PRINCIPAL SEQUENCE PATTERN ANALYSIS: A NEW APPROACH TO CLASSIFYING THE EVOLUTION OF ATMOSPHERIC SYSTEMS

ROSA H. COMPAGNUCCI[a,*], DIEGO ARANEO[b] and PABLO O. CANZIANI[a,c]

[a] *Departamento de Ciencias de la Atmósfera, Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires/CONICET, Argentina*
[b] *Departamento de Ciencias de la Atmósfera, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina*
[c] *Laboratorio de Sistema Ecológicos y Ambientales, Facultad de Ciencias Naturales y Museo,
Universidad Nacional de La Plata, Argentina*

ABSTRACT

A new eigentechnique approach, Principal Sequence Pattern Analysis (PSPA), is introduced for the analysis of spatial pattern sequence, as an extension of the traditional Principal Component Analysis set in the T-Mode. In this setting, the variables are sequences of k spatial fields of a given meteorological variable. PSPA is described and applied to a sample of 256 consecutive daily 1000 hPa geopotential height fields. The results of the application of the technique to 5-day sequences demonstrate the advantages of this procedure in identifying field pattern sequences, thereby allowing the determination of the evolution and development of the systems, together with cyclogenesis and anticyclogenesis processes.

In order to complete the study, the more traditional Extended Empirical Orthogonal Function (EEOF) analysis, which is the S-mode equivalent of the PSPA, was applied to the same dataset. For EEOF, it was not possible to identify any real sequences that could correspond to the sequences of patterns yielded by the EEOF. Furthermore, the explained variance distribution in the EEOF was significantly different from that obtained with PSPA. Conversely, the PSPA approach allowed for the identification of the sequences corresponding to those sequences observed in the data. Using diagrams of the squares of the component loadings values, as a function of time, the study of the times of occurrence of dominant field characteristics was also possible. In other words, successful determination of periods where the basic flow was dominant and times when strongly perturbed transient events with a significant meridional component occurred, was facilitated by PSPA. Copyright © 2001 Royal Meteorological Society.

KEY WORDS: atmospheric circulation; extended empirical orthogonal functions; principal component analysis; principal sequence pattern analysis; synoptic climatology; T-mode approach

1. INTRODUCTION

One of the major aims of synoptic climatology since the beginning of the 20th century has been to determine and classify atmospheric circulation patterns. Early in the century, the methods employed toward this end were subjective. The investigator would determine, by his own subjective knowledge, the main characteristics of each circulation pattern, making it difficult to compare the outcomes from different sources or reproduce ones work.

In recent years, these methodologies have been replaced by objective techniques such as the Lund (1963) correlation methodology. However, methods which are more objective and mathematically efficient are derived from the eigentechniques, which Lorenz (1956) first applied to meteorological studies.

Lately, the Empirical Orthogonal Functions (EOF) and the Principal Component Analysis (PCA) have rapidly proliferated in meteorological publications, particularly in synoptic-climatic researches. A

---

* Correspondence to: Departamento de Ciencias de la Atmósfera, FCEN-UBA, Pabellón II-2do Piso, Ciudad Universitaria, 1428 Capital Federal, Republica Argentina; e-mail: rhc@at.fcen.uba.ar

description of PCA and its application to meteorological and climatological variables is included in Barry and Perry (1973), then in Preisendorfer (1988) and more recently in Yarnal (1993). Richman (1981) first discussed PCA from the point of view of its applications to circulation typing.

Cattell (1952) first developed the idea of taking a three dimensional problem, posing it as a data cube and creating two dimensional 'slices' of that cube. The various slices are called modes and the analyses bases on each slice, are known as T-mode and S-mode, for example. There are two types of PCA used to analyse a variable observed in either space or time. The former, named S-mode, involves the analysis of time series corresponding to different space-points. That is, the statistical variable under study is the time series of a meteorological variable that can be considered continuous in space and time. This application is the most commonly used in refereed literature. The S-mode attempts to isolate subgroups of grid-points that covary similarly (Richman, 1986). The latter, named T-mode, is applied in order to analyse spatial fields in different times. In this case, the involved statistical variable corresponds to the spatial field of the meteorological variable in question. By doing so, it isolates subgroups of fields with a similar spatial structure.

While the S-mode may allow for the identification of homogeneous regions with respect to time variability (e.g. the rainfall analysis in Austria performed by Ehrendorfer, 1987), the T-mode may be applied in hope of classifying the atmospheric spatial fields (e.g. case of surface pressure fields analysed by Compagnucci and Salles, 1997, among others). Richman (1983), Vargas and Compagnucci (1983), Compagnucci and Ruiz (1992) and Huth (1996) found that the T-mode proved to be a useful tool for extracting and reproducing the circulation types, quantifying their frequency and showing the dominant weather periods in them.

Yarnal (1993, pp. 82–83) stated that: 'eigenvector based map-pattern classification targets the main modes of spatial variation of just one variable; usually surface pressure or geopotential height. This is S-mode eigenvector analysis'. He added that: 'Buell (1975 and 1979) demonstrated that in an S-mode analysis, unrotated principal component manifest a regular sequence of loadings map that is unrelated to the spatial variation in the data (see also, Richman, 1986 and White *et al.*, 1991). These patterns are simply statistical artefacts. Hence, it is necessary to rotate the components'. That is to say, apply rotation (orthogonal or oblique) to the results obtained by means of the S-mode, so as to obtain the instantaneous circulation types or a snapshot of the actual circulation. These concepts are frequently used in the examples found in the literature. However, according to Huth (1993) 'The principal component analysis in S-mode is inapplicable for classifying the circulation. In other words principal components resulting from such an analysis cannot be identified with any dominant circulation pattern and should not be considered as circulation types.' Alternately, T-mode results should not be confused with teleconnection patterns since, for such a task, the S-mode is frequently useful.

For this reason, when using a group of data in the S-mode, PCA obtains a spatial regionalization of the main characteristics of the time evolution of one parameter and a T-mode PCA determines the main spatial types in a parameter. The results obtained by these two modes are completely different. Examples of these differences can be seen in work by Compagnucci and Vargas (1986) and Vargas and Compagnucci (1986), who applied both modes to the daily surface pressure for July from 1973 to 1983 over the region of the Southern Cone (southern portion of America) in order to find the main spatial patterns and regionalize the time characteristics of the variable, respectively. Also Drosdowsky (1993a and 1993b) analysed the monthly rainfall anomalies in Australia for the period 1950–1987 to find the spatial patterns and the time variability along with its connection to the Southern Oscillation, respectively.

Another important task in synoptic climatology is to classify sequences of circulation patterns. The circulation patterns, obtained as 'snapshots' through the application of the T-mode analysis, are valuable in identifying the different types of flow. Nevertheless, in order to reach a proper understanding of the development of systems, i.e. cyclogenesis and anticyclogenesis, their trajectory, behaviour and dissipation, it is also important to have adequate tools to determine the dynamic evolution of the atmospheric system. The capability to carry out such an analysis hinges on the statistical inference based on how the covariance or correlation are captured by the eigenanalysis. The T-mode analysis can lead to the

determination of frequent synoptic situations, improving the basic knowledge essential to weather forecasting, among other things. The application of such a tool to a wide range of processes, ranging from the daily synoptic developments to the monthly or annual mean developments is valuable for an ample set of atmospheric processes, including both daily variability and climate fluctuations and change.

We propose here an extension of the traditional T-mode PCA methodology in order to classify, as an example, sequences of k-days atmospheric circulation fields. The aim is to obtain the dominant types of a system's evolution in order to study the tracks and behaviour of synoptic systems in a specific area. We have named this new approach Principal Sequence Pattern Analysis (PSPA) given that the principal components (PCs) obtained by the analysis are patterns of k-sequence which are the types involved in the input set of spatial fields sequences.

Previously, Weare and Nasstrom, (1982) had proposed the Extended Empirical Orthogonal Functions (EEOF) analysis as an extension of the common utilized EOFs in the traditional S-mode. Interesting results have been obtained with this technique, e.g. Tangang *et al.*, (1998).

2. METHODOLOGY

The PSPA that is shown in the present work is an extension of PCA, in the T-mode, where the mathematical input variable is a sequence of spatial fields at k successive times (i.e. field sequences of length k, from now on referred to as sequences of k-spatial fields). Data are placed in a matrix $\mathbf{X}$ in which the columns are formed by the ordered sequence of a k-fields subset, i.e. with spatial fields at n consecutive times there are $(n - k + 1)$ columns. Each of the $n - k + 1$ successive column is built by lagging the k-field subset in one time step, i.e. the j column corresponding to time j is composed by the fields for times j to $j + k - 1$. Therefore, there are k rows for each of the m grid points (mathematical observations), each of which has a lagged time series with lags from 0 to $k - 1$, resulting in k_m rows and $\mathbf{X}$: $(k_m) \times (n - k + 1)$

$$
\mathbf{X} = \begin{bmatrix}
x_{1,1}\, x_{1,2}\, x_{1,3} \,................\, x_{1,n-k+1} \\
x_{2,1}\, x_{2,2}\, x_{2,3} \,...............\, x_{2,n-k+1} \\
....\\
x_{m,1}\, x_{m,2}\, x_{m,3} \,..............\, x_{m,n-k+1} \\
x_{1,2}\, x_{1,3}\, x_{1,4} \,................\, x_{1,n-k+2} \\
x_{2,2}\, x_{2,3}\, x_{2,4} \,...............\, x_{2,n-k+2} \\
....\\
x_{m,2}\, x_{m,3}\, x_{m,4} \,..............\, x_{m,n-k+2} \\
....\\
....\\
x_{1,k}\, x_{1,k+1}\, x_{1,k+2} \,..............\, x_{1,n} \\
x_{2,k}\, x_{2,k+1}\, x_{2,k+2} \,.............\, x_{2,n} \\
....\\
x_{m,k}\, x_{m,k+1}\, x_{m,k+2} \,............\, x_{m,n}
\end{bmatrix} \tag{1}
$$

The elements $x_{i,j}$ of the **X**-matrix correspond to the data of the meteorological variable under study, where i corresponds to observation point and j the time at which it is observed.

Before proceeding further, it must be noted that the same input matrix structure (i.e. a sequence of spatial fields at k successive times in columns) have been used in EEOF analysis (Weare and Nasstrom, 1982, from now on WN). In WN, the eigenvalues are obtained from the covariance matrix which corresponds to covariance between rows (i.e. $\mathbf{R} = \mathbf{XX'}$, $\mathbf{R}$: $(m_k) \times (m_k)$ is a lagged covariance matrix). This is equivalent to standardizing **X** by rows (i.e. correlation or covariance between rows) and corresponds to the S-mode approach.

In the PSPA, the mathematical variables are the sequence of k-spatial fields in the T-mode approach. Therefore, the eigenvalues and eigenvectors are obtained from the correlation matrix, which correspond to correlation between columns:

$$\mathbf{R} = \mathbf{X'_s X_s}/(m_k - 1) \qquad \mathbf{R}: (n - k + 1) \times (n - k + 1) \tag{2}$$

where $\mathbf{X_s}$ is matrix **X** standardized by columns (apostrophe represents a transposed matrix). The element $r_{p,q}$ of the **R**-matrix is the correlation coefficient between the columns p and q of **X** (i.e. correlation between the sequence p and the sequence q)

The difference in the standardization criterion (i.e. correlation matrix) leads to the major differences in the outputs PC matrix structures and values results. The latter can be seen in Section 4, where the results of both PSPA and EEOF are displayed and discussed for the same input matrix. Due to such differences, a brief description of the procedure for PSPA is now introduced showing the basic resulting equations, the matrix dimensions and their geometric meaning. The equations following Green's (1978) terminology where further details on PCA can be found, including the differences between the T-mode and S-mode approach.

The PSPA results can be obtained from Singular Value Decomposition (SVD) of **R**:

$$\mathbf{R} = \mathbf{U} \, \mathbf{D} \, \mathbf{U'} \tag{3}$$

where $\mathbf{U}$: $(n - k + 1) \times (n - k + 1)$ is an orthonormal matrix ($\mathbf{U'U} = \mathbf{UU'} = \mathbf{I}$), whose columns are the eigenvectors of **R** which can be viewed as sets of direction cosines that rotate $\mathbf{X_s}$ to PC orientation or new axes **Z**, and $\mathbf{D}$: $(n - k + 1) \times (n - k + 1)$ is a diagonal matrix whose entries are the eigenvalues in decreasing order (λ_j, $1 \leq j \leq n - k + 1$) of **R** and explain the variances of the point projections along the new axes or PCs of $\mathbf{X_s}$ (i.e. the variance of the new variables).

Alternatively, PSPA can be found from the SVD of $\mathbf{X_s}$ (**X** standardized by columns):

$$\mathbf{X_s} = \mathbf{P \Delta Q'} \tag{4}$$

where $\mathbf{P}$: $(k_m) \times (k_m)$ is the matrix of eigenvectors of $\mathbf{X_s X'_s}$ and $\mathbf{Q}$: $(n - k + 1) \times (n - k + 1)$ is the matrix of eigenvectors of $\mathbf{X'_s X_s}$ and $\mathbf{D} = \mathbf{\Delta^2}/(m_k - 1)$ and with $\mathbf{U} = \mathbf{Q}$. Further theoretical aspects about the unique SVD of any matrix **X** can be found in Green (1978), and its application to climate data in Bretherton *et al.* (1992), among others.

From Equation (3) or Equation (4), two sets of data result: matrix $\mathbf{Z_s}$ with the Standardized PC Scores and matrix **F** with the Component Loadings.

$$\mathbf{Z_s} = \mathbf{X_s} \, \mathbf{U} \, \mathbf{D}^{-1/2} \qquad \mathbf{Z_s}: (k_m) \times (n - k + 1) \tag{5}$$

The columns of $\mathbf{Z_s}$ are the new standardized variables and represent the principal patterns of the k-spatial field sequences.

The variance percentage V, explained by the j-component, is $V_j = (\lambda_j/\Sigma\lambda)100$ where $\Sigma\lambda$ is the **D** trace.

$$\mathbf{F} = \mathbf{U} \, \mathbf{D}^{1/2} \qquad \mathbf{F}: (n - k + 1) \times (n - k + 1) \tag{6}$$

Each column of **F** is the $\mathbf{f}_j$ component loadings vector and means the correlation of the original variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{(n-k+1)}$ (columns of **X**) and the j-principal component $\mathbf{z}_{sj}$ (j-column of $\mathbf{Z_s}$).

$$\mathbf{f}_j = \mathbf{X'_s z}_{sj}/(m_k - 1) \tag{7}$$

The original standardized data matrix $\mathbf{X_s}$, can be reproduced by postmultiplying the full rank matrix of standardized component scores $\mathbf{Z_s}$ by the transpose of the full rank matrix of component loadings $\mathbf{F'}$.

$$\mathbf{X_s} = \mathbf{Z_s} \mathbf{F'} \tag{8}$$

By doing so, a standardized variable $\mathbf{x}_{\mathbf{s}j}$ can be represented as a linear composite of the standardized components as:

$$\mathbf{x}_{\mathbf{s}j} = \mathbf{f}_{j,\,1}\,\mathbf{z}_{\mathbf{s}1} + \mathbf{f}_{j,\,2}\,\mathbf{z}_{\mathbf{s}2} + \cdots + \mathbf{f}_{j,\,n-k+1}\,\mathbf{z}_{\mathbf{s}(n-k+1)} \tag{9}$$

where the $\mathbf{z_s}$s are columns of $\mathbf{Z_s}$, the standardized components scores, and the $\mathbf{f}_j$s are the component loadings for the jth variable.

The sum of the squares of the elements in each row of $\mathbf{F}$ is equal to 1, $\Sigma_j f_{ij}^2 = 1$ and the sum of the squares of the elements belonging to each column of $\mathbf{F}$ is equal to the corresponding $\mathbf{R}$ eigenvalue (the variance of the component scores), $\Sigma_i f_{ij}^2 = \lambda_j$. Thus, the quantity f_{ij}^2 can be taken as the explained variance of the j-th component score over the i-th original variable.

3. DATA

In order to better explain the above methodology, it was applied to a set of 1000 hPa geopotential heights, obtained from the diagnostic model run by the Servicio Meteorológico Nacional (SMN). The dataset includes the daily values at 12 universal time (UT) in a regular 178-point grid, which covers the southern part of South America, the Antarctic Peninsula and adjacent seas.

The grid used is based on a map projection in which the scale varies over the grid. The points are then spaced uniformly on this grid, but are not uniform in latitude and longitude (see Figure 1 with the grid point and the region covered). As a consequence, there is no weighting of one sub-region in favour of another one. In other words, no sub-region of the analysed field has a greater information density than the rest and, thus, all have a similar contribution to the statistical analysis.

The period under study begins on 14 March 1997 and ends on 3 December of the same year, covering a sequence of 265 consecutive days. The dates were identified using a Julian calendar applied specifically to this sequence, i.e. day 1 corresponds to 14 March and day 265 to 3 December. Hence, the dataset contained $n = 265$ daily 1000 hPa geopotential height spatial fields, with $m = 178$ observational points each. With such a set, sequences of $k = 5$ days were determined and the $\mathbf{X}$ data matrix dimensions were $(k_m) \times (n - k + 1)$, i.e. 890×261.
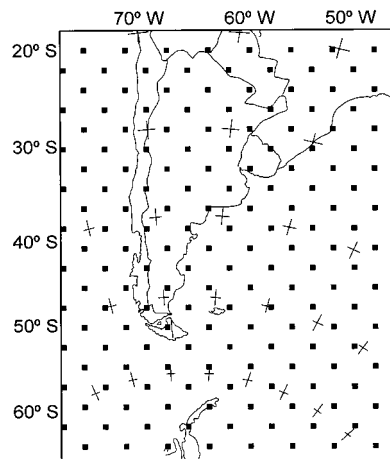

Figure 1. Grid showing the 178 locations spaced uniformly over southern South America and the adjoined oceans

Table I. Explained variance (%) and accumulated variance (%) for the first 14 components yielded by the PCA (traditional PCA with T-mode approach), by the PSPA, by the EOF analysis, i.e. S-mode approach, using correlation input matrix, and by the EEOF analysis (with correlation input matrix)

Component order	PCA		PSPA		EOF		EEOF	
	Variance (%)	Accum. var. (%)	Variance (%)	Accum. var. (%)	Variance (%)	Accum. var. (%)	Variance (%)	Accum. var. (%)
1	62.231	62.231	60.400	60.400	23.323	23.323	17.132	17.132
2	12.209	74.440	8.579	68.979	18.251	41.574	13.455	30.586
3	6.561	81.001	5.676	74.655	14.959	56.533	11.872	42.459
4	4.136	85.137	2.623	77.278	9.599	66.131	6.046	48.505
5	3.326	88.463	2.225	79.503	6.354	72.486	5.141	53.646
6	2.189	90.652	2.053	81.556	5.224	77.710	4.599	58.245
7	1.814	92.466	1.905	83.461	4.676	82.386	4.019	62.264
8	1.372	93.838	1.733	85.194	3.228	85.614	3.202	65.466
9	0.981	94.819	1.314	86.508	2.060	87.674	2.921	68.386
10	0.769	95.588	1.149	87.656	1.882	89.556	2.552	70.939
11	0.703	96.291	1.001	88.657	1.402	90.958	2.381	73.320
12	0.540	96.831	0.851	89.508	1.254	92.212	2.071	75.391
13	0.406	97.237	0.765	90.273	0.912	93.123	1.656	77.046
14	0.387	97.623	0.750	91.024	0.775	93.899	1.608	78.645

In cases where $k > 5$ was attempted, the coherency of the results decreased as it was difficult to fit a series of orthogonal vectors. This is related to the limit of predictability in the study region. In cases where $k < 5$, the synoptic development (i.e. translation) was not captured. Therefore, in other studies, one should use an autocovariance or autocorrelation function to help determine an optimal k. This is an important and necessary decision when applying PSPA.

4. RESULTS

4.1. Use of the PSPA approach

Through the application of the PSPA technique, a set of patterns of principal sequences, from now on referred to as PS (PS are the component scores or PCs, see Equation (5), and correspond to the $\mathbf{Z_s}$ columns), were obtained. Each of these models involved a series of five spatial patterns. Though only the first three will be shown here, more than three of these PS appear to be significantly different from noisy patterns due to the close similarity between the PS patterns and some of the real 5-day sequences involved in the data set. Those chosen are visibly distinct from noise, and together explain ∼ 74.6% of the total variance (see Table I). The (a) panels in Figures 2–4 show the PS patterns corresponding to the first three columns of $\mathbf{Z_s}$. According to the linearity inherent to PC, each PC can be multiplied by -1 to achieve a flip-flop property where either the patterns given by the PS in Figures 2–4 or their inverse could represent real 5-day sequences.

As explained in Section 2, the component loading matrix $\mathbf{F}$ represents the time series of correlation coefficients between the original sequences (mathematical variables) and each of the PS patterns. Therefore, the component loadings time series must be viewed together with the corresponding PS patterns in order to reach a full understanding of the results. They are shown in Figure 5 for each of the first three PS patterns in full line.

The *first PS* (Figure 2(a)), which explains 60% of the variance (see Table I), shows a quasi-stable field with little changes during the duration of the sequence. In all cases it represents a zonal flow, south of 40°S. To the north of the region it shows three systems, two of the same sign over the western and eastern coasts of the continent and one of opposite sign over northern central Argentina, Paraguay and Bolivia. It must
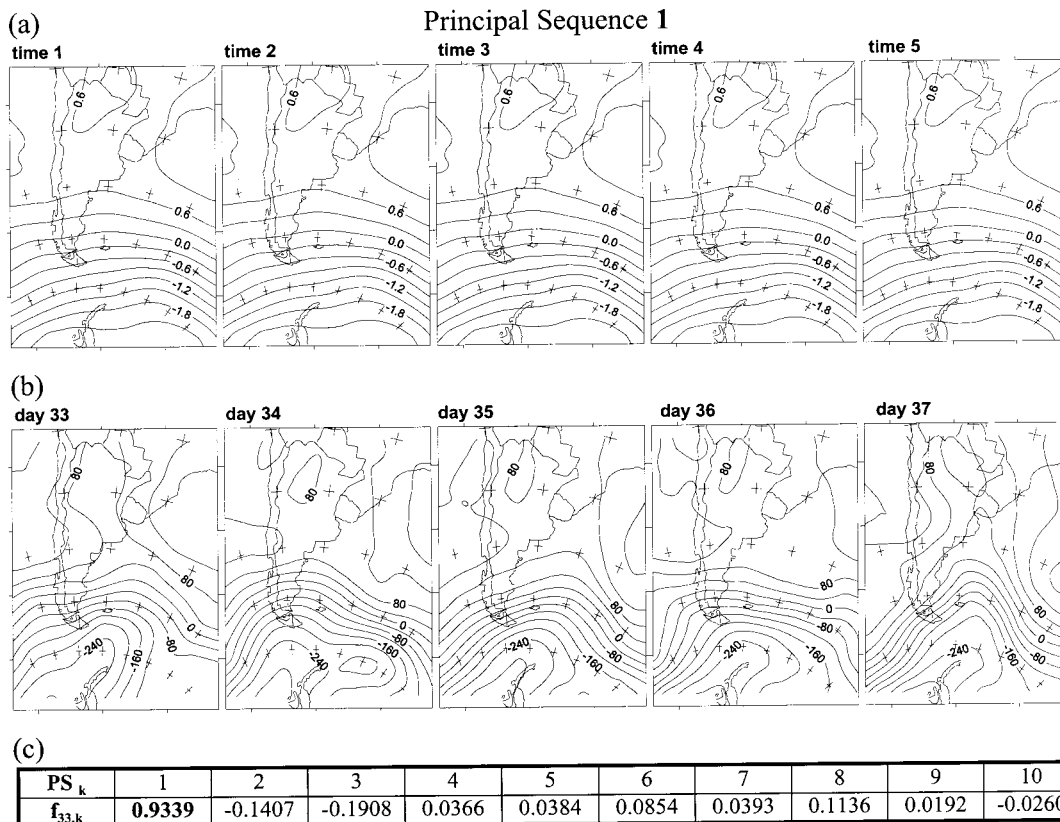
(a)   Principal Sequence **1**

(b)

(c)

PS$_k$	1	2	3	4	5	6	7	8	9	10
$f_{33,k}$	**0.9339**	-0.1407	-0.1908	0.0366	0.0384	0.0854	0.0393	0.1136	0.0192	-0.0260

Figure 2. (a) First PS patterns corresponding to the first column of the $\mathbf{Z_s}$ matrix. (b) Real 33rd sequence of surface fields including days 33–37. The isolines are plotted every 40 gpm. (c) Component loadings values for the 33rd sequence and the first ten PS patterns (i.e. $f_{33,k}$ with k between 1 and 10)

be noted that in this sequence of fields each frame has a structure very similar to that of the mean surface pressure field described by the classic climatology (e.g. Taljaard *et al.*, 1969; Satyamurty *et al.*, 1998).

The PC loadings time series for the first PS is in Figure 5(a) showing high correlation coefficients (near one) with many of the original sequences. Nevertheless, the absolute maximum is observed for sequence 33 (days 33–37, centre in day 35, i.e. 35th value of the time series). The component loading (i.e. correlation value) for that sequence is 0.94 and, thus, the first PS explains 87.2% of the total variance for the real sequence corresponding to days 33–37. In Figure 2(b), the original sequence for days 33–37 is shown. During the 5 days of the sequence, there is a strong dominant westerly flow south of 40°S. Such a flow is almost constant during the sequence as for the first PS, except for small high frequency perturbations, which introduce some deformations to the zonal flow. North of 40°S, over the continent a small low-pressure centre can be seen, which remains almost all the time over northern Argentina, as well as two high-pressure systems over the oceans. In Figure 2(c), the values for the first ten component loadings corresponding to the 33rd sequence are shown. These coefficients show that the contribution from the first PS is superior to the rest of the PS contributions for the real 33rd sequence. This PS pattern is the most frequent in the analysed sample. Since all the values are positive for the first PC loadings time series in Figure 5(a), then the 5-day sequence corresponding to the inverse of the Figure 2(a) with easterlies south of 40°S, does not occur as a real case.

The *second PS* (Figure 3(a)) which explains 8.6% of the variance (see Table I), shows the development of a pair of systems of opposite sign with a westerly displacement at mid latitudes, south of 35°S. In the first field (frame 1), a system with positive centre is located approximately at 57°S, 90°W and the negative one has its centre close to 50°S, 40°W. The eastward displacement of the positive centre can be clearly observed as it moves over the southern part of the continent. At the same time, the system displaces the
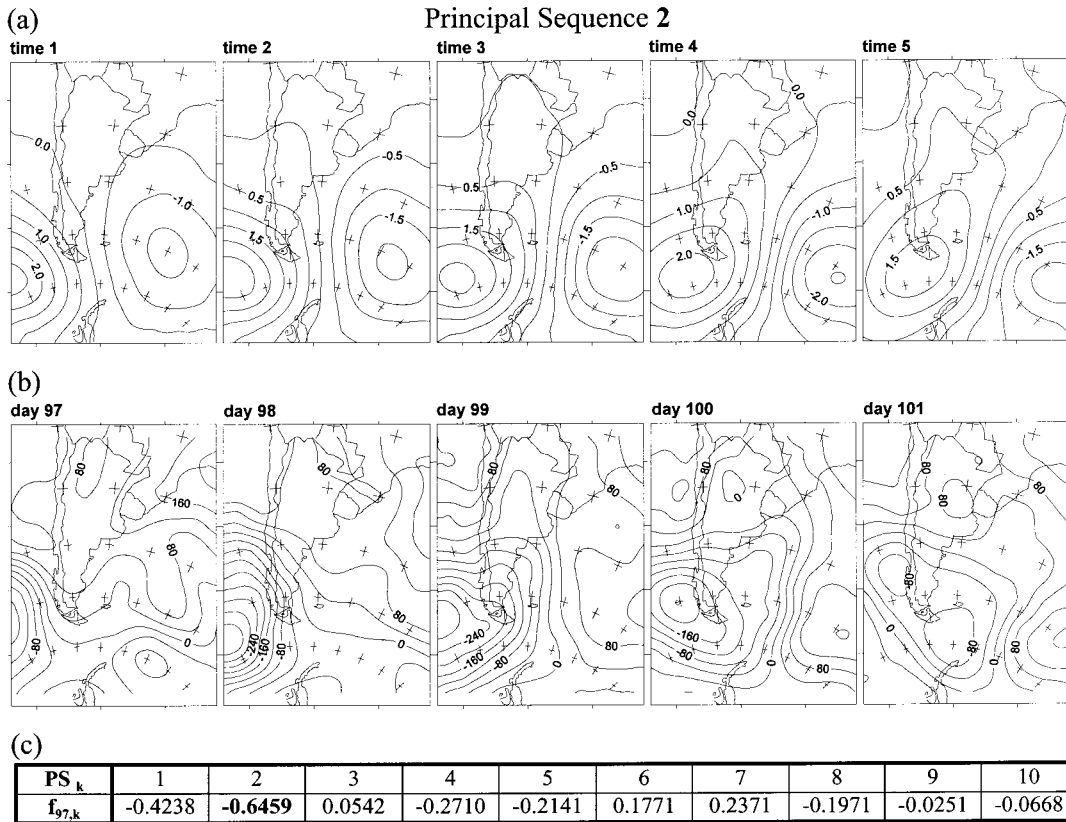
(a) Principal Sequence **2**

(b)

(c)

PS $_k$	1	2	3	4	5	6	7	8	9	10
$f_{97,k}$	-0.4238	**-0.6459**	0.0542	-0.2710	-0.2141	0.1771	0.2371	-0.1971	-0.0251	-0.0668

Figure 3. (a) Second PS patterns corresponding to the second column of the $\mathbf{Z_s}$ matrix. (b) Real 97th sequence of surface fields for days 97–101. The isolines are plotted every 40 gpm. (c) Component loadings values for the 97th sequence and the first ten PS patterns (i.e. $f_{97,k}$ with k between 1 and 10)

other one. In frame 3, the positive system covers the Argentine territory, while the negative centre continues its displacement along 50°S and is now found at 30°W. This displacement continues throughout the sequence until the positive one covers the whole southern part of the continent. These systems move along the 50°–60°S latitude band and have an eastward displacement of 15° in 5 days.

The PC loadings time series corresponding to the second PS, shown in Figure 5(b), yields a number of positive/negative extreme values that represent real sequences associated with the PS patterns given by the direct/inverse of the second PS. In Figure 3(b), one example of a real sequence that is fitted well by the inverse of the second PS patterns is shown. It is the real 97th sequence, which has component loading near −0.65 and, therefore, the second PS explains 47% of the total variance for that sequence. The 97th sequence shows a low-pressure centre which moves towards the east over the southern part of the continent, between 50° and 60°S. On days 97 and 98, this very deep low-pressure system is at ~55°S. While the location and motion of this system is very similar to that of the second PS, albeit of opposite sign, the PS patterns show a system in the south Atlantic with centre at 50°S, 40°W which cannot be clearly seen for days 97 and 98 in the real sequence. It does become distinct on day 99, however. At this stage, on the third frame, the flows given by the 97th sequence and second PS are practically the same with opposite sign. A high-pressure system can be observed over the Atlantic, almost co-located with the negative system described by the second PS, and the meridional flow over the Atlantic can observed as well. In the second PS and corresponding to the 97th sequence, during the following time step, both systems continue with their eastward motion and the depression covers the whole of the continent. During the last day of the 97th sequence, the low-pressure centre is located somewhat more to the east than the corresponding system in the second PS pattern. At the same time, the high-pressure system vanishes on
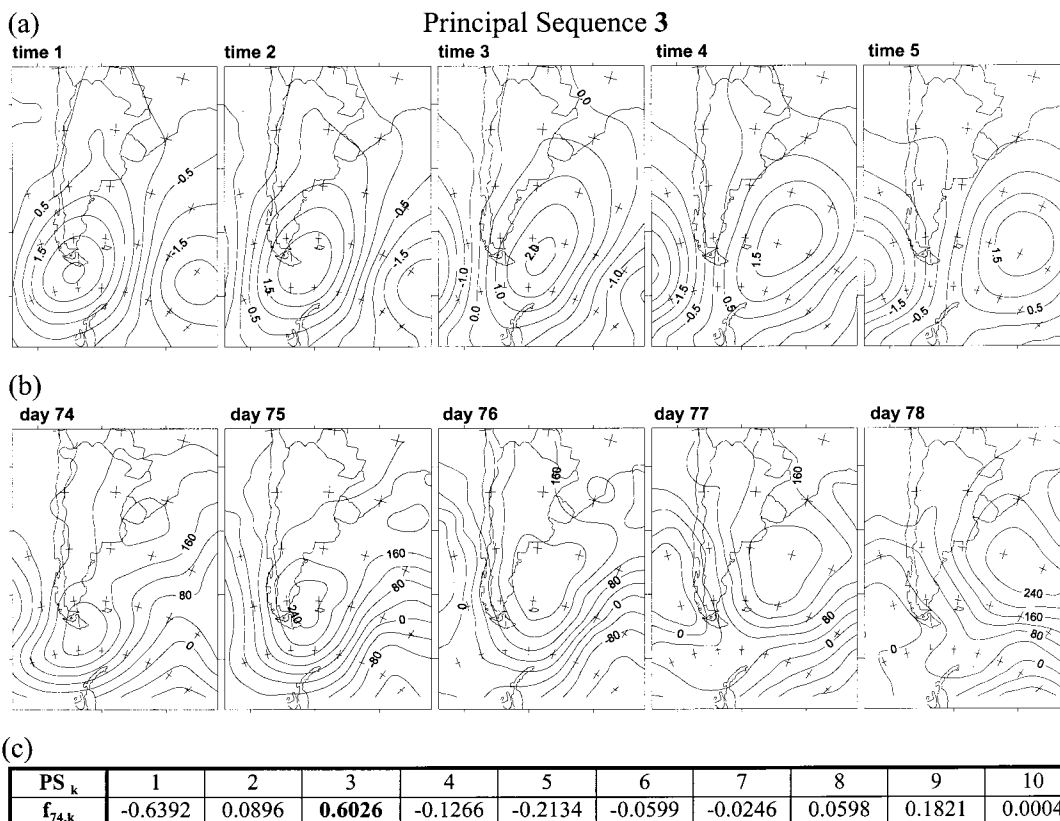
Figure 4. (a) Third PS patterns corresponding to the third column of the $\mathbf{Z_s}$ matrix. (b) Real 74th sequence of surface fields for days 74–78. The isolines are plotted every 40 gpm. (c) Component loadings values for the 74th sequence and the first ten PS patterns (i.e. $f_{74, k}$ with k between 1 and 10)

the eastern flank of the region under study for day 101, in agreement with the behaviour of the second PS for the fifth frame. In Figure 4(c), the component loadings values for the 97th sequence and the first ten PS patterns (i.e. $f_{97, k}$ with k between 1 and 10) are shown. The correlation coefficient between both for the second PS is ~ -0.65 and the negative correlation is in agreement with the inverse behaviour observed between the 97th sequence and the second PS systems. The same description may be carried out for other real sequences that are fitted well by the patterns (or the inverse patterns) depicted by the second PS.

The *third PS* (Figure 4(a)) that explains $\sim 5.6\%$ of the variance (see Table I) shows the northeastward displacement of a positive centre, originally located at 57°S, 70°W. This positive region extends as far as 30°S. At the beginning, it is also possible to observe a negative centre located at 50°S, 30°W, which is displaced in subsequent days, and actually goes out of the analysed region on the third day of the sequence, as another negative centre makes its appearance along the western border of the region under study. The large positive system continues its northeastward displacement throughout the period. At the end of the 5-day sequence, the new negative system is located to the southwest of the continent. Note that this configuration of systems has a displacement of more than 30° in longitude in 5 days along the East–North–East (ENE) axis between 50° and 60°S.

It would appear unreasonable to propose that second PS patterns and third PS patterns show more or less the same situation, albeit with a temporal difference. Nevertheless, when the displacement speeds of the systems are compared, it is important to note the different speeds involved in each sequence. The speed in the third PS is approximately twice that of the second PS. Furthermore, there is another significant difference between the two models. The separation between the system centres is not the same,
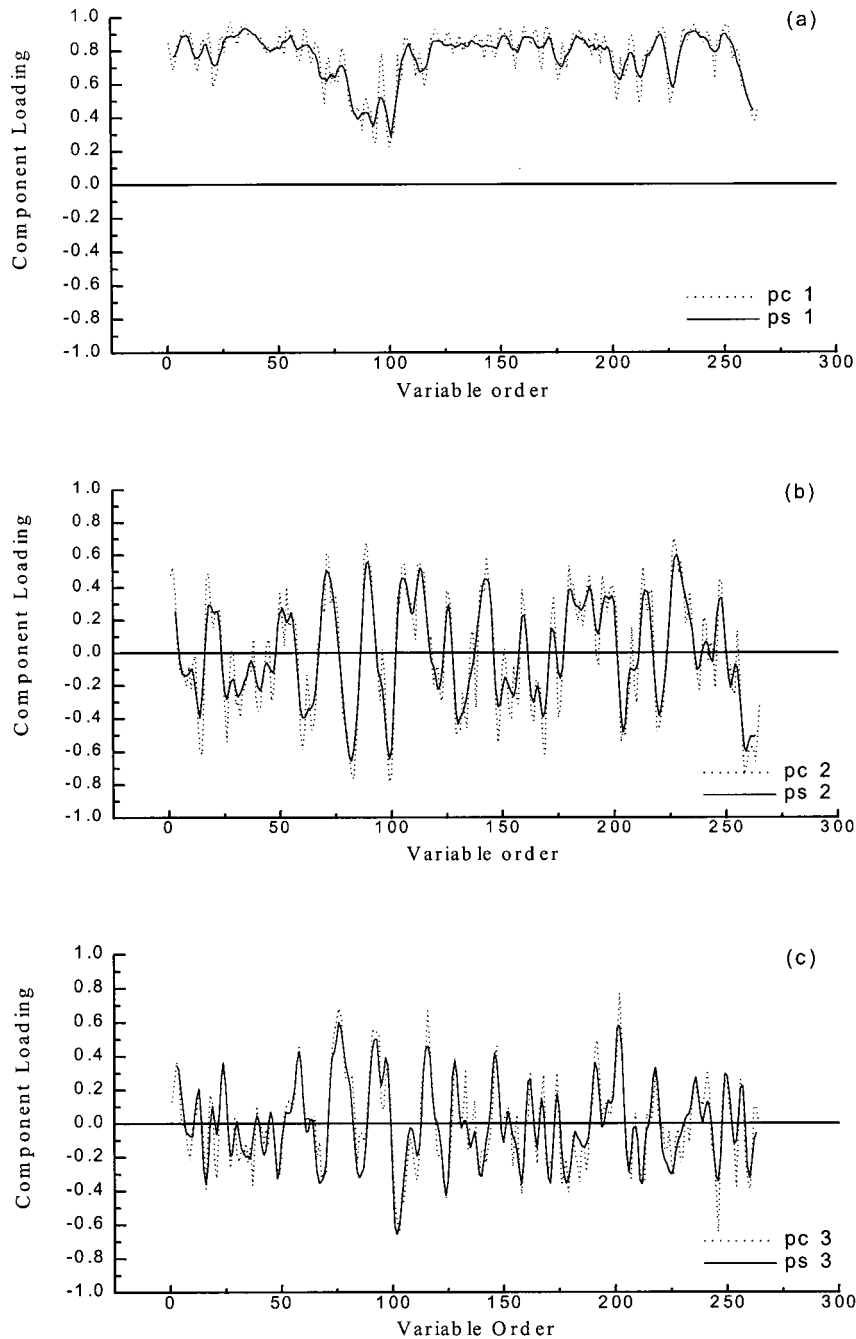
Figure 5. First three PC loading time series (a, b, c), showing the traditional PCA T-mode approach (dotted line: pc), and PSPA (solid line: ps). These values also are the correlation coefficients between each of the three PC patterns and each of the original variables for both methods

being smaller in the case of the third PS patterns. This implies a stronger flow than for the second PS patterns. Finally, the displacement itself of the systems is not the same. While the second PS presents a more zonal westerly motion, in the third PS the main system moves towards lower latitudes along a northeast axis. At any rate, both model sequences point to major alterations in the persistence of the basic flow, with the appearance of a significant meridional flow component.

The corresponding PC loading time series shown in Figure 5(c) with a bold line, has a number of values outside the limit ± 0.35 suggested by Richman and Gong (1999). Therefore, we could hope that some real sequences are well fitted by the third PS patterns as well as by the inverse. As an example, the real 74th sequence event corresponding to a positive maximum is discussed. In Figure 4(a and b), the third PS patterns and the 74–78 fields sequence, respectively, are shown. Day 74 presents a high pressure centre over the south extreme of the continent, similar to that observed in the third PS patterns during its initial frame. The other system in the third PS is a negative centre at 50°S, 30°W, coinciding with a trough in the actual data. The anticyclone has a southwesterly motion and on day 76 is located over the Argentine Sea, approximately 5° further north than the positive centre in the third PS at that time. Meanwhile, the trough over the Atlantic almost vanishes, both in the observations and the model and the anticyclonic circulation dominates the flow. On day 77, a low-pressure system makes its appearance on the western side of the region under study. Together with the anticyclone, this situation generates a strong northerly flow over the western side of the continent and the third PS in the fourth frame shows an almost identical pattern. This pattern moves eastward in the third PS as in the real sequence but located 4° further north, i.e. with the same zonal displacement velocity. During the 78th day, the anticyclone centre is located at 40°S, 45°W, establishing a NNW flow over the Argentine Sea. The ridge axis extends down to the Antarctic Peninsula. The low-pressure system generates a northwesterly circulation over central and southern Chile and from the northeast south of 60°S, on the western sector of the Antarctic Peninsula as seen in the last frame of the third PS. Furthermore, on the 78th day there exists a trough over the continent between 30° and 40°S and the third PS pattern is again in close agreement with this behaviour. Figure 5(c) shows a ~ 0.603 correlation between these two sequences (this implies that the third PS patterns explain the variance of 36.3% in the real 74th sequence). It must be noted that, while the correlation between the 74th sequence and first PS patterns is higher, the similarities between the observations and third PS are far more prominent.

A similar analysis can be carried out for other PSs of order higher than three, up to the PS whose patterns remain different from that produced by noise. Diverse techniques have been proposed in the literature, for the determination of the number of PCs with information significant different from noise. A wide range of these can be found in Preisendorfer *et al.* (1981) and Preisendorfer (1988). Nevertheless, according to Richman *et al.* (1992) the tests give a diverse number of components to retain. Therefore, the noise sequence patterns may be determined by inspection to the component loadings time series. According to Richman and Gong (1999), as no component loadings reside inside the hyperplane (with correlation-base PC loading, a value between zero and ± 0.20–0.35 according to the sample size) the patterns could be noise.

Further complementary information about the variances of the real sequences explained by the PS patterns can be obtained by using a time-variance diagram where the square of each of the entries for the $\mathbf{f}_j$ component loadings vector (Equation (7)) are plotted versus time. In Figure 6, the time-variance diagram for the first ten component loadings are presented. Such an approach provides insights into the temporal behaviour of the different components, periods when a particular PS is particularly relevant and even possible long-term linkages between the occurrences of the different PSs. The vertical axis shows the time from 1 (real first sequence) to 266 (real 266th sequence), while the horizontal axis shows the order of the PS from the second PS to tenth PS. *The first PS has been omitted since it represents the most frequently occurring situation and, hence, carries a very large part of the variance. Its inclusion obscures the details of all the remaining PS patterns.* The variances (f_{ij}^2) explained by the j PS for each real i sequence are shown by contours of equal variance. As it is possible to see in the accumulated variance for the low order PS (Table I), the components with order higher than ninth cannot contribute significantly to the time-variance diagram. The tenth PS, which has all loadings within the boundaries of the hyperplane, can not be distinguished from mathematical noise.

The advantage obtained by the use of such a diagram (Figure 6) can be seen, not only because it corroborates the conclusions that can be gleaned from the component loadings time series (first three in Figure 5), but also because it is possible to observe, as proposed above, the contribution from each PS to the real sequences. Thus, for example, when discussing Figure 5(a), it became clear that between days 70
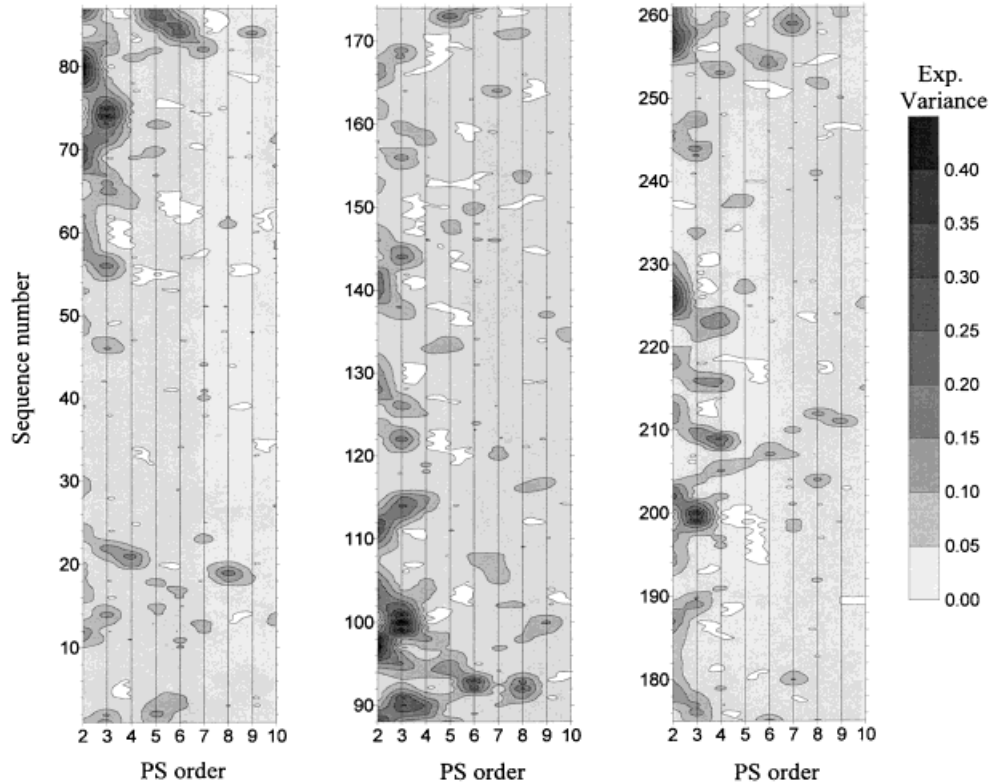
Figure 6. Time-variance diagram for the first ten PS yielded by PSPA: the square of each of the entries for the $\mathbf{f}_j$ component loadings vector (with $j = 1 \ldots 10$) are plotted versus time. The time evolution of the explained variance (f_{ij}^2) explained by the j PS for each real i sequence are shown by contours of equal variance (plotted every 5%). The first PS has been omitted since it represents the most frequently occurring situation and, hence, carries a very large part of the variance. Its inclusion obscures the details of all the remaining PS patterns

and 110 there existed a strong perturbation to the basic flow persistence represented by a non-zonal flow south of 40°S. The time-variance diagram in Figure 6 also yields that result, since it shows that other PSs than the first PS, such as second and third PS, dominate during that period with important strong meridional components, and actually specifies which are contributing most as time evolves. In other words, Figure 5(a) shows when the first PS is weak or irrelevant while Figure 6 indicates which other PS patterns become significant at those times.

Such a graphic presentation can also be useful in determining how many components should be considered in order to properly reproduce the behaviour of the variable under study, since it does show whether they are ever significant or not. In particular, for the higher order components or sequences that would be erroneously considered noise representation if a test as North *et al.* (1982) criterion were applied. According to Rutan and Smith (1992), high order PCs can contain signals, despite the fact that most eigenvalue tests would place them in the noise spectrum. This is because these tests examine only the total variance explained by each PC, which is indistinguishable from noise, but ignore the structure which gives rise to that variance. As a matter of fact, this analysis shows an interesting example. There are some specific times in Figure 6 where the eighth PS appear to have significant variance with high values that could be important for the description of the real sequence.

The *eighth PS* (Figure 7(a)) that explains only $\sim 1.7\%$ of the variance (see Table I) fits several real sequences very well, one of them the 212th sequence, shown in Figure 7(b). Prominent similarities can be observed between both the PS and the real example. In the first PS frame, a system with positive values over the Pacific can be seen, similar to the high-pressure system observed for the 212nd day in the
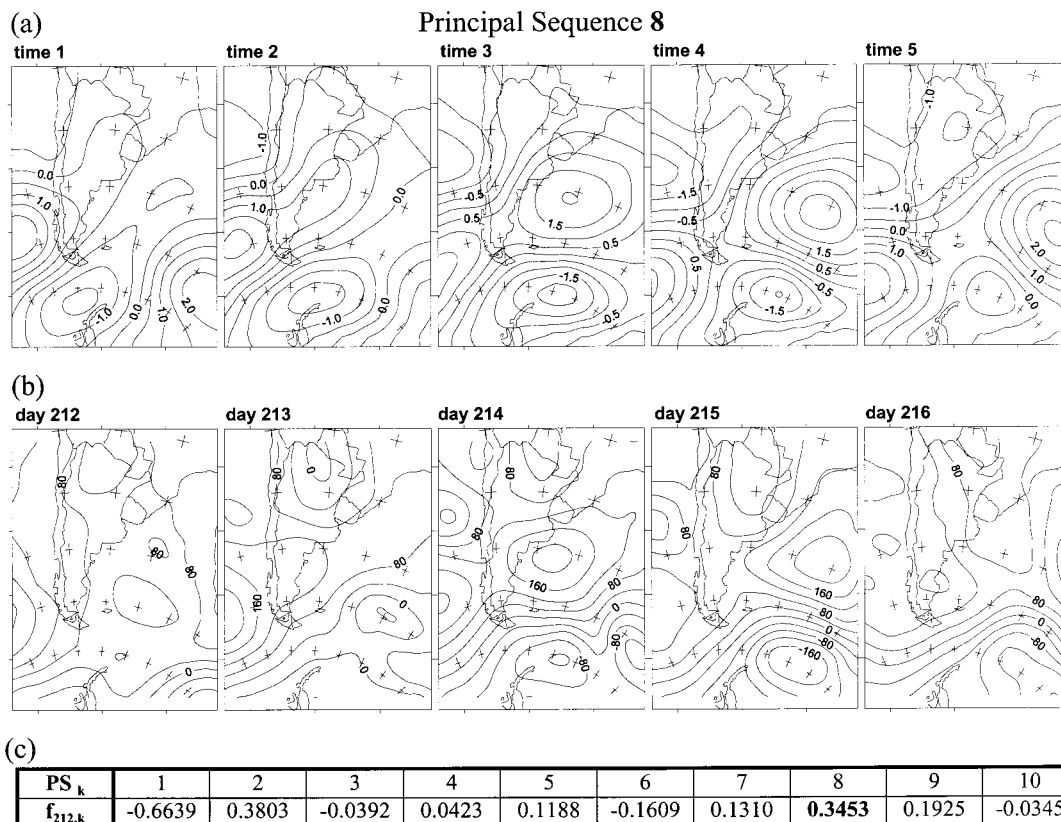
Figure 7. (a) Eighth PS corresponding to the eighth column of the $\mathbf{Z_s}$ matrix. (b) Real sequence of surface fields for days 212–216. The isolines are plotted every 40 gpm. (c) Component loadings values for the 212nd sequence and the first ten PS patterns (i.e. $f_{212, k}$ with k between 1 and 10)

example. Furthermore, the PS presents a negative system centred over the northwest of the Antarctic Peninsula with a northeast axis along the Argentine Sea. This system can also be observed in the real sequence as a trough over the Antarctic Peninsula as well as a low pressure area over the Argentine Sea. These systems establish a southwesterly flow over southern Chile, both in the example and the PS. During the next frame, the PS shows the positive centre moving eastward with an axis crossing the continent towards the Atlantic. This is similar again to the anticyclone and ridge observed on the 213rd day. At the same time, a negative centre is observed at 30°S, 80°W, which is faithfully reproduced by the PS pattern. Furthermore, a PS negative centre over the north of the Antarctic Peninsula seen in the PS, appears as a weak trough for the 213th day in the observations. At this stage both the real sequence and the PS yield similar circulation patterns over the Chilean coast and over the southern part of the continent. During the third frame in the PS and the 214th day in the real sequence, the similarities are particularly striking. At this time a ridge, south of 30°S, extending over the Atlantic has evolved into a high-pressure system to the east of the Argentine coast. It is associated with another anticyclone located to the southwest of the Chilean coast. These systems are almost in the same position both in the PS and real sequences. Furthermore, there are low-pressure centres, compatible with the negative centres present on the PS, located to the northwest of the Antarctic Peninsula and over the Pacific Ocean at 34°S, 80°W. This circulation pattern is almost identical in the observations and the PS pattern. During the next frame, the Atlantic systems have moved eastward, while those over the Pacific are quasi-stationary. This behaviour is again similar in both the observations and pattern. On day 216, all systems have a rapid eastward motion, resulting in some differences between the real sequence and corresponding PS frame. This is due to fact that the displacements in the PS are smaller.

In Figure 7(c), with component loadings for the first ten PSs on the 212th sequence shows a value of ~ 0.345 for the correlation with the eighth PS and, thus, the explained variance is only $\sim 11.9\%$ in spite of the close similarities between the PS pattern and the example. Furthermore, the component loadings (i.e. variances explained) for the first and second PS are far higher, but the actual fields for the 212th sequences are very different from the first PS patterns (see Figure 2(a)) and from that given by the second PS patterns. Again, the same unexpected result emerges as above between the third PS pattern and the real 74th sequence. This would imply that, if the component loading for two or more PS patterns are similar and outside the hyperplane threshold, the best fit to the observations could be given by the higher order PS.

The above results show how this method indeed determines the primary patterns of atmospheric system evolution, in this case short-term weather development for the southern part of South America. It is interesting that for the period under study, three sequential patterns suffice to explain a large portion of the variance, as indicated above. This does not mean, however, that higher order patterns are not representative of real situations as was also shown, but rather that the first three patterns are dominant.

Before proceeding further, it is instructive to compare the above results with more traditional PCA techniques in the T-mode approach (i.e. analysing individual snapshots). The first three PCs obtained by the traditional PCA T-mode technique are shown in Figure 8(a). Note the strong similarities between these modes and the central frames of the corresponding first three PS patterns, in Figure 2(a), Figure 3(a) and Figure 4(a), obtained with PSPA. The concomitant three PC loading time series obtained by PCA are shown in Figure 5 as the dotted line. The PSPA PC loadings are depicted by a bold line, which seems to be smoothed corresponding to PCA results.
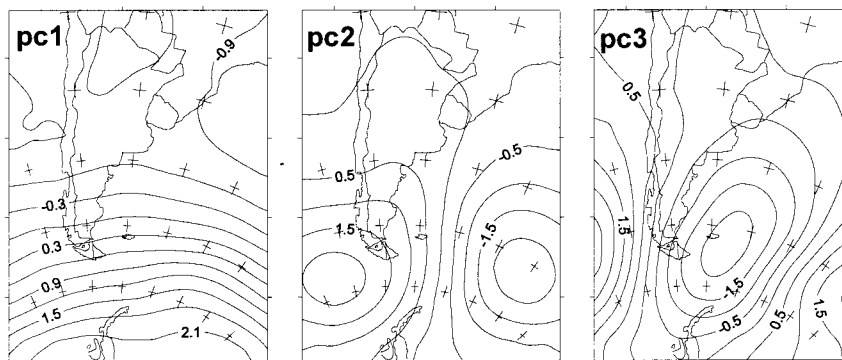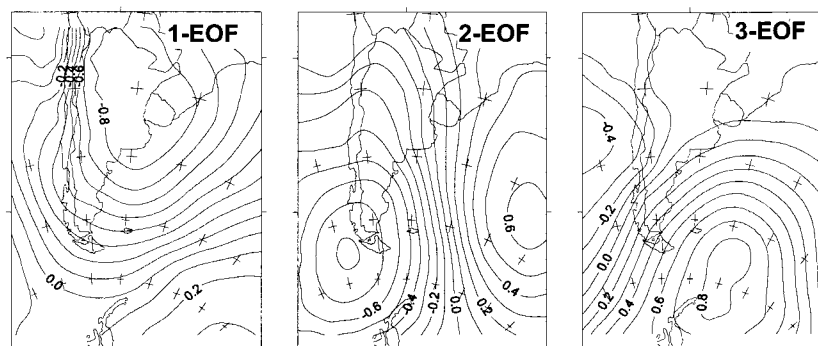

Figure 8. (a) The first three PC scores patterns obtained by the traditional PCA in T-mode approach (with correlation input matrix). (b) The first three PC loadings patterns (EOFs) obtained by the traditional EOF analysis (S-mode approach with correlation input matrix)

The explained variance for the first 14 PCs (Table I) obtained by the traditional PCA and the PSPA approach, together with the cumulative explained variance show a similar distribution with less explained variance for the PSPA than for the PCA results for the first to the sixth component. Nevertheless, when considering components with order higher than six the inverse occurs, i.e. less explained variance for the PC patterns than for the PS patterns. While traditional PCA requires only six PCs to account for 90% of the variance, the first 13 PS patterns are necessary to explain the same variance by the PSPA. This is reasonable, since in the PCA the results are independent patterns that can occur in any order in a given period, while the PSPA yields the more frequent sequences of evolution of the systems. Hence, as explained in the above examples, situations that look fairly similar at the outset may evolve differently after the first few time frames.

The similarities between the typical PCA in T-mode approach and the PSPA results could be expected. Since PCA in T-mode isolated group of spatial fields which co-vary in a similar fashion, the PSPA isolates subgroups of spatial sequences of fields which are of similar behaviour.

4.2. Application of EEOF

It is valid to enquire what the consequences would be of applying the EEOF method, as proposed by WN, to the same data matrix studied above, for the same aim, i.e. to classify temporal sequences of synoptic fields. The following results emerge from EEOF analysis as was described by WN, but using the correlation matrix between rows (WN used covariance matrix between rows, see Section 2), in order to compare the results with PSPA displayed above. Furthermore, the typical EOF analysis with correlation matrix between the time series of the grid point as input matrix is applied.

The variance distribution obtained though the application of EEOF (see Table I) is distinctly different from that obtained both by the PSPA approach and by the classical PCA. However, if the EOF with order $(i - 1)$ is compared with the EEOF with order (i) up to $i = 7$ the variance distribution between both results seem to be similar. Only the first three EEOF are clearly outside of the noise 'tail' in the LEV diagram (not shown here, this plots logarithmic eigenvalue against ordered root number (Craddock and Flood, 1969)), which is due to the small differences between eigenvalues with order equal to or higher than four. These can only explain 42.5% of the variance, which would suggest that the remaining components only correspond to noise. In the typical EOF, the first four EOF seem to be outside the 'tail' of noise patterns. The variance distribution appears to show that the variance accounted for by significant PCs obtained from the EEOF or the EOF is lower than that obtained by the PSPA and by the PCA.

In EEOF analysis, the component scores or PCs are time series as is the case of standard EOF analysis or S-mode PCA approach. The first three temporal series corresponding to the first three EEOF are shown in Figure 9. These are clearly different from the PC loadings time series obtained by the PSPA (Figure 5) as would be expected.

In order to understand the meaning of the spatial patterns given by the EEOF analysis, the component loadings were plotted for the first three components or EEOFs (Figure 10). As previously noted (WN), the central frame in the spatial sequence of map given by the EEOF is similar to that obtained by the EOF analysis (Figure 8(b)). The first EEOF (Figure 10(a)) shows a centre over the continent north of 45°S with a northeastward displacement. After 5 days, it is found over southern Brazil. The second EEOF (Figure 10(b)) yields a spatial structure, which at first glance looks similar to the second PS or third PS (Figures 3 and 4). However, a less superficial analysis, with a closer inspection of the plots, i.e. taking into account the corresponding time series, shows that this preliminary conclusion is false. Indeed the time series corresponding to the second and third PS are definitely different from that obtained for the second EEOF. In particular, the component scores time series corresponding to the third EEOF (Figure 9(c)) bears no resemblance whatsoever with any of the concomitant component loadings time series for the PS sequences in the PSPA (Figure 5), even when considering all of them up to the tenth PS (not shown here). However, if the fact that the component score time series corresponding to the third EEOF has an equal number of negative and positive values is not taken into account when comparing it to the component loadings time series for the first PS, which only has positive values, it could be argued that both series
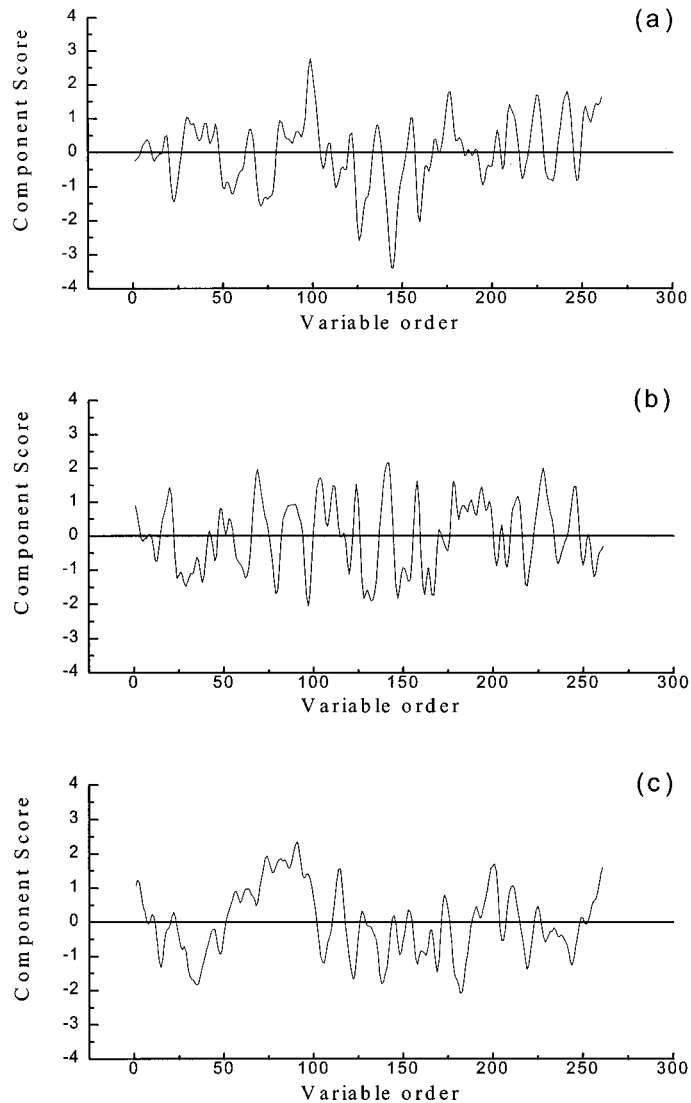
Figure 9. The first three PC scores time series (a, b, c) obtained by EEOF analysis with correlation input matrix

have a fairly similar evolution, but of opposite phase. Yet when comparing the spatial sequence patterns for the third EEOF (Figure 10(c)) and first PS (Figure 2(a)), no relationship can be found between them. While the first PS presents a zonal flow south of 40°S in all the frames, the third EEOF shows a centre over the northern tip of the Antarctic Peninsula. Furthermore, north of 40°S both sequences show different features.

In the previous section, it was shown that the sequence patterns obtained by the PSPA approach are representations of the real observed sequences of fields, illustrating the evolution in time of the synoptic systems. In order to clarify what happens with the particular case of the third EEOF, the same procedure was carried out to locate a similar sequence in the real sample. To accomplish this, the component scores time series was used to select a group of days when the series has extreme values. The maximum corresponds to days 91–95 of the series. In Figure 11(a), the frames showing the third EEOF pattern evolution is shown. In Figure 11(c), the sequence of real fields for that period (days 91–95) is shown too. As can be easily observed, these fields are different from the third EEOF sequence. The third EEOF patterns appear to suggest a flow induced by a high pressure over Antarctica which affects almost the
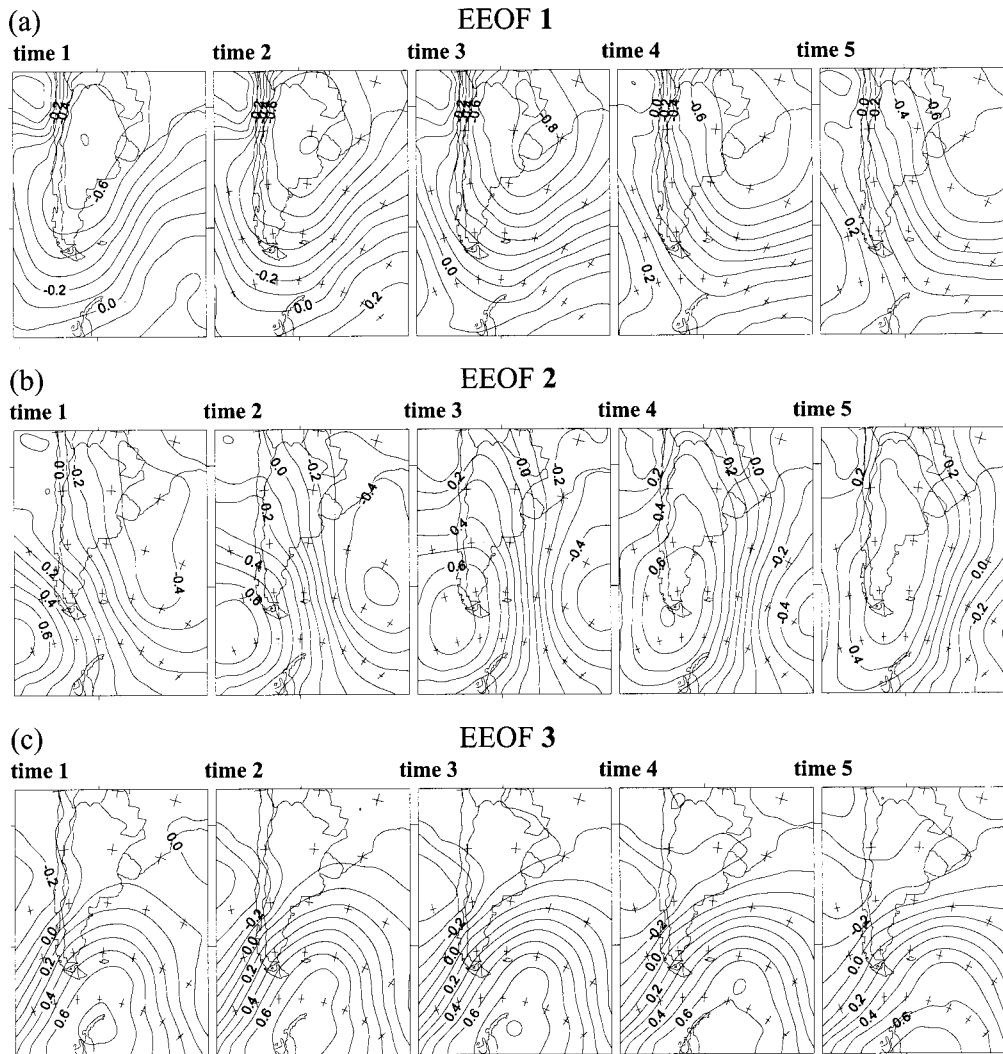
Figure 10. First three PC loadings sequence or EEOFs (a, b, c) obtained by EEOF analysis with correlation input matrix

whole of the continent, with a northeasterly flow over the southeast coast of southern South America (i.e. positive values of component scores during 91–95 days) as well as the persistence of this system for 5 consecutive days. When considering the real sequence, which in principle would have been best fitted by the third EEOF, it can be seen that the observed system is far more complex and changes over the span of the sequence. On the other hand, according to the time-variance diagram (Figure 6), the PS pattern from PSPA, which would be the best fit to this real field sequence, is the eighth PS shown in Figure 11(d). When comparing days 92–96 with eighth PS from frames 1–5, it is possible to see that the fit is very good and the systems can be distinctly followed. Hence, if only the eighth PS were available together with the time-variance diagram, it would still be possible to have a fairly good picture of the real synoptic behaviour for that period.

To gain a better idea of the differences between the EEOF and PSPA techniques when applied in order to analyse the time evolution of atmospheric systems, Figure 11(c) includes days 96 through 103. Figure 11(b) shows two other EEOF sequences, which would in principle better fit the real sequences for that period. These are the sixth EEOF for days 94–98 and first EEOF for days 99–103. These EEOF sequences were determined by the 'traditional' approach of choosing those EEOF with the largest values
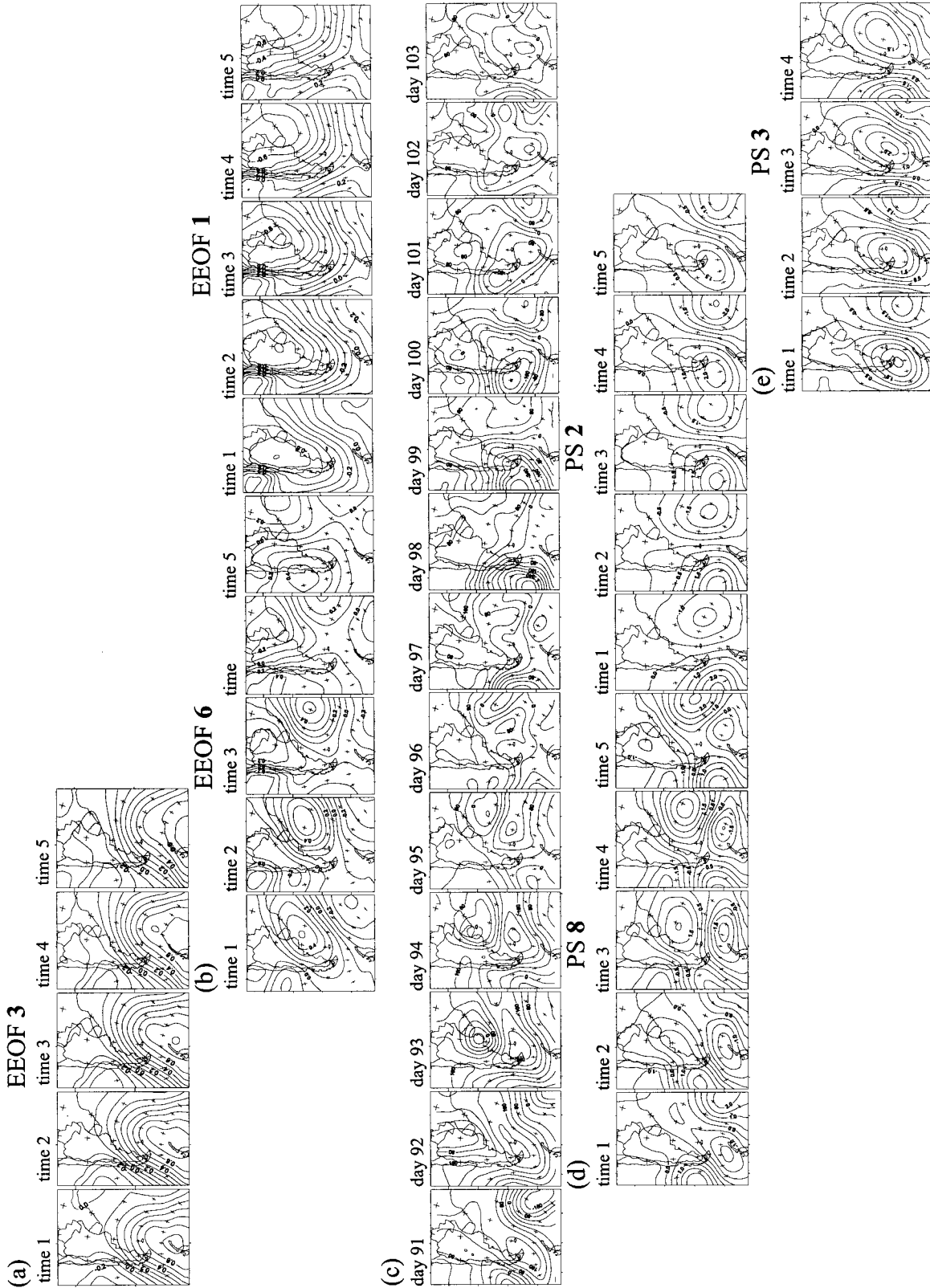
Figure 11. Comparison between results for the PSPA, EEOF analysis and the observed real sequences. (a) Third component loading sequence patterns (third EEOF). (b) Component loading sequence patterns for sixth EEOF and first EEOF. (c) Real sequence for days 91–103. (d) PS patterns for eighth PS and second PS given by the PSPA. (e) PS patterns for third PS given by the PSPA

in the component score time series, known in the literature as 'amplitude time series', following Kutzbach (1967) for the traditional EOF analysis. Figure 11(d and e) show the frames corresponding to the PS patterns that best fit the real fields, again determined by the time-variance diagram (Figure 6). These are the second PS for days 97–101 and the third PS for days 100–103. Again, it is clear that the days 100–103 bear no resemblance at all to the spatial pattern given by the first EEOF. Yet the third PS has an excellent fit for this period (101–102), where the centre southwest of the continent is well represented by it as well as by the second PS. Days 102 and 103 are reasonably well fit by the last frames of the third PS.

Therefore, it appears that PSPA can be reliably expected to yield a realistic evolution of synoptic flow patterns, whereas EEOF can not. However, EEOF may be useful in other types of applications, better suited to S-mode analyses.

5. CONCLUSIONS

The PSPA methodology is a useful approach to obtain spatial patterns and their time evolution. The results of the technique produce a good quality fit of the real fields. It is possible to identify model PS, which fit a real sequence of fields. Again, it is also possible to locate in time the observed sequence of fields best represented by a given PS patterns. This is done using the component loading time series, which are equivalent to correlation between the component score patterns and the real sequences. A better synthesis of the behaviour of the circulation is obtained through the use of a time-variance diagram. Such an approach allows the determination of links between the different sequences modelled by the PSs, persistence of the basic flows and perturbed periods in the circulation. It is possible to determine which sequences in the PSs are more frequent, i.e. how one system characteristically evolves into the next. Additionally it is possible to quantify the frequency of such sequences, using the explained variance for each PS. Therefore, while PSPA may be helpful in showing a preferred sequence of patterns, further application of dynamical techniques will be useful to provide a proper understanding of their evolution.

Alternately, EEOF analysis does not appear to yield this kind of information, and might lead to erroneous conclusions about system pattern evolution, their strength and temporal persistence. Furthermore, its application to synoptic climatology could result in wrong conclusions regarding the number and kind of significant sequences, as given by the eigenvalues. In consequence, the importance and evolution of a given system would be inadequately determined using the explained variance of the corresponding EEOFs. This does not happen if the latter analysis is used to study coherent areas for a temporal signature and the temporal evolution of the teleconnections. An example of such an application is Tangang *et al.*, (1998). In this work, the evolution over a 2-year period of the teleconnections generated by El Niño–Southern Oscillation (ENSO) over the Equatorial Pacific, during an ENSO cycle. In this manner the EEOF methodology yields important and significant results.

The advantage of the PSPA, when compared to the more traditional PCA T-mode, is given by the fact that it can be used to determine the kinds of evolution of meteorological systems. This is a valuable contribution for the study of dynamic climatology. By determining the types of system evolution, it is possible to study the variability of system evolution and climate change trends produced by circulation modifications resulting from mid to long-term changes in the development of meteorological systems. Therefore, careful consideration should be made to determine which technique is optimal.

Post-processing, by orthogonal or oblique rotation can be applied to the component loading matrix obtained in the PSPA method. This procedure does not modify by any means the meaning of the results. As a matter of fact, it could allow a better fit of the PS patterns to the real field sequences in the sample (Richman, 1986). In this work, no rotated results are shown because the aim was to present the method and an example of its application through a particular example. If, with the unrotated results, it was possible to determine the usefulness of the method and the meaning of the results, it can be expected that the application of rotation this quality would be further enhanced. However, it must be noted that the implementation of rotation could be more complex to operationalize since it would be necessary to determine the number of components that are to be used in the rotation.

REFERENCES

Barry RG, Perry AH. 1973. *Synoptic Climatology: Methods and Applications*. Methuen: London.
Bretherton CS, Smith C, Wallace JM. 1992. An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate* **5**: 541–560.
Buell CE. 1975. The topography of empirical orthogonal functions [*preprints, Fourth Conference on Probability and Statistics in Atmospheric Science*]. American Meteorological Society: Tallahassee, FL. pp. 188–193.
Buell CE. 1979. On the physical interpretation of empirical orthogonal functions [*preprints, Sixth Conference on Probability and Statistics in Atmospheric Science*], Banff, Alberta. American Meteorological Society: Boston, Massachusetts. pp. 112–117.
Cattell RB. 1952. *Factor Analysis*. Harper and Row: New York, NY.
Compagnucci RH, Vargas WM. 1986. Patterns of surface pressure field during July 1972–1983 in southern South America and the Antarctic peninsula. In *Proceedings of the Third International Conference on Statistical Climatology*, Cehak K (ed.). Australian Society of Meteorology: Vienna; 1–14.
Compagnucci RH, Salles MA. 1997. Surface Pressure Patterns during the year over Southern South America. *International Journal of Climatology* **17**: 635–653.
Compagnucci RH, Ruiz NE. 1992. On the interpretation of Principal Component Analysis applied to meteorological data. In *Proceedings of the Fifth International Meeting on Statistical Climatology*. Atmospheric Environmental Service of Canada: Toronto; 241–244.
Craddock JR, Flood CR. 1969. Eigenvectors for representing the 500 mb geopotential surface over the Northern Hemisphere. *Quarterly Journal of the Royal Meteorological Society* **95**: 576–593.
Drosdowsky W. 1993a. An analysis of Australian seasonal rainfall anomalies: 1950–1987. I: Spatial patterns. *International Journal of Climatology* **13**: 1–30.
Drosdowsky W. 1993b. An analysis of Australian seasonal rainfall anomalies: 1950–1987. II: temporal variability and teleconnection patterns. *International Journal of Climatology* **13**: 111–149.
Ehrendorfer M. 1987. A regionalization of Austria's pricipitation climate using principal component analysis. *Journal of Climatology* **7**: 71–89.
Green PE. 1978. *Analyzing Multivariate Data*. The Dryden Press: Hinsdale, Illinois.
Huth R. 1993. An example of using the obliquely rotated principal components to detect circulation types over Europe. *Meteorol. Z., N.F.* **2**: 285–293.
Huth R. 1996. An intercomparison of computer-assisted circulation classification methods. *International Journal of Climatology* **16**: 1–30.
Kutzbach JE. 1967. Empirical eigenvectors sea-level pressure, surface temperature and precipitation complexes over North America. *Journal of Applied Meteorology* **6**: 791–802.
Lorenz EN. 1956. Empirical orthogonal functions and statistical weather prediction. Scientific Report 1, Statistical Forecasting Project, Department of Meteorology, Massachusetts Institute of Technology: Cambridge, Massachusetts.
Lund IA. 1963. Map pattern classification by statistical methods. *Journal of Applied Meteorology* **2**: 56–65.
North GR, Bell TL, Cahalan RF. 1982. Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review* **110**: 699–706.
Preisendorfer RW. 1988. *Principal Component Analysis in Meteorology and Oceanography*. Elsevier: Amsterdam.
Preisendorfer RW, Zwiers FW, Barnett TP. 1981. *Foundations of Principal Component Selection Rules*. Scripps Institute of Oceanography, La Jolla, California. SIO Ref. Series 81-4 (NTIS PB); 83–146613.
Richman MB. 1981. Obliquely rotated principal components: an improved meteorological map typing technique? *Journal of Applied Meteorology* **20**: 1145–1159.
Richman MB. 1983. Specification of complex modes of circulation with T-mode factor analysis [*preprints, Second International Conference on Statistics and Climate*]. National Institute of Meteorology and Geophysics, Lisbon. pp. 511–518.
Richman MB. 1986. Rotation of principal components. *Journal of Climatology* **6**: 293–335.
Richman MB, Angel JR, Gong X. 1992. Determination of dimensionality in eigenanalysis [*preprints, Fifth International Meeting on Statistical Climatology*]. Environment Canada, Atmospheric Environment Service, Toronto. pp. 229–235.
Richman MB, Gong X. 1999. Relationships between the definition of the hyperplane width to the fidelity of principal component loadings patterns. *Journal of Climate* **12**: 1557–1576.
Rutan D, Smith L. 1992. Significance (tests) of EOFs for geophysical fields [*preprints, 12th Conference on Probability and Statistics in the Atmospheric Sciences*], Toronto. American Meteorological Society: Boston, Massachusetts. pp. 288–293.
Satyamurty PC, Nobre A, Silva Dias PL. 1998. South America. In *Meteorology of the Southern Hemisphere*. In: Meteorological Monographs, vol. 27, No. 49, Karoly DJ, Vincent DG (eds). American Meteorological Society: Boston, Massachusetts; 119–139.
Taljaard JJ, van Loon H, Crutcher HL, Jenne RL. 1969. *Climate of the Upper Air: Southern Hemisphere. Vol. I, Temperatures, Drew Points, and Heights at Selected Pressure Levels*. NAVAIR 50-1C-55, Chief Naval Operations: Washington DC.
Tangang FT, Tang B, Monahan AD, Hsieh WW. 1998. Forecanting ENSO events: A neural network-extended EOF approach. *Journal of Climate* **11**: 29–41.

Vargas WM, Compagnucci RH. 1983. Methodological aspects of principal component analysis in meteorological fields [preprints, Second International Conference on Statistical Climatology]. National Institute of Meteorology and Geophysics, Lisbon. pp. 531-539.

Vargas WM, Compagnucci RH. 1986. Homogeneous subregions according surface pressure variability during July in southern South America and the Antarctic peninsula. In *Proceedings of the Third International Conference on Statistical Climatology*, *Vienna*. Australian Society of Meteorology; 15–21.

Weare BC, Nasstrom JS. 1982. Examples of extended empirical orthogonal function analyses. *Monthly Weather Review* **110**: 481–485.

White D, Richman M, Yarnal B. 1991. Climate regionalization and rotation of principal components. *International Journal of Climatology* **11**: 1–25.

Yarnal B. 1993. *Synoptic Climatology in Environmental Analysis*. Belhaven Press: London.