

A method for studying the influence of intra-seasonal variability on the inter-annual variability of climate fields

C. S. Frederiksen* X. Zheng†

(Received 8 August 2003, revised 30 January 2004)

Abstract

A computational technique is proposed for studying patterns of interannual variability of climate seasonal means that arise from intraseasonal variability. The method, using monthly means of climate data, is more computationally efficient than other current methods that use daily means to estimate the spatial covariability of the intraseasonal component. Removing this component, it is possible to estimate the long-range potentially predictable patterns of climate variability. The method is applied to a study of the Southern Hemisphere winter circulation and compared with a method using daily data to show that the results are very similar.

*Bureau of Meteorology, GPO Box 1289K, Melbourne, Victoria 3001, AUSTRALIA.
<mailto:c.frederiksen@bom.gov.au>

†National Institute of Water and Atmospheric Research, Wellington, NEW ZEALAND
See <http://anziamj.austms.org.au/V45/CTAC2003/Fred> for this article, © Austral. Mathematical Soc. 2004. Published June 8, 2004. ISSN 1446-8735

Contents

1 Introduction	C379
2 Methodology	C382
3 Example	C385
4 Conclusions	C388
A Principal component analysis	C388
References	C389

1 Introduction

The potential for long-range (in advance of three months) predictability of regional or global climate is often predicated on a knowledge and understanding of the interannual variability of seasonal means of climate fields. Even for seasonal means, a significant proportion of the interannual variability can be attributed to internal dynamical processes which are inherently unpredictable, on the seasonal time-scale, and are related to the weather within the season. Thus, a common theoretical framework is to consider a decomposition of the temporal variability into an unpredictable part called the “weather noise” variability and a part which is assumed to be at least potentially predictable [4, 2, 5, 6, for example]. This latter component is more likely to be related to external forcings (such as, for example, sea surface temperature anomalies, sea-ice anomalies, varying greenhouse gas concentrations) and very low frequency (interannual to supra-annual) internal atmospheric variability which are themselves potentially predictable.

Recently, Frederiksen et al. [2] proposed a method for extracting spatial

patterns of interannual variability associated with both the weather noise and potentially predictable components of seasonal mean time series. Their method uses daily time series within each season, over many decades, to approximate the spatial covariance matrix for the weather noise component. From this covariance matrix the spatial patterns can be derived using a standard Principal Component (PC) analysis (see Appendix A for a summary of the PC analysis technique or [3] for a detailed description). The potentially predictable patterns can then be derived from a PC analysis applied to a residual covariance matrix, which is the difference between the total covariance matrix, based on the seasonal means of the climate data, and the weather noise matrix.

In brief, Frederiksen et al. [2] assume that, after removing the mean annual cycle, a daily climate anomaly $x_{yt}(r)$ can be represented as

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

and the seasonal mean anomaly as

$$x_{yo}(r) = \mu_y(r) + \varepsilon_{yo}(r). \quad (2)$$

Here: $y = 1, \dots, Y$ is the year; $t = 1, \dots, T$ is the day in a season of length T days; $r = 1, \dots, R$ 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 $\{\varepsilon_{yt} \mid t = 1, \dots, T\}$ 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 . An average over any index will be represented by a circle. Thus, in Equation (2), $x_{yo}(r)$ is an average over T days, and $x_{oo}(r)$ is an average over T days and Y years. The symbol V will be used to denote the covariance of two variables.

Frederiksen et al. [2] show that the covariance $V(\varepsilon_{yo}(r_1), \varepsilon_{yo}(r_2))$ of seasonal weather noise between two locations r_1 and r_2 can be estimated by

$$V(\varepsilon_{yo}(r_1), \varepsilon_{yo}(r_2)) \simeq \frac{1}{YT^2} \sum_{y=1}^Y \Re \left(\left[\sum_{t=1}^T x_{yt}(r_1) e^{it2\pi/T} \right] \left[\sum_{t=1}^T x_{yt}(r_2) e^{-it2\pi/T} \right] \right). \quad (3)$$

The covariance $V(x_{yo}(r_1), x_{yo}(r_2))$ of the seasonal mean anomaly, which we will refer to as the total covariance matrix, can be estimated by the sample covariance

$$V(x_{yo}(r_1), x_{yo}(r_2)) = \frac{1}{Y-1} \sum_{y=1}^Y [x_{yo}(r_1) - x_{oo}(r_1)] [x_{yo}(r_2) - x_{oo}(r_2)]. \quad (4)$$

Thus, from Equation (2), a residual covariance matrix can be defined as

$$\begin{aligned} V(x_{yo}(r_1), x_{yo}(r_2)) - V(\varepsilon_{yo}(r_1), \varepsilon_{yo}(r_2)) &= V(\mu_y(r_1), \mu_y(r_2)) \\ &+ V(\mu_y(r_1), \varepsilon_{yo}(r_2)) + V(\mu_y(r_2), \varepsilon_{yo}(r_1)). \end{aligned} \quad (5)$$

When $\mu_y(r)$ and $\varepsilon_{yo}(r)$ are independent, the residual matrix reduces to $V(\mu_y(r_1), \mu_y(r_2))$, the covariance matrix for the potentially predictable component. Even when this is not the case, the spatial patterns associated with the residual covariance matrix can be shown to be more potentially predictable than those from the total covariance matrix, because the weather noise component has been largely removed. Equations (3–5) can be used to construct the corresponding covariance matrices from which the spatial patterns can be derived using PC analysis.

The crucial step is the estimation of the covariance $V(\varepsilon_{yo}(r_1), \varepsilon_{yo}(r_2))$. In this paper, we propose a simpler method using only monthly data. This is possible because the power spectrum of large-scale atmospheric motions is fundamentally red [6]. Consequently, a non-zero seasonal mean of the weather noise is largely contributed to by intra-seasonal weather events with monthly or longer time scales, including such phenomena as persistent blocking events in the extratropics and the Madden-Julian Oscillation in the tropics. In other words, the seasonal mean operator is a very effective filter of the higher frequency daily weather events such as midlatitude storms.

There are a number of reasons why a method using monthly means is desirable. Firstly, it is computationally more efficient, dealing with only 3 monthly means for a typical season, compared with 90 daily means. Also, the estimation in Equation (3) requires an assumption of normality for the daily time series. This will not be true for many meteorological variables. Finally, the availability of daily time series is limited in comparison with that of monthly time series.

2 Methodology

A monthly anomaly $x_{ym}(r)$ is represented by the simple linear model [6]

$$x_{ym}(r) = \mu_y(r) + \varepsilon_{ym}(r), \quad (6)$$

where, m (1, 2, or 3) denotes a month within a given 3 month season, and $\varepsilon_{ym}(r)$ the intraseasonal weather noise. The other indices and variables have the same meaning as for Equation (1). The annual cycle for monthly data has been estimated using the method of [1]. Equations (2), (4) and (5) also hold with the understanding that the average over the day index is replaced by an average over the 3 months. The vector $\mathbf{e}^T(r) = (\varepsilon_{y1}(r), \varepsilon_{y2}(r), \varepsilon_{y3}(r))$ is assumed to comprise a stationary and independent annual random vector. The linear regression form (Equation (6)) implies that month-to-month fluctuations, or intra-seasonal variability, arise entirely from $\mathbf{e}(r)$ (for example, $x_{y1}(r) - x_{y2}(r) = \varepsilon_{y1}(r) - \varepsilon_{y2}(r)$).

An estimate of $V(\varepsilon_{yo}(r_1), \varepsilon_{yo}(r_2))$, using monthly means, can be made with the following assumptions. Since the daily time series of a climate variable, within a season, is in general assumed to be stationary, so are the monthly statistics. In particular, the covariance between two locations is assumed to be independent of months. That is,

$$V(\varepsilon_{y1}(r_1), \varepsilon_{y1}(r_2)) = V(\varepsilon_{y2}(r_1), \varepsilon_{y2}(r_2)) = V(\varepsilon_{y3}(r_1), \varepsilon_{y3}(r_2)). \quad (7)$$

The same is assumed to be true for the inter-monthly covariance, such that,

$$V(\varepsilon_{y1}(r_1), \varepsilon_{y2}(r_2)) = V(\varepsilon_{y2}(r_1), \varepsilon_{y3}(r_2)). \tag{8}$$

Also, because daily weather events are unpredictable beyond a week or two, we assume that the intraseasonal components are uncorrelated if they are a month or more apart. That is,

$$V(\varepsilon_{y1}(r_1), \varepsilon_{y3}(r_2)) = 0. \tag{9}$$

With the assumptions in Equations (7–9),

$$E(\mathbf{e}(r_1)\mathbf{e}^T(r_2)) + E(\mathbf{e}(r_2)\mathbf{e}^T(r_1)) = 2\alpha \begin{pmatrix} 1 & \beta & 0 \\ \beta & 1 & \beta \\ 0 & \beta & 1 \end{pmatrix}, \tag{10}$$

where E denotes the expectation value based on all years and

$$\alpha = V(\varepsilon_{ym}(r_1), \varepsilon_{ym}(r_2)), \quad m = 1, 2, 3, \tag{11}$$

and

$$\begin{aligned} \beta &= \frac{1}{2\alpha} [V(\varepsilon_{y1}(r_1), \varepsilon_{y2}(r_2)) + V(\varepsilon_{y1}(r_2), \varepsilon_{y2}(r_1))] \\ &= \frac{1}{2\alpha} [V(\varepsilon_{y2}(r_1), \varepsilon_{y3}(r_2)) + V(\varepsilon_{y2}(r_2), \varepsilon_{y3}(r_1))]. \end{aligned} \tag{12}$$

Following [6], we constrain α to lie within the interval $[0, 0.1]$ in order to reduce the estimation error. Using Equations (11–12),

$$\begin{aligned} &E \left\{ \begin{pmatrix} \varepsilon_{y1}(r_1) - \varepsilon_{y2}(r_1) \\ \varepsilon_{y2}(r_1) - \varepsilon_{y3}(r_1) \end{pmatrix} \begin{pmatrix} \varepsilon_{y1}(r_2) - \varepsilon_{y2}(r_2) \\ \varepsilon_{y2}(r_2) - \varepsilon_{y3}(r_2) \end{pmatrix}^T \right\} \\ &+ E \left\{ \begin{pmatrix} \varepsilon_{y1}(r_2) - \varepsilon_{y2}(r_2) \\ \varepsilon_{y2}(r_2) - \varepsilon_{y3}(r_2) \end{pmatrix} \begin{pmatrix} \varepsilon_{y1}(r_1) - \varepsilon_{y2}(r_1) \\ \varepsilon_{y2}(r_1) - \varepsilon_{y3}(r_1) \end{pmatrix}^T \right\} \\ &= 2\alpha \begin{pmatrix} 2 - 2\beta & 2\beta - 1 \\ 2\beta - 1 & 2 - 2\beta \end{pmatrix}. \end{aligned} \tag{13}$$

From Equation (6), $x_{y1}(r) - x_{y2}(r) = \varepsilon_{y1}(r) - \varepsilon_{y2}(r)$ and $x_{y2}(r) - x_{y3}(r) = \varepsilon_{y2}(r) - \varepsilon_{y3}(r)$. Therefore, the left hand side of Equation (13) can be evaluated using the given data $x_{ym}(r)$. Hence, solving for α and β

$$\alpha = a + b, \tag{14}$$

and

$$\beta = \frac{a + 2b}{2(a + b)}, \tag{15}$$

where

$$a = \frac{1}{2} \left\{ \frac{1}{Y} \sum_{y=1}^Y [x_{y1}(r_1) - x_{y2}(r_1)] [x_{y1}(r_2) - x_{y2}(r_2)] + \frac{1}{Y} \sum_{y=1}^Y [x_{y2}(r_1) - x_{y3}(r_1)] [x_{y2}(r_2) - x_{y3}(r_2)] \right\} \tag{16}$$

$$b = \frac{1}{2} \left\{ \frac{1}{Y} \sum_{y=1}^Y [x_{y1}(r_1) - x_{y2}(r_1)] [x_{y2}(r_2) - x_{y3}(r_2)] + \frac{1}{Y} \sum_{y=1}^Y [x_{y2}(r_1) - x_{y3}(r_1)] [x_{y1}(r_2) - x_{y2}(r_2)] \right\} \tag{17}$$

Using these estimates for α and β in Equation (10),

$$\begin{aligned} & \frac{1}{2} [V(\varepsilon_{y1}(r_1), \varepsilon_{y2}(r_2)) + V(\varepsilon_{y1}(r_2), \varepsilon_{y2}(r_1))] \\ &= \frac{1}{2} [V(\varepsilon_{y2}(r_1), \varepsilon_{y3}(r_2)) + V(\varepsilon_{y2}(r_2), \varepsilon_{y3}(r_1))] \\ &= \alpha\beta. \end{aligned} \tag{18}$$

Thus, taking into account assumption Equation (9), an estimate for the covariance of the intraseasonal weather noise at locations r_1 and r_2 , using

monthly data, is

$$\begin{aligned}
 V(\varepsilon_{yo}(r_1), \varepsilon_{yo}(r_2)) &= \frac{1}{2}[V(\varepsilon_{yo}(r_1), \varepsilon_{yo}(r_2)) + V(\varepsilon_{yo}(r_2), \varepsilon_{yo}(r_1))] \\
 &= \frac{1}{18} \sum_{m,n=1}^3 [V(\varepsilon_{ym}(r_1), \varepsilon_{yn}(r_2)) + V(\varepsilon_{ym}(r_2), \varepsilon_{yn}(r_1))] \\
 &= \frac{\alpha(3 + 4\beta)}{9}.
 \end{aligned} \tag{19}$$

In the next section, we show that using either Equation (3) or Equation (19) gives very similar results.

3 Example

Here, we apply both methods (Equation (3) and Equation (19)) to a study of the Southern Hemisphere (SH) wintertime (June-July-August (JJA)) atmospheric circulation. The data we use is the 500 hPa geopotential height taken from the National Centers for Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR) re-analysis data. This field effectively gives the height of the 500 hPa pressure level and is often used to study spatial patterns, or teleconnections, of interannual variability in the atmospheric circulation. The data runs from latitude 90°S–20°S on a 5° × 5° latitude/longitude grid and is for the period 1958–1996.

Figure 1 shows the first four dominant intraseasonal weather patterns using both daily (left column) and monthly (right column) data. Both sets of patterns are remarkably similar with only subtle differences reflected in high pattern correlations of 0.99, 0.98, 0.93 and 0.93 for patterns 1–4, respectively. Using either Equation (3) or Equation (19) in Equation (5) for the residual matrix, also produces remarkably similar potentially predictable patterns (Figure 2) with similar high pattern correlations of 0.99, 0.99, 0.94

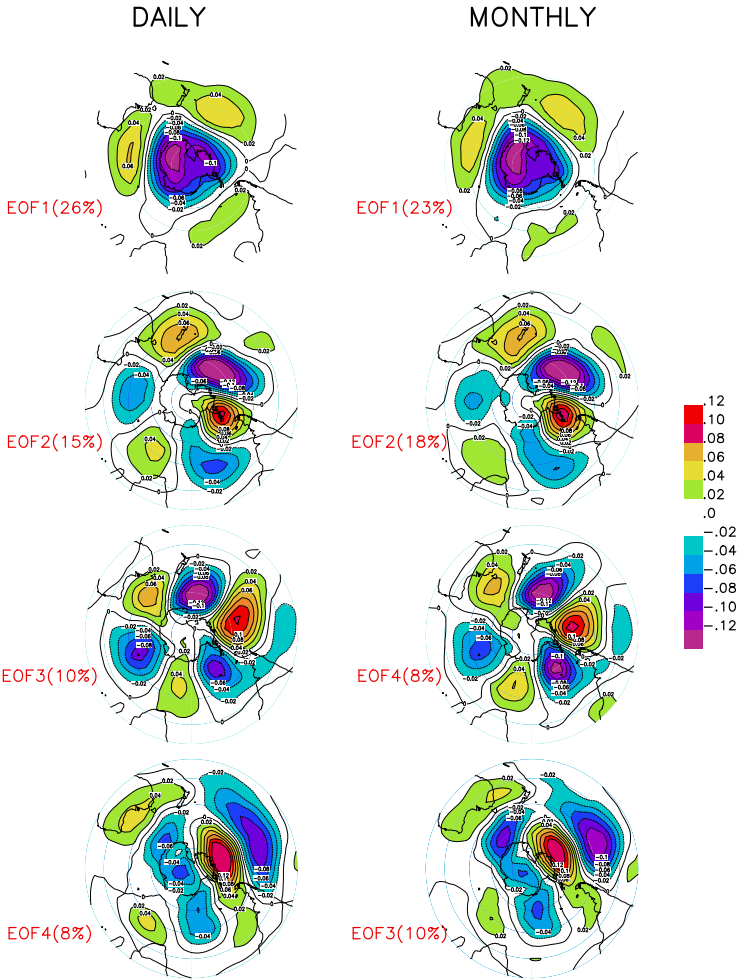


FIGURE 1: The first four dominant JJA patterns of the intraseasonal component of SH geopotential height variability using daily data (left column) and monthly data (right column). The percentage of the total intraseasonal weather variability explained by each pattern appears in brackets.

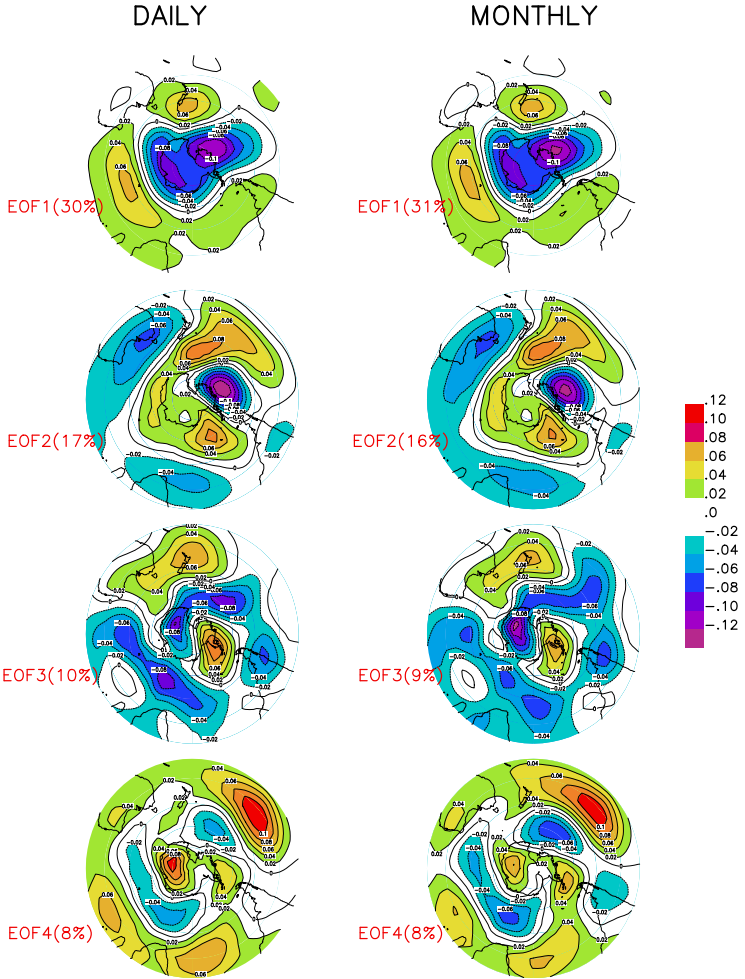


FIGURE 2: The first four dominant JJA patterns of the potentially predictable component of SH geopotential height variability using daily data (left column) and monthly data (right column) The percentage of the total potentially predictable variability explained by each pattern appears in brackets.

and 0.92 for patterns 1–4, respectively. The differences are due to sampling errors (daily versus monthly) and estimation errors (Equation (3) versus Equation (19)). Clearly, these are not large.

4 Conclusions

We have proposed and examined a technique for extracting weather noise patterns of interannual variability from meteorological seasonal mean fields. The technique uses monthly mean, as opposed to the currently used daily, data to estimate the covariance of the weather noise component at two different geographical locations. The method is more computationally efficient and relies on the fact that the major contribution to the seasonal mean of the weather noise component is from weather events at the monthly or longer time scales, that is, intraseasonal phenomena. The dominant intraseasonal patterns of the SH wintertime circulation derived from this method and one using daily data are essentially the same, with minor differences due to sampling and estimation errors. This is also true of the potentially predictable patterns of interannual variability, derived after the removal of the intraseasonal variability.

A Principal component analysis

Here we provide a brief summary of the PC analysis technique used to produce spatial patterns of variation from a given covariance matrix. The interested reader is referred to [3] for more details.

Let $\mathbf{u}_y = \{u_{ry} \mid r = 1, \dots, R\}$ ($y = 1, \dots, Y$) be a multivariate column vector of data at R locations, in a given year y of Y years and with sample covariance matrix \mathbf{V} . That is, it represents R time series of length Y years.

A PC analysis seeks to define another vector $\mathbf{w}_y = \{w_{py} \mid p = 1, \dots, R\}$ ($y = 1, \dots, Y$) with

$$w_{py} = \sum_{r=1}^R c_{pr} u_{ry} = \mathbf{c}_p^T \mathbf{u}_y \quad (20)$$

defined as the p th PC whose sample variance can be shown [3] to be $\mathbf{c}_p^T \mathbf{V} \mathbf{c}_p$. The PCs are chosen in such a way as to maximize their sample variance with the largest variance assigned to the first PC, the second largest variance to the second PC and so forth. If the coefficients \mathbf{c}_p are constrained so that

$$\mathbf{c}_p^T \mathbf{c}_q = \begin{cases} 1, & p = q, \\ 0, & p \neq q, \end{cases} \quad (21)$$

then it can be shown [3] that this can be achieved by solving the eigenproblem

$$\mathbf{V} \mathbf{c} = \lambda \mathbf{c}, \quad (22)$$

with the eigenvalues λ_p giving the sample variances of the PCs and the eigenvectors (\mathbf{c}_p) providing the coefficients in Equation (20). The eigenvectors also provide the spatial patterns of variability and are sometimes referred to as “empirical orthogonal functions”.

References

- [1] E. S. Epstein. On obtaining daily climatological values from monthly means. *J. Climate*, 4:365–368, 1991. [C382](#)
- [2] C. S. Frederiksen, A. P. Kariko and X. Zheng. A technique for extracting potentially predictable patterns from climate data. *ANZIAM J.*, 44(E):C160–C179, 2003. [Online] <http://anziamj.austms.org.au/V44/CTAC2001/Fred>. [C379](#), [C380](#)
- [3] I. T. Joliffe. Principal Component Analysis. *Springer Verlag*, 271pp, 1986. [C380](#), [C388](#), [C389](#)

- [4] R. A. Madden. Estimates of the natural variability of time averaged sea level pressure. *Mon. Wea. Rev.*, 104:942–952, 1976. [C379](#)
- [5] X. Zheng and C. S. Frederiksen. Validating interannual variability in an ensemble of AGCM simulations. *J. Climate*, 12:2386–2396, 1999. [C379](#)
- [6] X. Zheng, H. Nakamura and J. A. Renwick. Potential predictability of seasonal means based on monthly time series of meteorological variable. *J. Climate*, 13:2591–2604, 2000. [C379](#), [C381](#), [C382](#), [C383](#)