

Circulation Regimes and Low-Frequency Oscillations in the South Pacific Sector

Andrew W. Robertson* and Carlos R. Mechoso
Department of Atmospheric Sciences, University of California, Los Angeles,
Los Angeles, California

*Current affiliation: International Research Institute for Climate Prediction,
Palisades, NY.

November 14, 2002

MWR, sub judice

Correspondence address:

Dr. Andrew W. Robertson
International Research Institute for Climate Prediction
PO Box 1000
Palisades, NY 10964
Tel: (845) 680 4491
Fax: (845) 680 4865
E-mail: awr@iri.columbia.edu

Abstract

The characteristics of sub-seasonal circulation variability over the South Pacific are examined using 10-day lowpass filtered 700-hPa geopotential height NCEP/NCAR Reanalysis data. We determine to what extent the variability in each season is characterized by recurrent geographically-fixed circulation regimes and/or oscillatory behavior. Two methods of analysis (a *K*-means cluster analysis and a cross-validated Gaussian mixture model) both indicate 3–4 geographically-fixed circulation regimes in austral fall, winter and (to some extent) spring. The spatial regime structures are found to be quite similar in each season; they resemble the so-called "Pacific-South American" (PSA) patterns discussed in previous studies and often referred to as PSA 1 and PSA 2. Oscillatory behavior is investigated using singular spectrum analysis. This identifies a predominantly stationary wave with a period of about 40 days and a spatial structure similar to PSA 1; it is most pronounced in winter and spring and exhibits a noticeable eastward drift as it decays. The power spectrum of variability is otherwise well approximated by a red spectrum, together with enhanced broader-band 15–30-day variability.

Our results indicate that low-frequency variability over the South Pacific is not dominated by a propagating wave whose quadrature phases are PSA 1 and PSA 2, as hitherto described. Rather, we find that the variability is well described by the occurrence of 3–4 geographically-fixed circulation regimes, with a (near) 40-day oscillation that is predominantly stationary in space. The potential sub-seasonal predictability implied by this duality is discussed. Only during austral spring is a strong correlation found between El Niño and the frequency-of-occurrence of the circulation regimes.

1. Introduction

Two distinct approaches have been used to study the "coarse-grain" structure of atmospheric low-frequency variability ($10 < T < 100$ d): the episodic or intermittent and the oscillatory or periodic. Ghil and Robertson (2002) have reviewed studies of the Northern Hemisphere in these terms. The intermittency approach describes geographically-fixed multiple-flow (or weather) regimes, their persistence and recurrence, and the Markov chain of transitions between them. The periodicity approach studies intraseasonal oscillations and their predictability. Plaut and Vautard (1994) described low-frequency variability in the Northern Hemisphere midlatitudes in terms of oscillatory phenomena as well as in terms of flow regimes defined by the most frequently occurring patterns. Their results reveal both oscillatory and episodic features. For example, they found that a Euro-Atlantic blocking regime is strongly favored by, although not systematically associated with a particular phase of a 30–35-d oscillatory component. This type of relationship has potentially important practical implications due to the higher inherent predictability of oscillatory behavior.

The extratropical circulation of the Southern Hemisphere, being much more zonally-symmetric than that of the Northern Hemisphere, is deceptively simpler. A closer look hints at a higher complexity since the variance spectrum of empirical orthogonal functions (EOFs) is flatter and the leading modes are characterized by higher zonal wavenumbers, thus suggesting a system with a higher number of degrees of freedom (Mo and Ghil 1987). The leading EOFs of low frequency variability constructed over the entire Southern Hemisphere consist of a high-latitude vacillation in the strength of the polar vortex and Rossby wave trains over the South Pacific (Kidson 1988). Mo and Ghil (1987) identified a wavetrain-like pattern arching poleward from the subtropical central Pacific to Argentina, and then refracting equatorward into the Atlantic. Szeredi and Karoly (1987a,b) identified a similar wave pattern, but with a 90° zonal phase lead.

These wave-like "modes" have since been referred to as Pacific-South American (PSA) patterns, numbers 1 and 2 respectively. PSA 1 and 2 are reproduced well by the two leading EOFs of low-frequency variability over a domain restricted to the South Pacific extratropics (see Fig. 1, below). The nomenclature is motivated by analogy with the Pacific-North American pattern in the Northern Hemisphere (e.g., Wallace and Gutzler 1981).

The spatial phase-quadrature relationship between PSA 1 and 2, together with their near-degeneracy in EOF analyses, suggests a propagating wave. In an analysis of 200-hPa streamfunction, Mo and Higgins (1998) have reported evidence during austral winter from lag correlations and singular spectrum analysis (SSA) that PSA 1 and 2 represent an eastward propagating wave with a period of about 40 d. They also argued that forcing over the western tropical Pacific can create a favorable situation for a particular phase of the PSA modes to strengthen. Low-frequency variability in the Southern Hemisphere is also episodic in nature (Mo 1986). Mo and Ghil (1987) studied quasi-stationary events in the Southern Hemisphere using daily 500-hPa geopotential maps during 1972–1983 and found two types of geographically-fixed persistent anomalies, both with a strong zonal wavenumber 3 component and strongly resembling the two leading EOFs in the data set.

A non-dispersive propagating wave has no preferred phase so that is unclear how the PSA patterns can simultaneously be described as quadrature phases of a propagating wave, and geographically fixed circulation regimes. The aim of the present paper is to explore systematically the episodic/intermittent and oscillatory/periodic characteristics of the PSA modes, and to extend the work of Mo and co-workers to all four seasons. We examine the evidence for circulation regimes as well as low-frequency oscillations over the South Pacific sector from the NCEP/NCAR reanalysis data, and investigate their seasonal dependence. We apply three analysis methods to the 700 hPa geopotential

height dataset: a K -means clustering and Gaussian mixture model, as well as multi-channel SSA (MSSA). We then assess any relationships between the circulation regimes and oscillatory behavior that could lead to predictability of the former. All three methods use conventional EOF analysis only as a means of data reduction, thereby circumventing the potential problem of near-degeneracy of PSA 1 and 2 when defined in terms of the leading EOFs, which may complicate the physical interpretation of either one.

The paper is organized as follows. In section 2 we describe the data set and pre-processing steps. Section 3 contains the regime analysis for each of the four conventional 3-month meteorological seasons. In section 4 we use MSSA to isolate oscillatory behavior, and discuss its relationships with the clusters in section 5. This is followed by the conclusions in section 6.

2. Data and preprocessing

All analyses in this paper are restricted to the South Pacific sector 150°E – 60°W , 20° – 90°S . We use daily 700-hPa geopotential heights from the NCEP-NCAR Reanalysis data set 1948–1999 (Kalnay et al. 1996). The lower-tropospheric geopotential is chosen so as to emphasize the midlatitude circulation where 700 hPa corresponds to an approximate “steering level” (Blackmon et al. 1994). Data is very sparse in much of the domain of study, and a near-surface variable should be more closely controlled by surface observations which predominate. The reanalysis fields are on a 2.5° latitude-longitude grid, which was sub-sampled by omitting every other point to yield a $5^{\circ}\times 5^{\circ}$ resolution. The values were preprocessed by low-pass filtering (Blackmon and Lau 1980) at 10 days, followed by construction of the leading EOFs using the covariance matrix, unweighted by area. There is an error in the reanalysis dataset between 1979 and 1992 due to bogus Australian surface pressure data, but its impact on low-pass filtered data is expected to be minimal (Mo and Higgins 1998).

For the regime analysis (Sect. 3) we determine EOFs separately for the December–February (summer, DJF), March–May (fall, MAM), June–August (winter, JJA), and September–November (spring, SON) seasons during the 1948–1999 interval, which provides the maximum sample size. For the oscillatory analysis (Sect. 4), we constructed EOFs for the *entire* calendar year during the period 1958–1999, but (1) with the mean seasonal cycle subtracted on a daily basis (after the low-pass filtering), and (2) with the interannual variability suppressed by forming (deseasonalized) anomalies from annual means.

The two leading EOFs from the latter year-round filtered data are shown in Fig. 1 and compare closely with the PSA patterns constructed by Mo and Higgins (1998) from 500-hPa geopotentials for June–August. They explain comparable fractions of (low-pass filtered) variance (21.7% and 19.2%), and are well separated from the higher-ranked EOFs (EOF 3 explains 12.5%). The pairing between the first two empirical modes is less clear in the individual seasons (in which interannual variability was also retained). When interannual variability is retained PSA 1 exhibits a zonally-symmetric component over the pole, especially in summer. The leading 10 EOFs account for more than 85% of the (low-pass filtered) variance in all analyses.

3. Regime analysis

(i) Number of regimes

To analyze low frequency variability from the episodic point of view, we use primarily *K*-means clustering (MacQueen 1967). This is a straightforward and widely used partitioning method that classifies all days into a predefined set of *K* clusters, such as to minimize the spread within them. The number of clusters must be specified, but the sensitivity of the resulting centroids to the choice of initial seeds and the data sub-sample can be used as ad-hoc criteria for assessing the validity of the partitioning into distinct

clusters (Michelangeli et al. 1995). We performed the cluster analysis in the subspace of the leading 10 EOFs for each season in turn, as described in Appendix A.

To be most conservative, both robustness criteria are applied together. The K -means method then indicates multiple regimes in fall ($K=3$ or 5), spring ($K=4$) and marginally in winter ($K=3$); see Table A1. The sensitivity to initial seeds is considerably higher in summer, suggesting that the regime description is less valid then.

In order to obtain an independent measure of the extent to which the geopotential height data support the existence of multiple stationary flow regimes, we have also fitted a Gaussian mixture model to the probability density function (PDF) in the EOF subspace. This method seeks to fit the PDF with a small number of Gaussian components and enables a rigorous test for the existence of multimodality, by estimating the cross-validated likelihood of a single Gaussian versus a that of a mixture of several. The methodology follows the work of Smyth et al. (1999), and the details are given in Appendix B.

In fall and winter, the mixture model indicates multimodality with 3–4 Gaussians (Table B1), supporting the results of the K -means method. Figure 2 shows the data scatter in the subspace of EOFs 1 and 2 for fall, together with the locations of the three Gaussians. The positions of the three regime centroids given by the mixture model (circles) agree closely with those of the cluster analysis (squares). In summer the mixture model indicates unimodality, which is consistent with the much higher summertime sensitivity to initial seeds in the K -means analysis. In spring, the K -means analysis selects $K=4$ while the mixture model indicates unimodality. We will return to this apparent contradiction in Sect. 6.

Based on the above results, the regime description of low-frequency variability over the South Pacific is well justified in fall and winter, somewhat less certain in spring,

and probably not applicable in summer. For parsimony, in what follows we will consider the case of $K=3$ for fall, winter and spring, basing our analysis on the K -means results. Although the spring values in Table A1 suggest that $K=4$ is more appropriate, three of the cluster centroids obtained in spring are almost indistinguishable from those in winter when $K=3$ (the pattern correlation between the respective centroids exceeds 0.97), while the 4th centroid is less robust between the seasons.

(ii) Spatial structures and regime transitions

Figure 3 shows hemispheric geopotential height anomaly composites for each of the three clusters for fall, winter and spring. In all three seasons the regimes exhibit PSA-like patterns, characterized by meridionally elongated tripole Rossby wave patterns (Berbery et al. 1992). Regimes 1 and 3 resemble opposite polarities of EOF 1 although they are slightly phase-shifted, while regime 2 resembles EOF 2 (Fig. 1). The cluster analysis does not, however, simply select each polarity of the EOFs, as is clear from the off-axis positions of the centroids in Fig. 2 which represent a linear combination of EOF 1 and 2. Thus, EOF 1 can be viewed as the single spatial pattern that maximizes the variance contained in regimes 1 and 3.

For regime durations beyond a few days, the cumulative frequency distribution of residence times in each regime follows approximately a geometric distribution (not shown), similar to the findings of Dole and Gordon (1983) and Kimoto and Ghil (1993) in the Northern Hemisphere. Thus, the duration of events can be approximated by a first-order Markov chain. The composites in Fig. 3 are ordered in terms of their most frequently occurring transitions, so that the "circuit" $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 1$ is the preferred temporal ordering. This progression of spatial patterns suggests an eastward propagation. Simple counts of the number of times this eastward-propagating circuit occurs, compared to the opposite westward-propagating one ($1 \Rightarrow 3 \Rightarrow 2 \Rightarrow 1$) are tabulated in Table 1. In

each season the eastward-propagating circuit is more frequent than the reverse, but not dramatically so. The third row of the Table 1 gives the degree of asymmetry (a) in the transition matrices, where a value of $a=2$ means that transitions are twice as likely to have a preferred direction in time, while $a=1$ denotes equal likelihood of direction. The tendency toward propagation is strongest in spring with a rough estimate of the period of 30 days given by summing the average regime-durations. This apparent propagation motivates the analysis in Sect. 4 where we take an oscillatory approach to analyzing low-frequency variability over the South Pacific sector.

The results in this section were checked a posteriori for any severe long term trends and large seasonal variation in the frequency of regime occurrence. No large long-terms trends were found. The largest within-season variations in regime membership (averaged across the 51 years in the dataset) have a magnitude of about a factor of two, despite the rather similar regime spatial structures seen in different seasons.

4. Low-frequency oscillations

In this section we examine the oscillatory components of the 700-hPa geopotential height over the South Pacific, considering the whole calendar year but with the seasonal cycle subtracted. We begin by computing the year-round sub-annual EOFs using low-pass filtered data (Fig. 1), as described in section 2. Next, 5-day averages of *unfiltered* data, with the mean seasonal cycle subtracted are projected onto these EOFs to give pseudo principal component time series (referred to simply as the PCs in the following). Thus, in effect, we use the leading EOFs of low-frequency variability as "spatial filters" to focus on the low-frequency structures while retaining the full temporal spectrum. Finally, singular spectrum analysis (SSA and M-SSA) is applied to these time series, using the UCLA SSA-MTM Toolkit (Dettinger et al. 1995, Ghil et al. 2002).

We begin with a univariate analysis, and apply SSA to PCs 1 and 2 (i.e. the time series of the PSA 1 and 2 patterns plotted in Fig. 1) in turn. For an independent spectral estimate we also compute the multi-taper method (MTM) spectra. Figure 4 shows the SSA and MTM spectra for PC 1, with statistical significance estimated against the null-hypothesis of red noise, using the tests of Allen and Smith (1996) for SSA and Mann and Lees (1996) for MTM. Both spectra have a long-term “trend” component. This is probably due to interannual variability associated with ENSO’s influence on PSA 1 (Karoly 1989, Cazes et al. 2002). The MTM spectrum exhibits sub-seasonal peaks at periods of about 45, 36, 30, 20 and 15 d (all significant at the 95% confidence level). The SSA spectrum suggests oscillatory pairs of eigenvalues at 45, 31, 19 and 15 d, with the 45d and 31d components being the most statistically significant though at a somewhat lower level than those identified by MTM. The spectrum of PC 2 (Fig. 5) shows very pronounced oscillatory components at 48, 22 and 15 d in both spectral estimates. Mo and Higgins (1998) applied SSA to PSA time series derived from 200-hPa streamfunction and reported common periods in PSA 1 and PSA 2 of 36–40 d, 22–25 d and 16–18 d.

While the spectral peaks in PCs 1 and 2 have comparable periods, the respective spectra do not exhibit strongly commensurate peaks that would indicate a propagating wave, given the approximate spatial quadrature seen in EOF 1 and 2. Indeed, the maximum lag correlation between PCs 1 and 2 is only 0.14, and the spectra in Figs. 4 and 5 indicate a red background with rather modest oscillatory components superposed.

To avoid the potential problem of near degeneracy of PCs 1 and 2 and to examine more closely the spatio-temporal structure of the data, we next apply *multi-channel* SSA to the leading 6 PC time series, thereby using the spatial EOF analysis only as a data reduction tool. To focus on the intraseasonal band, variability with time scales greater than 65 d was removed at the outset using the SSA “detrending” procedure employed by

Robertson (1996), applied to each PC time series in turn. Similar results were obtained using only PCs 1 and 2.

The multichannel analysis identifies the 45–50d component present in both the univariate analyses of EOF 1 and 2 with a period closer to 42 d, while the behavior between 20 and 30 d appears to be more broad-band. Figure 6 shows the eigenvalue spectrum and the power spectra of the leading 8 temporal PCs identified by multichannel SSA. The leading two eigenelements form an oscillatory pair with a period of about 42 d. The next four eigenvalues are clustered together and are associated with a broad-band peak in the temporal PCs with periods of 20–30 d; these are followed by an oscillatory 18-d pair.

Each eigenelement of the MSSA is associated with an evolving spatio-temporal structure, such that the sum of all reconstructs the original data set. The reconstructed contribution of the 42-d pair is illustrated in Fig. 7. The time series of channels 1 and 2 (i.e. PSA 1 and 2) are plotted in Fig. 7a over the 1997–2000 interval. The PSA 1 and 2 components of the wave do appear in phase quadrature, with PSA 2 leading PSA 1, indicative of eastward phase propagation also hinted at in the regime analysis of Sect. 3 (cf. also EOFs 1 and 2 in Fig. 1). However, the PSA 1 component has considerably higher amplitude than that of PSA 2.

Figure 7b shows the spatial structure of the oscillation through a composite half-cycle formed by compositing the reconstruction of the 42-d wave in phase intervals of $1/8^{\text{th}}$ period (i.e. about every 5 days), using the technique of Plaut and Vautard (1994). Here we have plotted the composites of the wave in the PC 1–2 subspace, so that the patterns are linear combinations of PSA 1 and 2 plotted in Fig. 1. To aid interpretation of the longitudinal phase progression with time through the cycle, Fig. 7c shows meridional averages over the band 50–60°S, together with the longitudes of the maxima of EOFs 1 and 2. The temporal evolution can be interpreted as a very gradual eastward propagation,

with rapid intensification of amplitude into category 1, where the oscillation resembles EOF 1 (i.e. PSA 1). This pattern persists while drifting very slowly eastward into category 2, before decaying rapidly into an EOF-2-like pattern (category 3). Thus, these phase composites indicate a predominantly stationary oscillation, with the peak amplitude corresponding to PSA 1 and the minimum amplitude corresponding to PSA 2. The cyclic evolution of the wave plotted in Figs. 7b and c suggests a highly dispersive wave, with an eastward group velocity much exceeding the phase speed. When the phase composites are plotted in the PC 1–6 subspace (in which the MSSA was computed), the peak phase changes little while the quadrature phase is no longer recognizable as PSA 2.

A slow modulation in the amplitude of the 42-d component is visible in Fig. 7a, but there is no detectable relationship with ENSO. To determine whether there is any marked seasonal variation, Fig. 8 shows the variance of the PSA 1 and 2 components of the 42-d wave for each season. PSA 1 clearly dominates the variance in all seasons, with largest values in winter-spring and a minimum in summer. Thus, low-frequency variability during summer appears less coherent, and is neither well characterized by oscillatory behavior nor regimes.

5. Relationships between regimes and LFOs

The oscillatory components identified in the previous section account for only a very small fraction of the variance, with the 42-d mode accounting for about 5% of the sub-annual variance of 5-d means over the South Pacific sector. In this section we explore whether or not this weak oscillation may nonetheless be related to the occurrence of the three circulation regimes constructed in Sect. 3. To do this, regime occurrence was simply counted for each of the eight phase categories of the 42-d oscillation taking the fall, winter and spring seasons in turn. Confidence limits for by-chance occurrence were computed by permutating the order of the regime-occurrence time series 100 times.

There is a highly statistically significant relationship between the 42-d oscillation and regimes 1 and 3 during winter and spring, as shown in Fig. 9 for regime 1; the frequency of occurrence of regime 3 is generally the inverse of regime 1 (Table 2). This comes about because regimes 1 and 3 resemble opposite polarities of PSA 1, which is also the spatial pattern that dominates the oscillation (Fig. 7b). Regime 2 also shows a significant relationship during several phases of the oscillation, though the changes in its frequency-of-occurrence are more moderate, consistent with weakness of the quadrature phase of the oscillation (Fig. 7b,c). There is a general regime progression of $1 \rightarrow 2 \rightarrow 3$ through the cycle of the 42-d oscillation, which is consistent with Table 1.

The relationship between the 42-d oscillation and the frequency-of-occurrence of the three regimes suggests that the latter may contain some predictability. To explore this further, we compute the probability of regime occurrence conditional on the lagged phase category of the oscillation, following Plaut and Vautard (1994). Thus, assuming that the phase category of the oscillation is known at some initial time, we aim to predict the regime occurrence at later times. Figure 10 shows the probability of regime 1 occurring during winter at lead times up to 50 days, conditional on phase category of the 42-d oscillation at the initial time. The results are shown for initial phase categories 1–4, with similar results obtained for the second half cycle. The conditional probabilities clearly exceed chance up to 30 days into the future. Similar results are found for regime 3, as well as for spring and fall. Regime 2 is found to be less predictable. Similar results are found with $K = 2$ regimes specified, in which case we recover EOF 1 (i.e. PSA 1).

The apparent predictability seen in Fig. 10 cannot, however, be interpreted as hindcast skill because the probabilities have been computed from the same data used to determine the regimes and the oscillation.

6. Discussion and conclusions

We have analyzed low frequency variability in the NCEP/NCAR Reanalysis data of lower troposphere (700 hPa) geopotential height field over the South Pacific sector in terms of geographically-fixed circulation regimes and oscillatory behavior. The regimes were identified using a K -means cluster analysis and a cross-validated Gaussian mixture model, while a multi-channel singular spectrum analysis (MSSA) was used to the search for oscillatory components.

The spatial structures identified in both types of analysis are similar to the Pacific-South American (PSA) wavetrains identified in previous studies. The two leading ($T > 10$ d) EOFs are usually referred to as PSA 1 and PSA 2 (Fig. 1). In relaxing the orthogonality constraints inherent in EOF analysis, our results confirm the physical relevance of EOF 1, since a similar pattern dominates our oscillatory analysis in terms of amplitude (Fig. 7b,c). Our analysis of circulation regimes (Fig. 3) also indicates that the leading two EOFs give reasonable approximations of spatial structure, although the regimes are linear combinations of them (Fig. 2).

Both methods of regime analysis show that low-frequency variability over the South Pacific sector is well described by three or four recurrent geographically-fixed circulation regimes in austral fall (MAM) and winter (JJA), and to some extent in spring. Both methods suggest that the regime description is less valid in summer. The spatial structures of the regimes (Fig. 3) are found to be very similar in fall, winter and spring, consisting of zonal wave trains across the South Pacific. These patterns are confined to midlatitudes and differ from the PSA wavetrains detected by Mo and Higgins (1998) in 200-hPa streamfunction that extend into the tropics. In further contrast to the latter study, we found no statistically significant relationships with tropical OLR, and thus conclude that the regimes identified here are intrinsic to the midlatitudes, similar to the

interpretation of Lau et al. (1994). On the other hand, the frequency-of-occurrence of regime 3 is highly correlated with ENSO during austral spring ($r=0.60$ with the Niño 3.4 index). This is consistent with the study of Cazes et al. (2002) which proposed that interannual ENSO teleconnections over southeastern South America could be interpreted in terms of the changes in the frequency of occurrence of intraseasonal circulation regimes during October–December.

The spectral analysis of PSA 1 and 2 reveals a predominantly red spectrum, which is consistent with episodic regime-like behavior where regime durations follow approximately a geometric distribution without a preferred duration. However, there is evidence of significant oscillatory peaks in the 40–50-d, and 20–30-d bands. The spectral peaks in PSA 1 and 2 do not match closely, suggesting no simple propagating wave. A multichannel analysis in the sub-seasonal range identifies a dominant peak at 42.5 d, which is slightly longer period than the 36–40 d peak found by Mo and Higgins (1998) in 200-hPa streamfunction. Phase composites of this component show that it is dominated by the PSA 1 spatial pattern, and is almost stationary in phase. A gradual eastward drift of this pattern accompanies its rapid attenuation, so that the quadrature phase is very weak in amplitude. The oscillation is present throughout the year but is most pronounced in austral winter and spring. No ENSO modulation was found.

Previous studies (e.g. Mo and Higgins 1998) have interpreted low-frequency variability over the South Pacific in terms of a propagating wave with a period of about 35–40 days, with an eastward progression characterized by the PSA 1 and PSA 2 patterns. Such a description would not be compatible with the geographically-fixed circulation regimes. Our analysis does find strong evidence for the latter in fall and winter. We also find evidence of an oscillatory component with a period of about 42 days. These two findings are consistent because the oscillation is found to be a predominantly stationary in space. Thus, we find that both the episodic and the

oscillatory viewpoints are consistent with each other, with the regimes characterizing the slow part of the cycle.

While we have stressed the geographically fixed nature of the PSA patterns, we do find some evidence of eastward propagation in austral spring. This is brought out by the transitions between the three regimes identified in the K -means analysis (Table 1). It is also hinted at by the increased robustness of the $K=4$ description in which the four regimes have spatial structures that are close to the PSA 1 and 2. The Gaussian mixture model indicates unimodality in spring, arguing against the regime description during spring. It is beyond the scope of this paper to stratify the oscillatory analysis by season, but it would be of interest to know if the 42-d wave has a stronger propagating component during austral spring.

The 42-day wave is weak only explaining about 5% of low-frequency variance so that its relevance is questionable. We investigated whether or not a weak oscillatory component could nonetheless influence regime transitions. According to Fig. 9 there is quite a strong (and highly statistically significant) relationship between the phase of the 42-d wave and regime occurrence during winter and spring. The frequency of occurrence of regime 1 (similar to PSA 1) changes by a factor of three between the extreme phases of the oscillation, suggesting that the oscillatory component is stronger than its variance indicates. One explanation for this apparent discrepancy would be that the oscillation is somewhat broader band than indicated by the MSSA. The finding that low-frequency variability over the South Pacific is characterized by both (a) geographically-fixed PSA-like circulation regimes, and (b) by oscillatory components has implications for potential predictability. We find that bias in the frequency-of-occurrence of each regime may be strong enough for the oscillation to be used as a predictor of the probability of regime occurrence, up to 30 days in advance in certain cases (Fig. 10), although further work is required to determine whether there is any useful skill using cross-validation. The

predictive nature of the oscillatory component found here is similar in extent to that reported for the North Atlantic-European sector by Plaut and Vautard (1994).

Acknowledgements: We would like to thank and H. Berbery, G. Cazes and K. Mo for helpful discussions, and the two reviewers whose comments lead to a substantial improvement of the manuscript. Many of the computations were performed by G. Kerneis and A. Loireau, two summer student visitors from the French Ecole Polytechnique. This work was supported by NOAA grants NA06GPO427, and NA06GPO511. The NCEP/NCAR Reanalysis data were provided through the NOAA Climate Diagnostics Center (<http://www.cdc.noaa.gov/>).

Appendix A: Cluster analysis

The K -means method is applied in the subspace of the leading 10 PCs following Michelangeli et al. (1995). An initial 10% random subset of days is used to determine the initial seeds, and the algorithm proceeds iteratively from the initial seeds, modifying the cluster centroids (i.e. the *means*) at each iteration. The clustering is then repeated 50 times to eliminate any sensitivity to initial seeds. The *reference partition* is defined from this set of 50 analyses to be the one whose cluster centroids are most similar to the remaining 49, in terms of pattern correlation. The sensitivity to the choice of initial seeds gives a measure of how classifiable the data set is for a particular pre-specified number of clusters K . The similarity between two partitions P_i and P_j can be quantified by the *smallest* pattern correlation between a centroid in P_i with its best analogue in P_j . A *classifiability index (CI)* can then be defined as the average of this similarity-value over

all pairs of partitions (Michelangeli et al. 1995). The CI is unity for a perfect match and zero for uncorrelated patterns.

The number of clusters K should also maximize the reproducibility of the patterns obtained from subsets of the data (Cheng and Wallace 1993). To quantify reproducibility, random subsets containing 50% of the days in the dataset are drawn 100 times. Splitting a data set into two equal halves is a common device for assessing robustness. The reference partition is computed for each 50%-subset, and its similarity with that of the full dataset calculated; averaging these similarity values from all 100 subsets defines a *reproducibility index* (R) for each value of K . Cheng and Wallace (1993) argue, on the basis of experience, that two *hemispheric* patterns bear a strong resemblance to each other if their pattern correlation is near or above 0.89. They increase this threshold for a sector of the hemisphere to account for the reduced number of spatial degrees of freedom, so as to retain a similar value of the Student t -statistic. For the 150° South Pacific sector considered here, the corresponding threshold would be 0.94.

Table A1 shows the classifiability and reproducibility indices (in percent) as a function of K for each 3-month season. Values of $K=3-4$ in Table A1 generally yield the best CI and R scores, although the highest values of these indices do not always clearly point to a particular value of K . The case of $K=2$ yields almost exactly EOF 1 in all seasons so that the K -means method does not yield any additional information to classical EOF analysis (Michelangeli et al. 1995).

Appendix B: The Gaussian mixture model

The method of cross-validated maximum likelihood is used to determine the number of component Gaussian distributions that provide the best fit to the data. The cross-validation here consists of randomly selecting 25 seasons, training the model on these seasons, and then validating on the remaining 25. The procedure is repeated 20

times. This method provides a rigorous test of any multimodality in the PDF, against the null-hypothesis of a single unimodal Gaussian.

The mixture model was applied to the same seasonally-stratified data sets used for the K -means analysis, restricting the data to the subspace of the leading two PCs. The results are shown in Table B1 as a function of the number of Gaussian components, k . The cross-validated log-likelihoods are relative measures of likelihood, with the maximum (i.e. smallest negative) value being the most likely. The estimated (posterior) probabilities of each value of k , given the dataset (i.e. $P(k|D)$), are also tabulated (see Smyth et al. 1999). The high posterior probability of $k=3$ in fall is consistent with the cluster analysis in which there is a unique coincidence of both CI and R values greater than 0.98 (Table A1) with three clusters during the fall.

References

- Allen, M. R., and L. A. Smith, 1996: Monte Carlo SSA: detecting irregular oscillations in the presence of coloured noise. *J. Climate*, **9**, 3373-3404.
- Berberly, E. H., J. Nogués-Paegle, and J. D. Horel, 1992: Wavelike Southern Hemisphere extratropical teleconnections. *J. Atmos. Sci.*, **49**, 155-177; Corrigendum, **49**, 2347.
- Blackmon, M. L., and N.-C. Lau, 1980: Regional characteristics of the Northern Hemisphere wintertime circulation: A comparison of the simulation of a GFDL general circulation model with observations. *J. Atmos. Sci.*, **37**, 497-514.
- Blackmon, M. L., N.-C. Lau, and H.-H. Hsu, 1984: Time variation of 500 mb height fluctuations with long, intermediate and short time scales. *J. Atmos. Sci.*, **41**, 981-991.
- Cazes, G., A. W. Robertson, and C. R. Mechoso, 2002: Seasonal Dependence of Teleconnections Over South America. *J. Climate*, in press.
- Cheng, X., and J. M. Wallace, 1993: Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns. *J. Atmos. Sci.*, **50**, 2674-2696.
- Dettinger, M. D., C. M. Strong, W. Weibel, M. Ghil, and P. Yiou, 1995: Software for singular spectrum analysis of noisy time series. *Eos Trans. AGU*, **76**(2), pp. 12,14,21.
- Dole, R. M., and N. D. Gordon, 1983: Persistent Anomalies of the Extratropical Northern Hemisphere Wintertime Circulation: Geographical Distribution and Regional Persistence Characteristics. *Mon. Wea. Rev.*, **111**, 1567-1586.
- Ghil, M., and K. Mo, 1991: Intraseasonal oscillations in the global atmosphere. Part II: Southern hemisphere. *J. Atmos. Sci.*, **48**, 780-790.

- Ghil, M., and A. W. Robertson, 2002: "Waves" vs. "particles" in the atmosphere's phase space: A pathway to long-range forecasting?, *Proc. Natl. Acad. Sci.*, **99** (Suppl. 1), 2493-2500
- Ghil M., R. M. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou, 2002: Advanced spectral methods for climatic time series, *Rev. Geophys.*, in press.
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph, 1996: The NCEP/NCAR 40-year Reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-470.
- Karoly, D. J., 1989: Southern Hemisphere circulation features associated with El Niño-Southern Oscillation events. *J. Climate*, **2**, 1239-1251.
- Kidson, J. W., 1988: Interannual variations in the Southern Hemisphere circulation. *J. Climate*, **1**, 1177-1198.
- Kimoto, M., and M. Ghil, 1993: Multiple flow regimes in the Northern Hemisphere winter. Part I: Methodology and hemispheric regimes. *J. Atmos. Sci.*, **50**, 2625-2643.
- MacQueen, J. (1967): Some methods for classification and analysis of multivariate observations. *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, pp. 281-297.
- Mann, M. E., and J. M. Lees, 1996: Robust estimation of background noise and signal detection in climatic time series. *Climatic Change*, **33**, 409-445.

- Michelangeli, P. A., R. Vautard, and B. Legras, 1995: Weather regimes: Recurrence and quasi-stationarity. *J. Atmos. Sci.*, **52**, 1237-1256.
- Mo, K. C., 1986: Quasi-Stationary States in the Southern Hemisphere. *Mon. Wea. Rev.*, **114**, 808–823.
- Mo, K. C., and M. Ghil, 1987: Statistics and dynamics of persistent anomalies. *J. Atmos. Sci.*, **44**, 877-901.
- Mo, K., and R. W. Higgins, 1998: The Pacific South American modes and tropical convection during the Southern Hemisphere winter. *Mon. Wea. Rev.*, **126**, 1581-1598.
- Lau, K. -M., P. -J. Sheu, I. -S. Kang, 1994: Multiscale Low-Frequency Circulation Modes in the Global Atmosphere. *J. Atmos. Sci.*, **51**, 1169-1193.
- Plaut, G., and R. Vautard, 1994: Spells of low-frequency oscillations and weather regimes in the northern hemisphere. *J. Atmos. Sci.*, **51**, 210-236.
- Robertson, A.W., 1996: Interdecadal variability in a multicentury climate integration. *Climate Dyn.*, **12**, 227-241.
- Smyth, P. J., K. Ide, and M. Ghil, 1999: Multiple regimes in Northern Hemisphere height fields via mixture model clustering. *J. Atmos. Sci.*, **56**, 3704-3723.
- Szeredi, I., and D. Karoly, 1987a: The vertical structure of monthly fluctuations of the Southern Hemisphere troposphere. *Aust. Meteor. Mag.*, **35**, 19-30.
- Szeredi, I., and D. Karoly, 1987b: The horizontal structure of monthly fluctuations of the Southern Hemisphere troposphere from station data. *Aust. Meteor. Mag.*, **35**, 119-129.
- Vautard, R., P. Yiou, and M. Ghil, 1992: Singular spectrum analysis: A toolkit for short, noisy, chaotic signals. *Physica D.*, **58**, 95-126.

Wallace, J. M., D. S. Gutzler, 1981: Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter. *Mon. Wea. Rev.*, **109**, 784–812.

Figure Captions

Figure 1: Leading two EOFs of 10-d low-pass filtered 700 hPa geopotential heights 1958–99, computed over the South Pacific sector with the mean seasonal cycle subtracted. Shown are hemispheric maps of regression coefficients with the respective principal component time series (PC). Amplitudes correspond to a one standard deviation of the PC, with each PC’s variance given (in percent). Contour interval: 10 gpm.

Figure 2: Scatter of daily geopotential heights (March–May) projected onto the PC-1/PC-2 plane, with every 5th point plotted. The solid squares denote the positions of the three *K*-means. The results of the Gaussian mixture model are plotted in terms of its centroids (solid circles) and covariance ellipses.

Figure 3: Regime composites of 700-hPa geopotential height obtained from the *K*-means analysis of the South Pacific sector for each season 1948–99. Contour interval: 10 gpm. Negative anomalies are shaded.

Figure 4: Spectral analysis of pseudo-PC 1. (a) SSA spectrum computed with $M=40$ 5-d means. The error bars give the 95% confidence interval of a red-noise process fitted to the time series. (b) MTM spectrum computed with 39 tapers. The background curves denote the 50, 95 & 99% thresholds of a red-noise null-hypothesis. Periodicities of interest are given above each spectrum (in days). The SSA used a window of $M=40$ 5-d means (an approximate spectral resolution of 0.0040 cy d^{-1}). For MTM, 39 tapers were used yielding a half-bandwidth spectral resolution of 0.0013 cy d^{-1} .

Figure 5: Spectral analysis of pseudo-PC 2. Details as in Fig. 4.

Figure 6: Multichannel SSA of detrended pseudo-PCs 1–6, using $M=20$. (a) Eigenvalue rank spectrum, in percent of total variance. (b) MEM spectra (5 poles) of the leading 8 temporal PCs resulting from the MSSA. The approximate periods are given above the curves. The MSSA was applied using the full covariance matrix (Plaut and Vautard 1994).

Figure 7: The sum of MSSA reconstructed components (RCs) 1 and 2. (a) Channel 1 (solid curve) and channel 2 (dotted curve) plotted over the 3-year interval 1997–1999. (b) Phase composites over the phase intervals (1) $0-\pi/4$, (2) $\pi/4-\pi/2$, (3) $\pi/2-3\pi/4$, and (8) $7\pi/4-2\pi$, with negative anomalies shaded. (c) Meridional averages $50-60^\circ\text{S}$ of the maps in panel b. Also shown (arrows) are the longitudes of the maxima of EOFs 1 and 2 (Fig. 1). The units are arbitrary. Panels b and c show the projection of the oscillation in the 2-D subspace of PCs 1 and 2 (Fig. 1).

Figure 8: The seasonal variation of the amplitude of the 42-d wave reconstructed components (RCs 1–2), in terms of the variance of channels (PCs) 1 (grey) and 2 (black).

Figure 9: Occurrence frequency of regime 1 for each phase category of the 42-d wave reconstructed components (RCs 1–2). The error bars show the 95% range of random sampling.

Figure 10: Conditional probability of occurrence of regime 1 during June–August, taking the phase category of the 42-d wave reconstructed components (RCs 1–2) as a predictor, as a function of lead time. Error bars give the 95% range of random sampling.

Table Captions

Table 1: Number of transitions that belongs to each temporal "circuit", and the average asymmetry in the transition matrix, defined by the arithmetic average of the ratios of the off-diagonal elements.

Table 2: The number of days that fall simultaneously into the three clusters (columns) and eight 42-d oscillation $\pi/4$ phase categories (rows). The clusters were determined separately for each season. Values smaller than the 2.5 %-ile of random sampling are given in italics, with those that exceed the 97.5 %-ile in bold face.

Table A1: Classifiability (CI) and reproducibility (R) indices for the K -means analysis as a function of K for each season, both in percent. The reproducibility is given as a range over all the K centroids in the partition. Values of 94% or above are significant, according to an ad-hoc threshold.

Table B1: Cross-validated log-likelihood (top) and estimated posterior probability (bottom) of the Gaussian mixture model as a function of k . The most likely values are highlighted in bold.

Table 1:

	MAM	JJA	SON
$W \rightarrow E$ $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 1$	275	266	315
$E \rightarrow W$ $1 \Rightarrow 3 \Rightarrow 2 \Rightarrow 1$	218	184	185
Transition matrix asymmetry, (a)	1.29	1.45	1.70

Table 2:

	MAM			JJA			SON		
	1	2	3	1	2	3	1	2	3
1	135	126	169	115	115	225	81	149	180
2	168	130	132	134	105	156	116	114	190
3	183	153	104	201	112	137	213	82	140
4	223	155	67	273	96	86	301	64	85
5	183	177	80	276	100	59	273	95	82
6	162	172	96	247	112	86	257	124	84
7	163	155	117	148	166	111	114	145	136
8	111	121	203	92	162	171	87	119	224

Table A1

	Summer (DJF)		Fall (MAM)		Winter (JJA)		Spring (SON)	
	<i>K</i>	<i>CI</i>	<i>R</i>	<i>CI</i>	<i>R</i>	<i>CI</i>	<i>R</i>	<i>CI</i>
3	82	94–99	98	98–99	92	95–97	81	91–95
4	80	97–99	80	88–95	90	98–99	96	98–99
5	80	98–99	97	97–99	82	91–98	93	93–99
6	75	88–98	66	88–98	84	81–99	75	82–97
7	70	95–97	63	80–92	84	90–99	71	88–97
8	64	87–93	65	78–92	68	82–94	76	67–96

Table B1:

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
Summer (DJF)	-1395.327 0.983	-1395.490 0.017	-1395.704 0.000	-1395.911 0.000	-1396.365 0.000	-1396.389 0.000
Fall (MAM)	-1457.226 0.000	-1456.876 0.040	-1456.749 0.951	-1456.933 0.010	-1457.155 0.000	-1457.273 0.000
Winter (JJA)	-1485.122 0.000	-1484.708 0.000	-1484.225 0.293	-1484.192 0.662	-1484.464 0.001	-1484.300 0.045
Spring (SON)	-1443.410 0.999	-1443.698 0.001	-1443.779 0.000	-1443.913 0.000	-1444.147 0.000	-1444.330 0.000

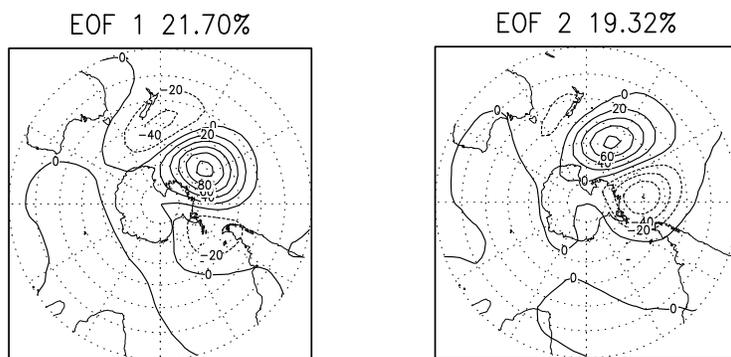


Figure 1

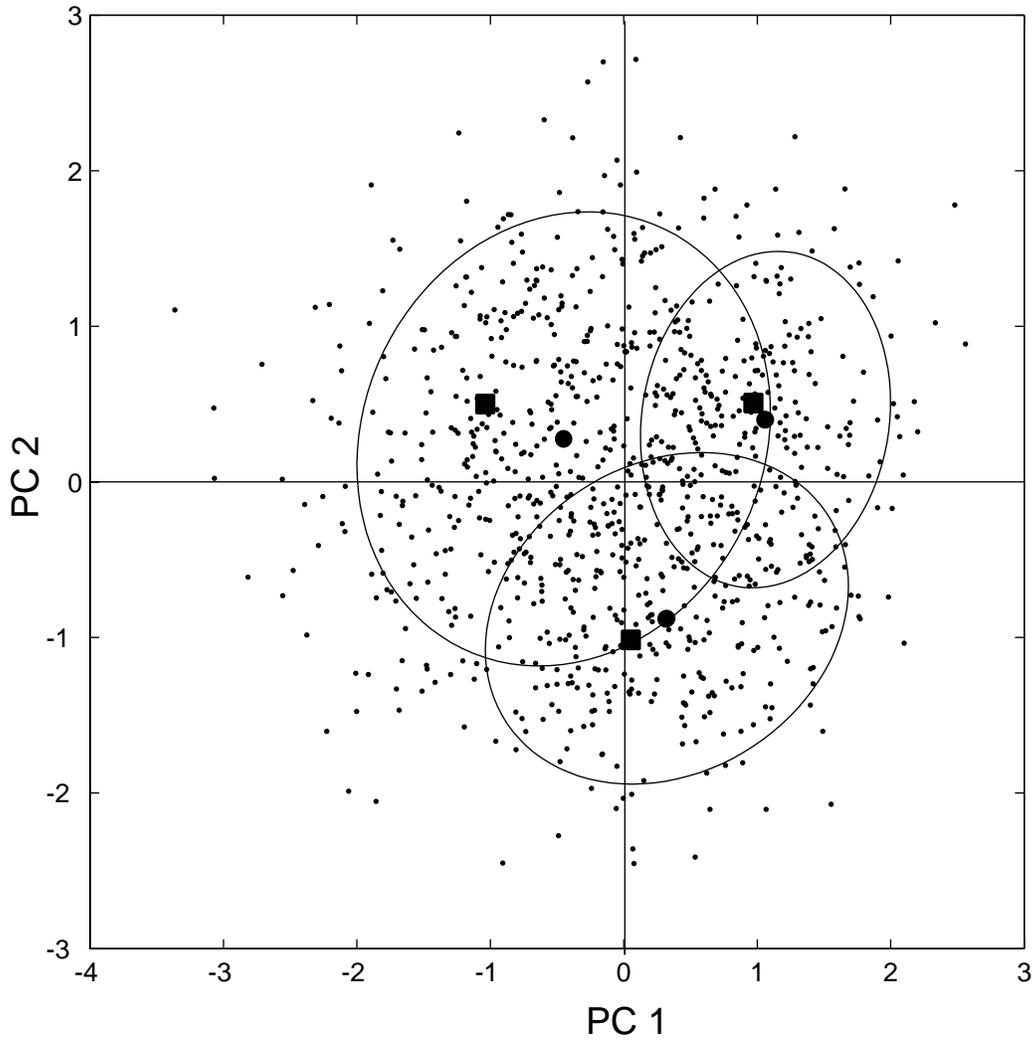


Figure 2

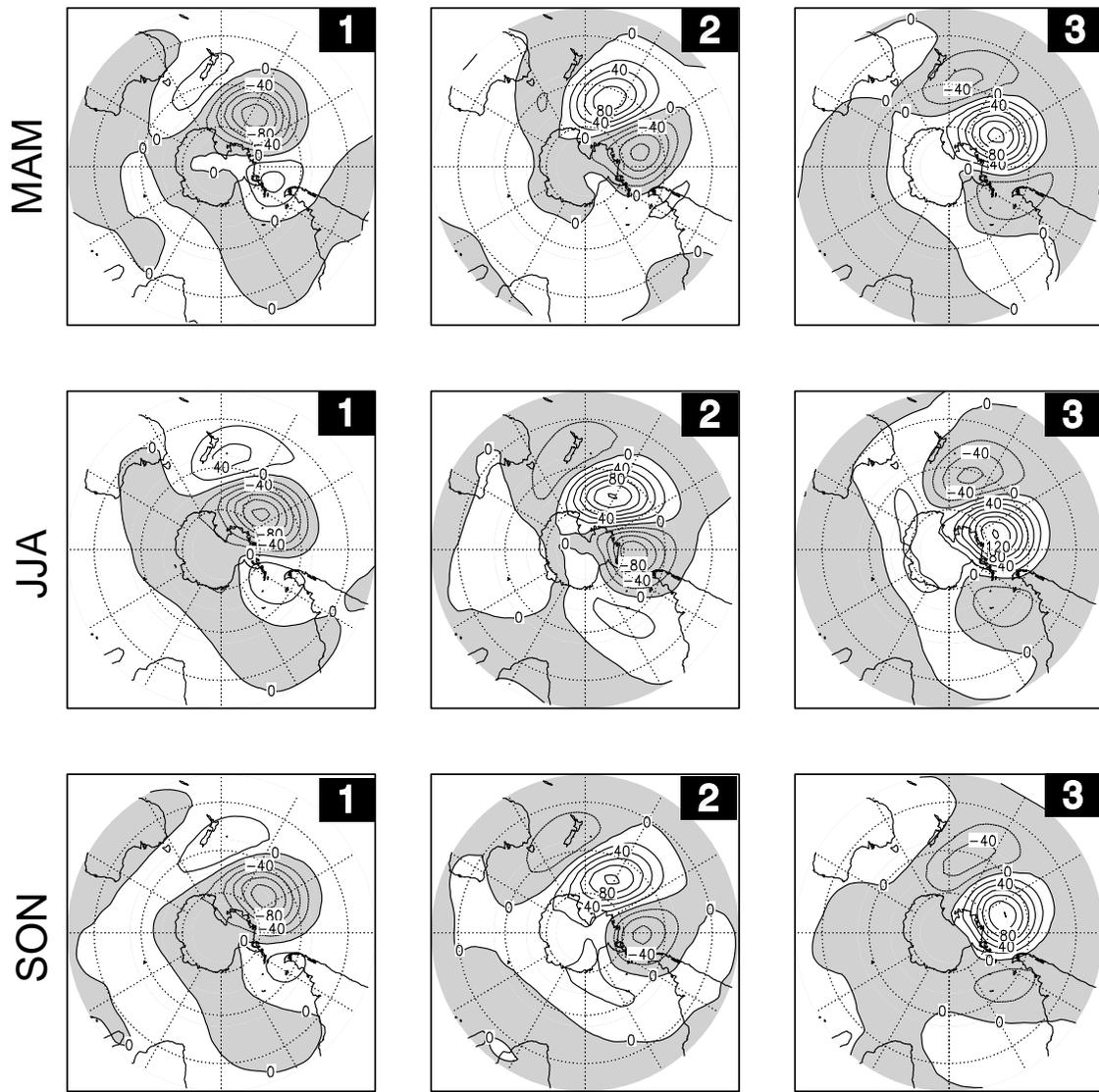


Figure 3

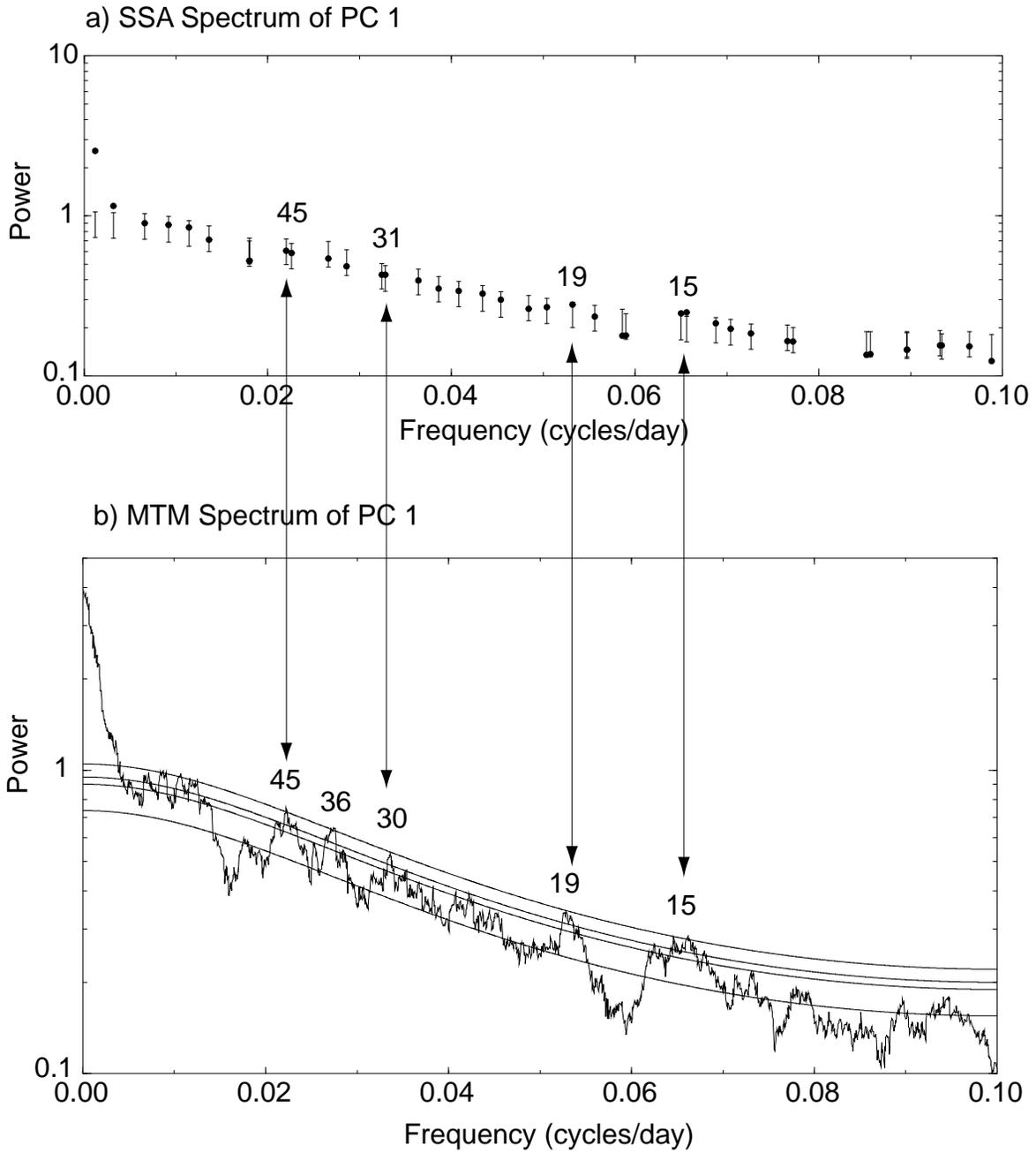


Figure 4

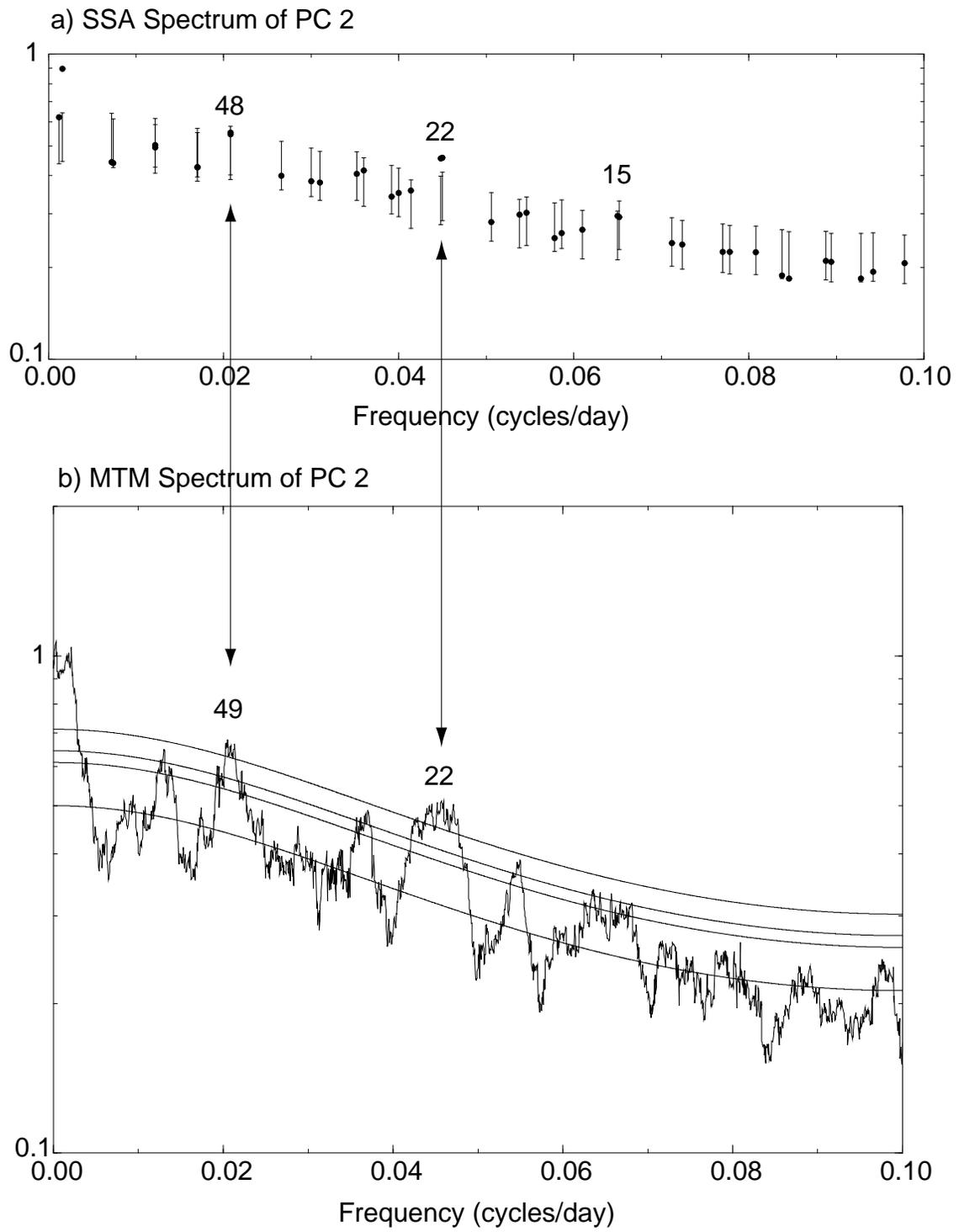


Figure 5

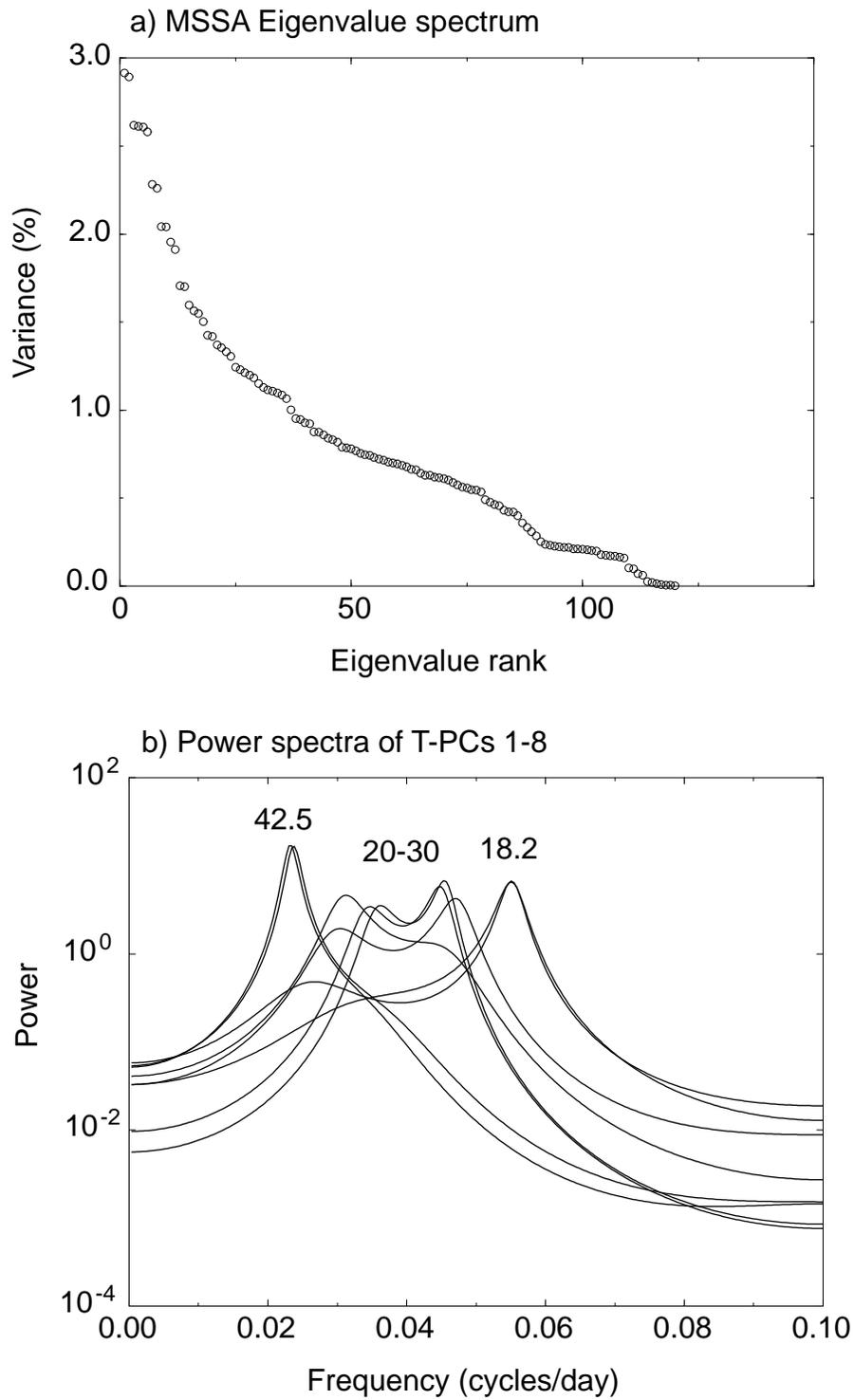
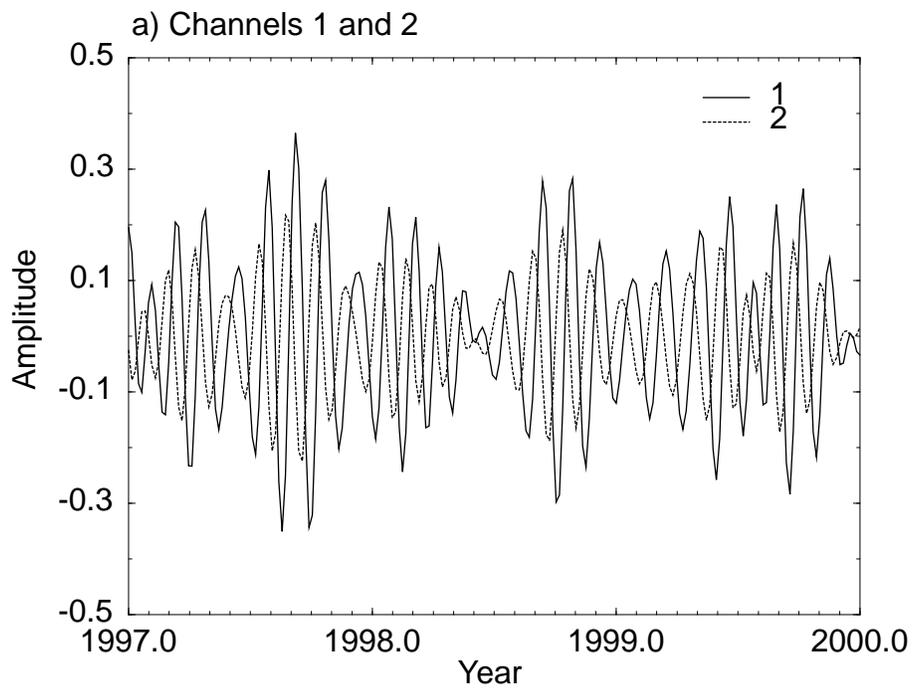


Figure 6



b) Phase composites (half cycle)

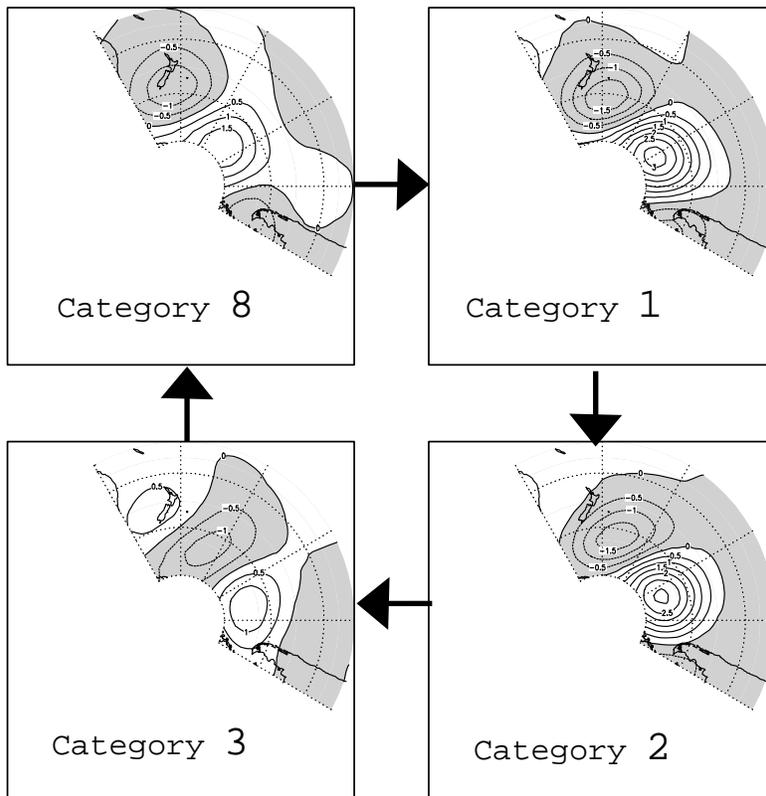


Figure 7a,b

c) Phase composites (averaged 50—60jS)

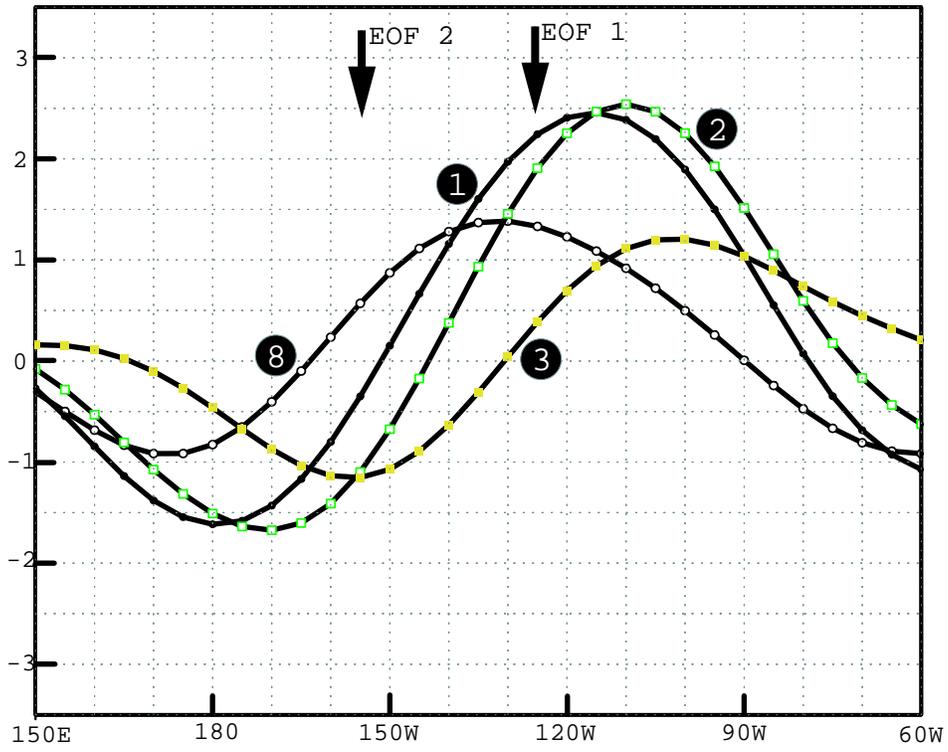


Figure 7c

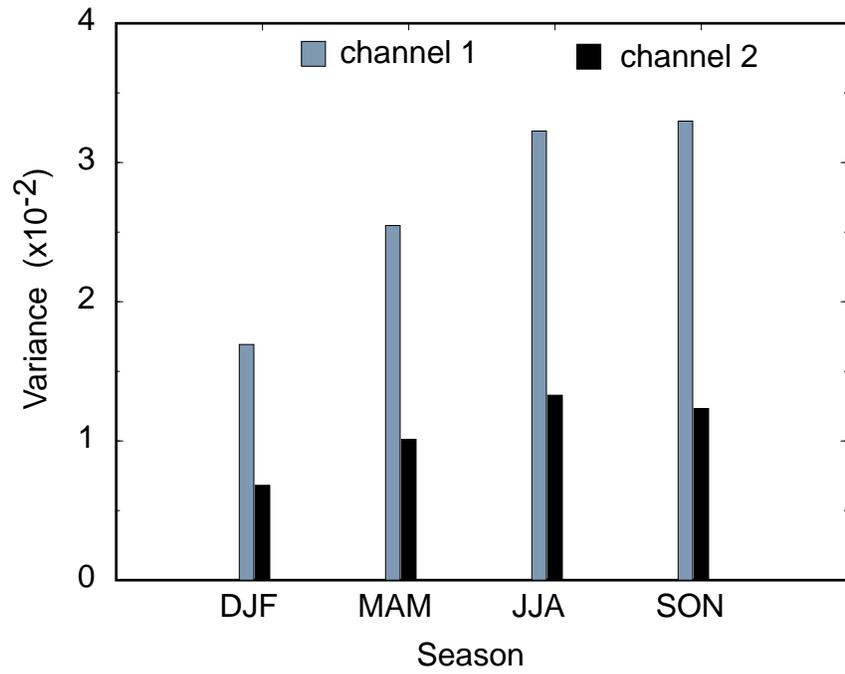


Figure 8

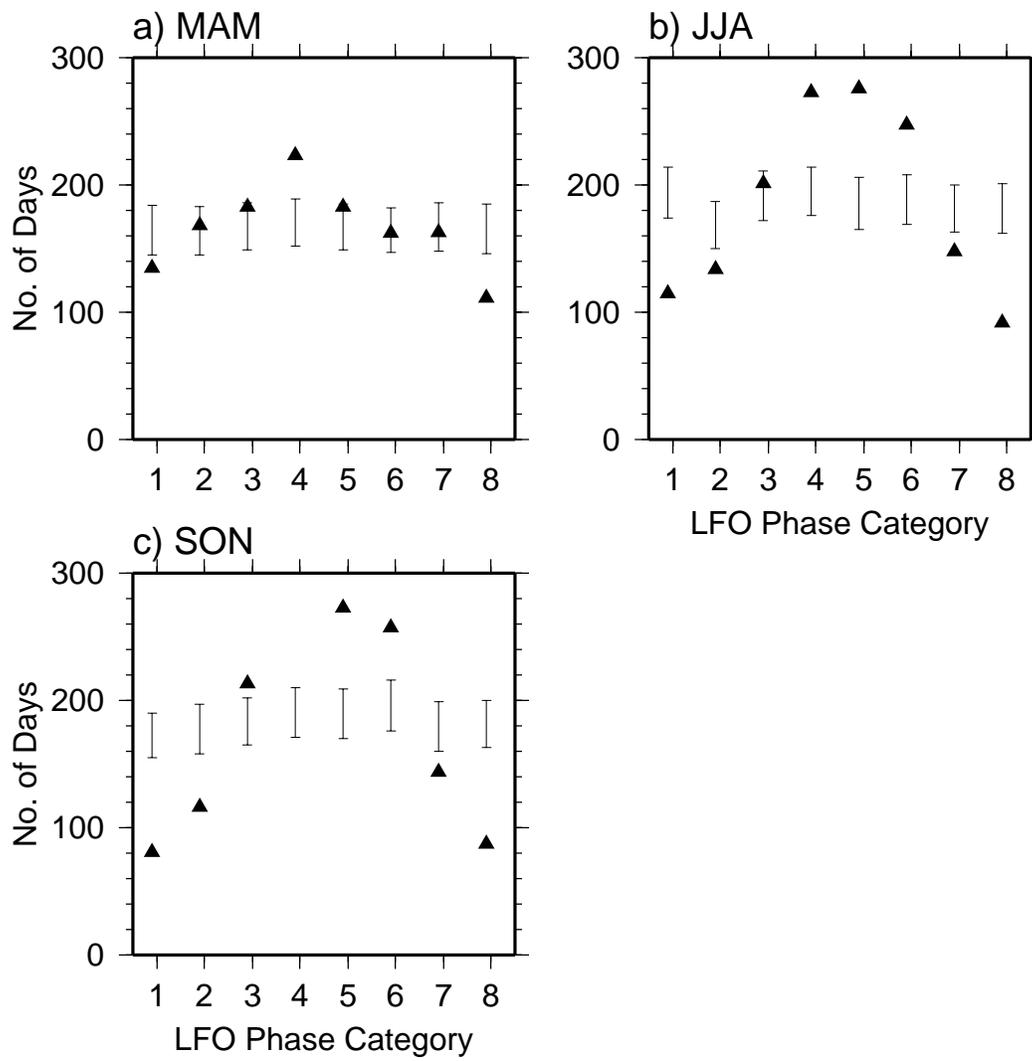


Figure 9

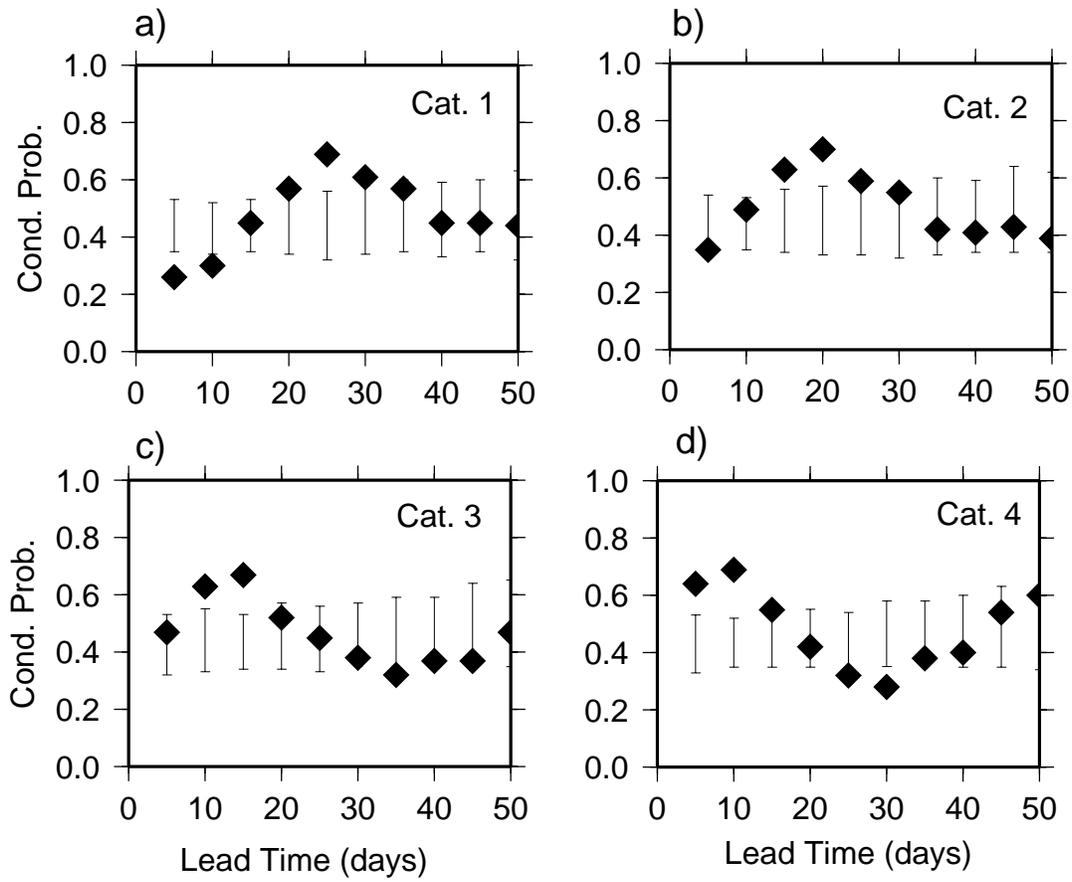


Figure 10