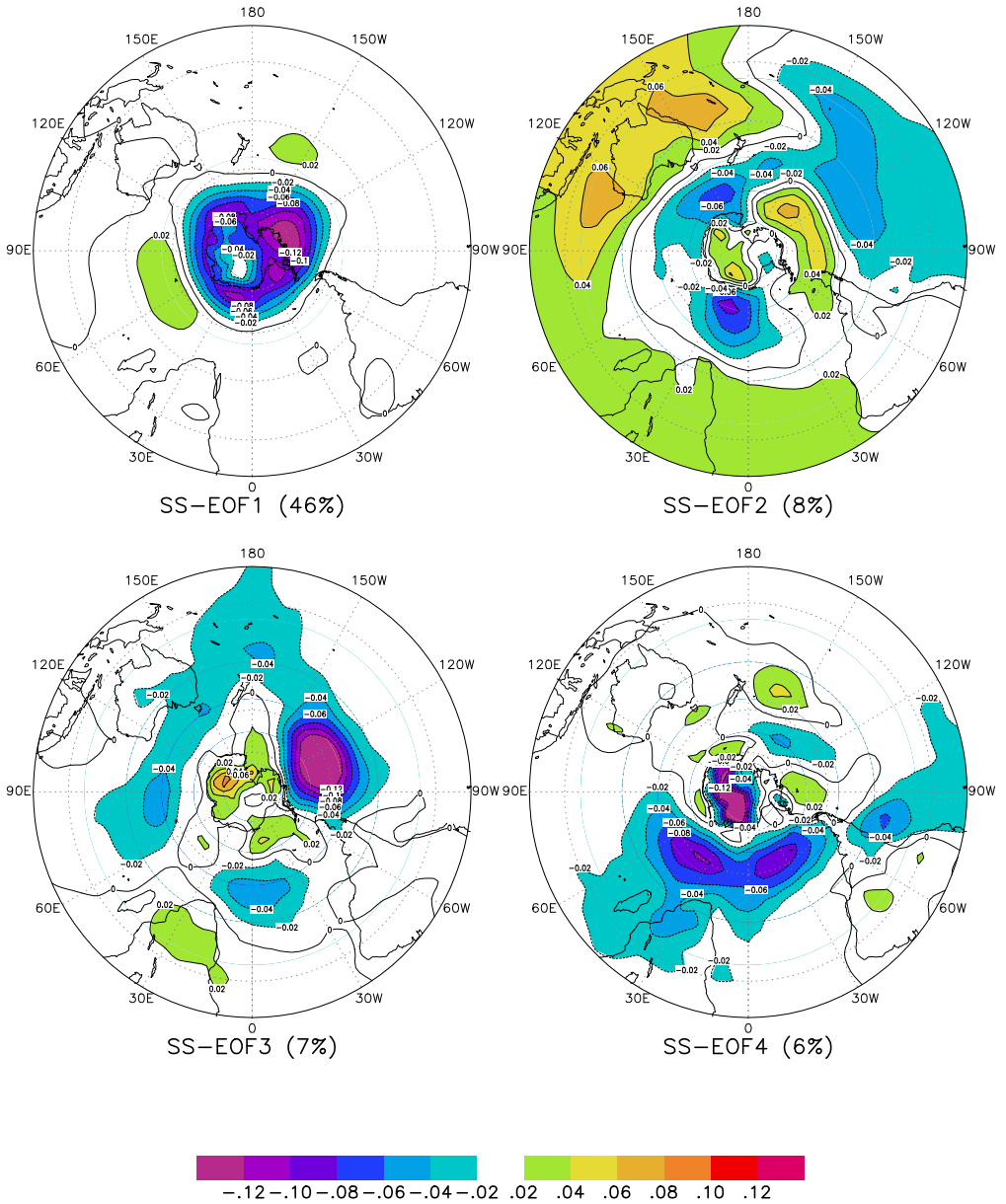
FIGURE 1: The four most dominant slow summer (SS) EOFs of interannual variability in the SDJF SH MSLP field.
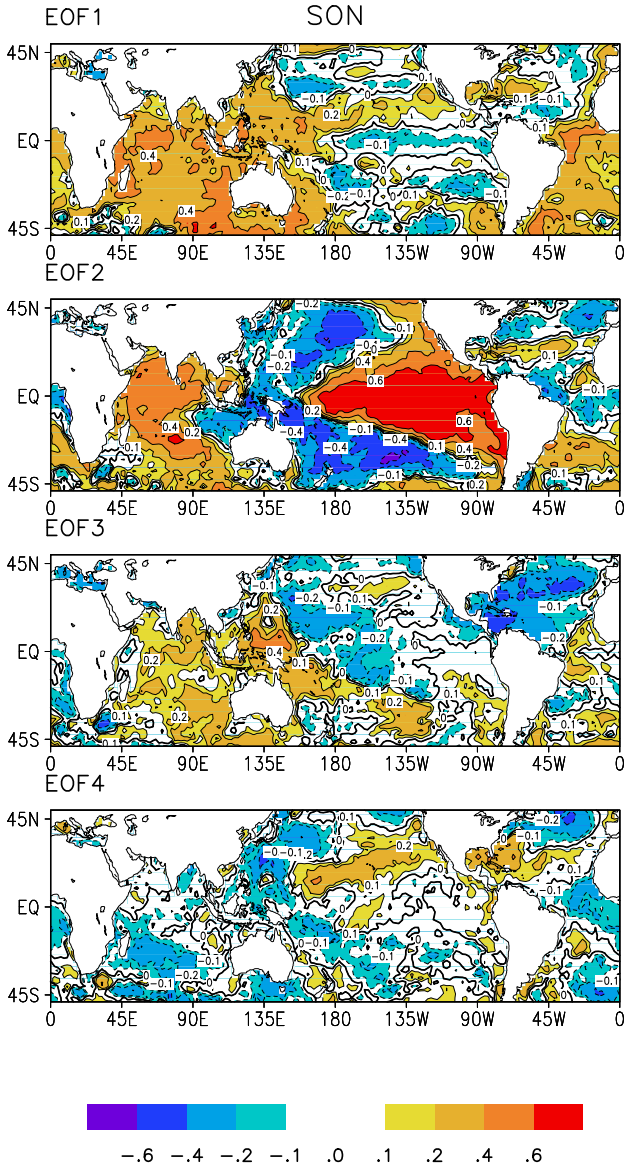
FIGURE 2: The correlation between observed SON sea surface temperature and the PCs of the slow components of SH MSLP variability.

TABLE 1: Statistical prediction

Predictand	Prediction Formulae
PC1	$-6.28587 + 0.71680 \times \text{trend} + 0.19973 \times \text{Nov.SAMI}$
PC2	$-0.60345 + 0.12945 \times \text{SON.NINO3.SST}$
PC4	$-32.7514 - 0.09602 \times \text{Nov.Coral Sea.SST}$

Mode (SAM) [6] that dominates the SH variability in the MSLP. Although the correlation between the PC time series (PC1) and SSTs is fairly high in the tropical Indian Ocean, this is mainly due to positive trends in both (not shown). Possible predictors for PC1 are the SON and November SAM index (SAMI)[1] and the linear trend calculated over the training period. In this case, the regression analysis shows (see Table 1) that the November SAMI and the trend are the best predictors.

SST correlations are highest for slow PC2, with the correlation map showing largest correlation in the eastern Pacific Ocean. Our analysis shows that the SON Nino-3 SST index (defined as the average SST anomaly over the region 150°W–90°W, 5°N–5°S) is the best predictor (see Table 1). For slow PC3, we were not able to find any significant SST or circulation predictors, so have left this out of our consideration. However, for slow PC4, November SST variations in the Coral Sea were found to be important and significantly related to a Coral Sea SST index (defined as the average SST over the region 150°E–180°E, 15°S–30°S).

For the verification period, Figure 3 shows the observed and predicted PC time series. The predicted PC time series appear to be well correlated with the observed, but with a negative bias in PC1. This bias is due to a smaller trend in the earlier training period than in the verification period. However, the skill of the forecasts of the MSLP anomaly, constructed from slow PC1, PC2 and PC4 and their corresponding slow EOFs, using equation (11), is quite high

---

[1]Obtainable via http://www.jisao.washington.edu/aao/ from Thompson and Wallace [6].
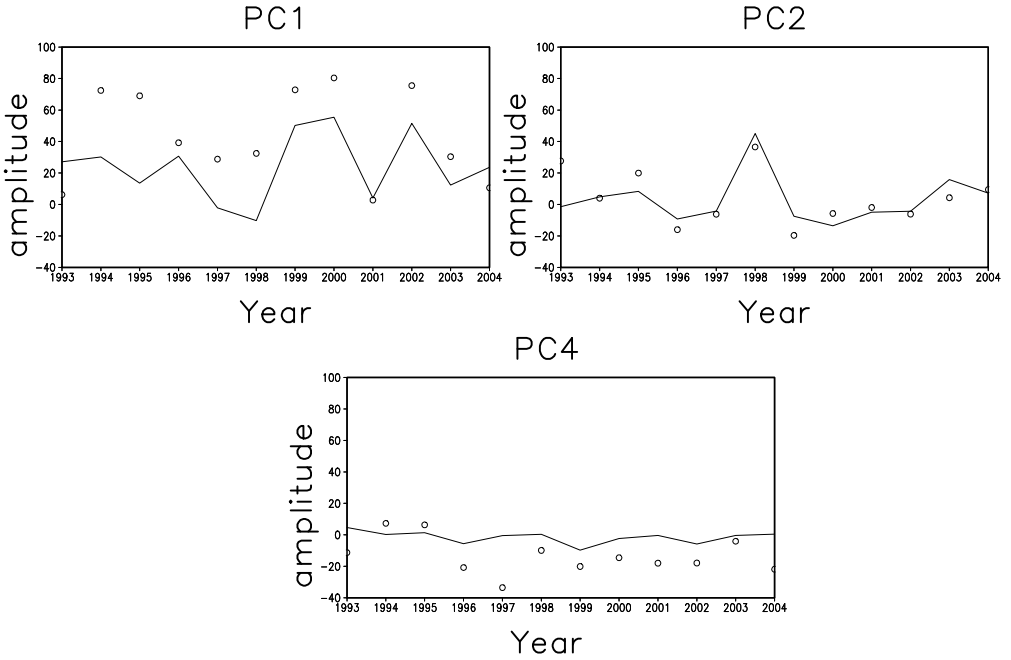
FIGURE 3: The observed (open circle) and predicted (solid line) PC values in the verification period for SDJF.

as seen in Table 2. Thus, for example, the percentage explained variance, above climatology, is 44.5% and the stc is 0.56. Also shown in Table 2 are the skill scores for different regions of the Southern Hemisphere, and for different combinations of the slow PCs. Generally, the skill is maximized when all PCs are used, although most of the skill appears to come from PC1 and PC2.

TABLE 2: Forecast skill scores

	% Explained Variance			STC		
	PC1	PC1,2	PC1,2,4	PC1	PC1,2	PC1,2,4
SH	44.0	43.5	44.5	0.37	0.54	0.56
0–22.5°S	29.0	45.8	46.9	0.37	0.59	0.60
17.5–32.5°S	7.2	39.4	40.3	0.36	0.43	0.44
32.5–62.5°S	36.7	41.6	42.3	0.38	0.39	0.41
62.5–90°S	55.3	22.3	22.9	0.26	0.56	0.58
Australia	21.6	43.9	44.6	0.38	0.63	0.63
New Zealand	1.1	41.5	42.0	0.40	0.40	0.41

4   Conclusions

We outlined a procedure for constructing skillful statistical forecast schemes based on extracting from climate data the EOF patterns of interannual variability in the slow component. By concentrating on this component, which has been shown in other studies to be more potentially predictable [7, 1, 2], there is the possibility of optimizing the skill of the forecasts. The method is based on the prediction of the slow PC time series, from which, together with the corresponding slow EOFs, it is possible to construct a forecast of the climate field itself. We illustrated the procedure by constructing a statistical forecast scheme for the SH MSLP, which has significantly higher skill (44.5% explained variance) than forecasts based on climatology. The forecasts also largely capture the spatial pattern of the MSLP anomaly during the verification period as indicated by a high spatio-temporal correlation of 0.56.

References

[1] C. S. Frederiksen and X. Zheng. Variability of seasonal-mean fields arising from intraseasonal variability. Part 2, application to NH winter

circulations. *Climate Dynamics*, 23:193–206, doi:10.1007/s00382-004-0429-6, 2004. C91, C101

[2] C. S. Frederiksen and X. Zheng. A method for studying the influence of intra-seasonal variability on the inter-annual variability of climate fields. *ANZIAM J.*, 45(E):C378–C390, http://anziamj.austms.org.au/V45/CTAC2003/Fred, 2004. C91, C94, C101

[3] E. M. Kalnay et al. The NCEP/NCAR 40-year reeanalysis project. *Bull. Amer. Meteor. Soc.*, 77:437–471, doi:10.1175/1520-0477(1996)077¡0437:TNYRP¿2.0.CO;2, 1996. C96

[4] N. A. Rayner, D. E. Parker, C. K. Folland, L. V. Alexander, E. B. Horton and D. P. Rowell. Globally complete analyses of sea-surface temperature, sea-ice and marine air temperature, 1871-2000. *J. Geophys. Res.*, 180:4007, doi:10.1029/2002JD002670, 2003. C96

[5] H. von Storch and F. W. Zwiers. Statistical Analysis in Climate Research. *Cambridge University Press*, 484pp, 1999. C91, C94, C96

[6] D. W. J. Thompson and J. M. Wallace. Annular Modes in the Extratropical Circulation. Part I: Month-to-Month Variability. *J. Climate*, 13:1000–1016, doi:10.1175/1520-0442(2000)013¡1000:AMITEC¿2.0.CO;2, 2000. C99

[7] X. Zheng and C. S. Frederiksen. Variability of seasonal-mean fields arising from intraseasonal variability. Part 1, methodology. *Climate Dynamics*, 23:177–191, doi:10.1007/s0038-004-0428-7, 2004. C90, C91, C92, C93, C94, C101

[8] X. Zheng and C. S. Frederiksen. A Study of Predictable Patterns for Seasonal Forecasting of New Zealand Rainfall. *J. Climate*, 19:3320–3333, doi:10.1175/JCLI3798.1, 2006. C91

Author addresses

1. **C. S. Frederiksen**, Bureau of Meteorology, GPO Box 1289, Melbourne, Victoria 3001, Australia.
   mailto:c.frederiksen@bom.gov.au

2. **X. Zheng**, National Institute of Water and Atmospheric Research, Wellington, New Zealand.