Natural Hazards and Earth System Sciences (2001) 1: 83–92

Natural Hazards and Earth System Sciences

Decomposing spatio-temporal seismicity patterns

C. Goltz

Department of Geophysics, University of Kiel, Germany

Received: 21 May 2001 – Accepted: 16 July 2001

Abstract. Seismicity is a distributed process of great spatial and temporal variability and complexity. Efforts to characterise and describe the evolution of seismicity patterns have a long history. Today, the detection of changes in the spatial distribution of seismicity is still regarded as one of the most important approaches in monitoring and understanding seismicity. The problem of how to best describe these spatio-temporal changes remains, also in view of the detection of possible precursors for large earthquakes. In particular, it is difficult to separate the superimposed effects of different origin and to unveil the subtle (precursory) effects in the presence of stronger but irrelevant constituents. I present an approach to the latter two problems which relies on the Principal Components Analysis (PCA), a method based on eigenstructure analysis, by taking a time series approach and separating the seismicity rate patterns into a background component and components of change. I show a sample application to the Southern California area and discuss the promising results in view of their implications, potential applications and with respect to their possible precursory qualities.

1 Introduction

The identification and description of characteristic seismicity patterns has been and remains the most important approach in the effort to understand the earthquake process and to predict large events (e.g. Wyss et al., 1999, for a comprehensive collection of recent research papers on the topic). In the past, the search for precursory signals focused on local patterns situated around or near the source region of the future earthquake. Mogi donuts (Mogi, 1969) are a famous early example for propositions of this kind. Other such examples, non-exhaustively, include precursory seismic quiescence (e.g. Wyss et al., 1996), seismic gaps (e.g. Haberman, 1981), characteristic earthquakes (e.g. Swan et al., 1980), foreshocks (e.g. Dodge et al., 1996) and temporal clustering (e.g. Fröhlich, 1987). More recently, the application of the concept of critical point dynamics to the earthquake process (e.g. Bufe and Varnes, 1993) and corresponding numerical simulations (e.g. Rundle, 1988) combined with first observational evidence (e.g. Brehm and Braile, 1998) has led to the expectation of non-local, premonitory patterns, spanning up to hundreds of kilometers.

Both approaches, local as well as non-local, however, suffer from the fact that an appropriate region for the hypothesised patterns to occur must either be known in advance or otherwise somehow be determined from data. The fundamental problem thus lies in the objective decompositon and quantification of observed seismicity as a whole, into independent characteristic constituents. So far, methods attempting this decomposition have been either been somewhat arbitrary or purely qualitative in nature. I suggest that a spatio-temporal principal components analysis can address this issue.

The paper is organised as follows: in Sect. 2, I present the methodical approach by first discussing digital change detection and then introduce the principal components transformation, extending it to incorporate temporal change in addition to spatial change. The earthquake catalogue data used in this study is described in Sect. 3. Section 4 details the preparation of the data into a suitable time series and the subsequent application of the spatio-temporal principal components analysis. Results are discussed in Sect. 5. Conclusions and an outlook are found in Sect. 6.

2 Digital change detection

Digital change detection is an active field of research primarily in connection with remote sensing satellite data. Essentially, it attempts to quantify the spatio-temporal effects, such as environmental change in multitemporal data. The digital vectorial nature of satellite data makes it easily amenable for computer-aided analysis. The analogy with seismicity data prompted the approach presented in this paper. Detect-

Correspondence to: C. Goltz (goltz@geophysik.uni-kiel.de)

ing change possibly involves a procedure to transform the data into a more appropriate representation and a technique to delineate areas of significant fluctuations. Data of different dates may be classified independently before comparison or multitemporal data may be analysed simultaneously. Singh (1989) gives a comprehensive review of the different techniques in use. As shown below, a simultaneous analysis of transformed multitemporal data is most appropriate for seismicity data as it allows for the decomposition of the superimposed effects which are not discernible in the raw data.

Detecting change in high-dimensional data is fundamentally a problem of the appropriate representation of the data. Orthogonal transformations, such as Fourier- or wavelet--transformations, are generally employed to preserve the shape (information) of the data so that the vector norm is preserved (e.g. Gershenfeld, 1999). Wavelets are a generalisation of the global trigonometric functions of a Fourier transformation in that they introduce locality. However, it is still assumed that the data is best represented in time-frequency space. As this is certainly not always the case, it is desirable to customise an optimal orthogonal transformation for a given particular data set. Principal Components Analysis (PCA) provides a way to find such a transformation.

While PCA is widely used in linear, multivariate statistics and has many different applications (e.g. Jackson, 1991), its application in remote sensing is most closely related to the problem at hand. Namely, it is most often applied to multispectral image data (i.e. reflectance data of a region obtained simultaneously in several frequency bands) to reduce redundancy between bands or to make evident features not discernible in the original data (e.g. Lillesand and Kiefer, 1994). There have also been applications to multitemporal data, such as the detection of change in land cover type or vegetation index (e.g. Eastman and Fulk, 1993).

2.1 Principal Components Analysis

The starting point for PCA is the variance/covariance matrix

$$\mathbf{C}_x = \langle (\boldsymbol{x} - \boldsymbol{\mu}_x)(\boldsymbol{x} - \boldsymbol{\mu}_x)^t \rangle, \tag{1}$$

where $\boldsymbol{x}$ denotes vectorial data, $\boldsymbol{\mu}_x$ is the mean vector or expectation of $\boldsymbol{x}$, t denotes transpose and $\langle\rangle$ is the expectation operator. An unbiased estimate of $\mathbf{C}_x$ may be obtained from

$$\mathbf{C}_x = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu}_x)(\boldsymbol{x}_i - \boldsymbol{\mu}_x)^t \tag{2}$$

to describe the variability of the data around $\boldsymbol{\mu}_x$ in an l-dimensional vector space. n is the number of individual vectors $\boldsymbol{x}_i$ and should not be confused with the dimension m of each observable (which is two in the case of images). PCA of m-dimensional variables in an l-dimensional space only requires the appropriate rearranging of the input and output. The rationale for using a vector ($\boldsymbol{x}$) to designate multidimensional data is explained in more detail in the Appendix. The correlation matrix $\mathbf{R}_x$ with the elements

$$R_{ij} = \frac{1}{\sigma_i \sigma_j} C_{ij}, \tag{3}$$

where C_{ij} are the elements of $\mathbf{C}_x$ and $\sigma_i = \sqrt{\mathbf{C}_{ii}}$ is the standard deviation of the ith observable may be used instead of $\mathbf{C}_x$. Using $\mathbf{R}_x$ is equivalent to standardising the data by subtracting the mean and dividing by one standard deviation before obtaining $\mathbf{C}_x$. PCA based on the correlation matrix is termed standardised principal components analysis. In the case of high correlation between observables, the diagonal terms of $\mathbf{C}_x$ (the variances) will be small in comparison to the respective off-diagonal elements (the covariances). Covariances will be close to zero if there is weak correlation. If none of the observables correlate with any other, then both $\mathbf{C}_x$ and $\mathbf{R}_x$ will be diagonal.

To obtain a diagonal matrix $\mathbf{C}_y$ of the transformed data

$$\boldsymbol{y} = \mathbf{T}\boldsymbol{x} \tag{4}$$

is precisely the objective of principal components transformation. In other words, one seeks a transformation into a new coordinate system, where the transformed variables $\boldsymbol{y}$ will be uncorrelated. It is (e.g. Richards, 1986)

$$\mathbf{C}_y = \mathbf{T}\mathbf{C}_x\mathbf{T}^t. \tag{5}$$

Hence, assuming an orthogonality state, the transformation matrix $\mathbf{T}$ is the transposed matrix of eigenvectors $\boldsymbol{\epsilon}_i$ of $\mathbf{C}_x$. $\mathbf{C}_y$ is then the diagonal matrix of eigenvalues λ_i of $\mathbf{C}_x$. The eigenvalues of $\mathbf{C}_x$ are thus the variances of the transformed variables. As detailed here, the transformation is equivalent to the Karhunen-Loève (Loève, 1955) or Hotelling transformation (Hotelling, 1933). To compute $\mathbf{T}$, one first obtains the eigenvalues of $\mathbf{C}_x$ by solving the characteristic equation

$$|\mathbf{C}_x - \lambda\mathbf{I}| = 0, \tag{6}$$

where $\mathbf{I}$ is the identity matrix. The eigenvectors then result from

$$(\mathbf{C}_x - \lambda\mathbf{I})\boldsymbol{\epsilon} = 0 \tag{7}$$

with the constraint that $\mathbf{T}$ must be orthogonal, i.e. $\mathbf{T}^t\mathbf{T} = \mathbf{I}$. As $\mathbf{C}_x$, and, $\mathbf{R}_x$ are real symmetric matrices, it is always possible to find such a set of orthonormal eigenvectors (e.g. Golub and Van Loan, 1996). The transformed variables are finally obtained by applying Eq. 4. The new variables are called (principal) components or, in the case of two-dimensional observables, component images, hereafter denoted by $\mathbf{C}_{i,i=\{1,2,3,\ldots\}}$. An instructive numerical example for the complete procedure may be found in Richards (1986). The transformation may equally be achieved by a singular value decomposition (SVD) (e.g. Golub and Van Loan (1996)) of the original data, instead of an analysis of the eigenstructure of the covariance matrix (e.g. von Storch and Navarra, 1995; Freire and Ulrych, 1988), but it would be computationally more intensive in the current application.

Several points are noteworthy: the eigenvectors $\boldsymbol{\epsilon}_i$ define the new component axes in terms of the original axes, spanning a rotated orthogonal coordinate system. Hence, the classification of the principal components transformation as a rotational transformation. It is carried out such that the first component axis points in the direction of maximum variance

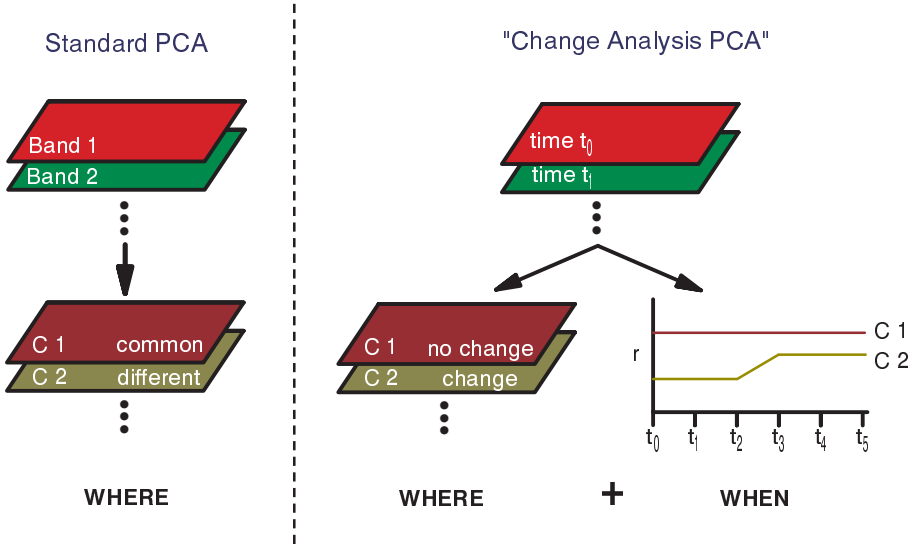

Fig. 1. Illustration of standard PCA (spatial, left) and "Change Analysis PCA" (spatio-temporal, right). Shown are (1) two, two-dimensional observables as a subset of higher dimensional hypothetical input data sets, (2) the resulting components with their properties and, (3) for the spatio-temporal case, the loadings graph which serves to identify the temporal significance of the respective components. $\mathbf{C}_1$ represents the stationary part, $\mathbf{C}_2$ is the change that occurs from t_2 to t_3 and persists during the period of observation.

(therefore, also the term "varimax rotation"), and the second axis points in the direction of second largest remaining variance, etc. It is in this sence that the principal components transformation is optimal concerning the decorrelation of original observables. By interpreting variance as information, the first (or principal) component $\mathbf{C}_1$ thus represents the majority of information contained in the data set, where the higher components contain increasingly smaller portions of the overall information. One may also say that $\mathbf{C}_1$ will contain what is common among observables (highly correlated, "characteristic" or "typical" features), while higher components will emphasize increasingly subtle differences (weakly or anti-correlated, "atypical" features).

2.2 Spatio-Temporal PCA

As detailed so far, PCA will only yield spatial information by means of the component images. Inspection or appropriate post-processing of the images will show *where* regions with common or different properties are located and how dominant these features are. To incorporate time into the analysis simply requires replacing, for example, different spectral bands by snapshots (time slices) of one "band" at different consecutive time intervals

$$[t_0, t_0 + \Delta t[, [t_0 + \Delta t, t_0 + 2\Delta t[, \ldots \tag{8}$$

The resulting component images will still give spatial information, but $\mathbf{C}_1$ will now show regions with temporal consistency, while the higher components will be increasingly subtle components of change. $\mathbf{C}_1$ thus represents the stationary or background process, and $\mathbf{C}_{i,i>1}$ represents the fluctuations. It is important to note that this only applies if the largest part of the total variance is due to the stationary constituent. Here, I use the terms stationary constituent and background process interchangeably and only in the sense above, i.e. not claiming stationarity of $\mathbf{C}_1$ in the strict statistical sense. The temporal information about the onset and duration of changes may be obtained from the so-called loadings graph, where the correlation r between each component and all time slices is obtained. If a particular component isolates a particular change effect, there must be a sudden increase in r in the case of rapid-onset change. For gradual change, r also should increase gradually. In fact, one might require spatio-temporal coherence of any change for it to be considered a significant change rather than an inconsequential variation. Coherence designates patterns have a tendency to persist over space and time (Eastman et al., 1994).

For an interpretation of a spatio-temporal PCA, both the component images and the loadings graph have to be regarded simultaneously. Note that for $\mathbf{C}_1$ to represent the stationary part, one would expect a consistent high r with time. The idea of spatio-temporal PCA in comparison with standard PCA is illustrated in Fig. 1.

Spatio-temporal PCA has been applied by Byrne et al. (1980) (land cover change), Howarth and Boasson (1983) (urban areas), Richards and Milne (1983) (bush fires and revegetation) and Eastman and Fulk (1993) (vegetation index). The work by Eastman and Fulk (1993) most resembles the approach taken here in that it compares data of more than only two different dates. Tiampo et al. (2000) and Rundle et al. (2000) applied PCA to seismicity patterns. While their approach is similar to the one taken here, their focus is on the expansion of seismicity patterns into physical modes and the subsequent prediction of large earthquakes associated with these modes. PCA components do not neccessarily represent physical modes, however, as will be discussed below.

3 Data

Seismicity data for the Southern California are obtained from the Southern California Earthquake Data Center. Data from January 1984 to July 2000, where the SCEDC catalogue is complete for events of magnitude $M \geq 1.8$, was selected, resulting in about 280 000 events. This catalogue contains several large events of different spatio-temporal qualities, as

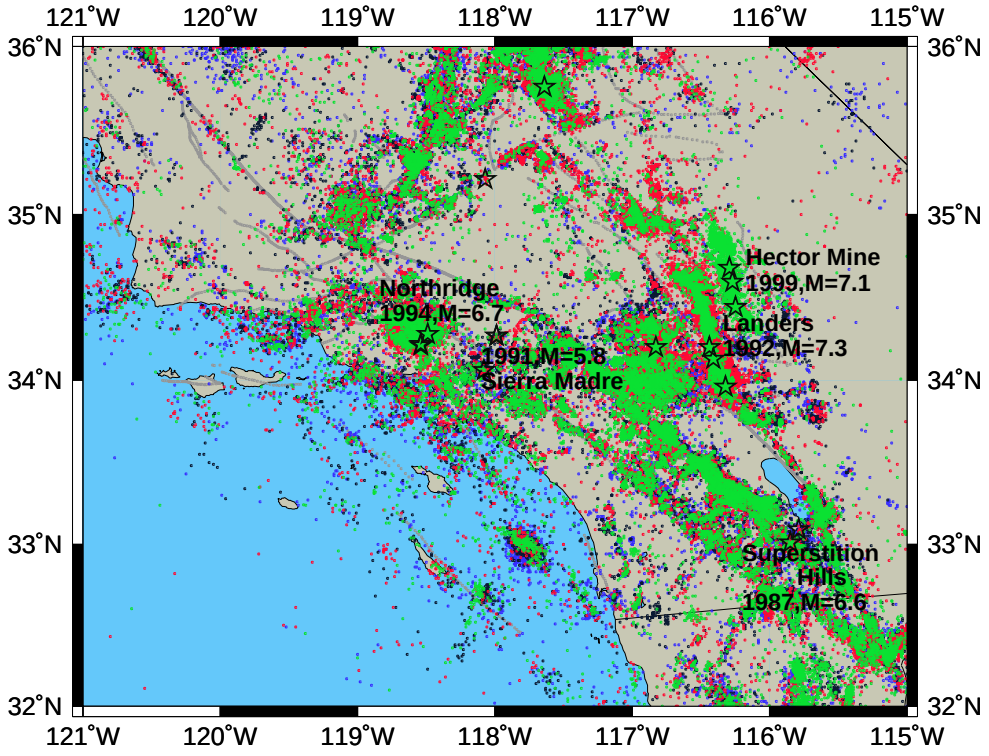

Fig. 2. Southern California seismicity from January 1984 to July 2000. The temporal colour-coding demonstrates that seismicity may be regarded as a stationary process with superimposed fluctuations (1984–1987: blue; 1988–1991: black; 1992–1995: red; 1996–2000: green).

well as quiet years. It contains no data gaps and is well documented (www.scecdc.scec.org). No further contraints (e.g. magnitude or depth restrictions) were imposed on the data and no declustering (e.g. Reasenberg, 1985) was attempted. All conventional declustering methods include arbitrary parameters, such as a pre-defined space-time window (the so-called aftershock window) and it is easy to over-decluster or under-decluster the catalogue. Linking the removal of dependent earthquakes with a large event under a space-time distance threshold does not remedy the situation (e.g. Zhuang et al., 2001). Avoiding declustering thus avoids the introduction of free parameters into the current analysis. Furthermore, spatio-temporal PCA may be regarded as a promising approach to declustering (in the sense of pattern separation) by itself, as will be seen below. More generally, PCA is to be tested on the earthquake process as a whole.

The vertical extent of the data set (average hypocentre depth is 20 km) is negligible in comparison to the horizontal dimensions (roughly 560 km $\times$ 450 km). The data may thus be regarded as essentially two-dimensional. In addition, the depth location error is, on average, two times larger than the horizontal one. Horizontal accuracy is primarily better than 5 km in the selected data. The following analysis will thus be constrained to an analysis of the spatio-temporal evolution of epicentre patterns. The data is shown in Fig. 2, where earthquakes of magnitude $M \geq 5.8$ are highlighted. Temporal colour-coding illustrates that the distributed earthquake process may be viewed as a stationary process of background seismicity with superimposed fluctuations due to mainshock-aftershock sequences and swarms (colour-coding in four year intervals: 1984–1987: blue; 1988–1991: black; 1992–1995: red; 1996–2000: green). A plot of magnitude versus time is part of Fig. 6. Background activity is again equated with stationarity here, as well as below, in the sense of the "no-change"-component, as detailed in Sect. 2.2. Seismologically speaking, background earthquakes are those which have no causal relation with other earthquakes, in contrast to events such as fore- and aftershocks, swarms and multiplets. While it is frequently impossible to discriminate between these events, many authors in fact agree that the background seismicity process should be regarded as a (Poissonian) stationary process (e.g. Wyss and Toya, 2000).

4 Application

Conducting a principal components analysis of epicentre distributions first requires turning the earthquake catalogue into vectorial data, i.e. a raster data set or matrix of spatial resolution Δl. For the spatio-temporal analysis, an additional separation into time slices of equal duration Δt is required (see Fig. 1 and Eq. 8).

In addition to the selection of a feasible space-time volume, i.e. the overall geographical bounds of the area of interest (here: $\Delta X = 115°$ to 121° lon W, $\Delta Y = 32°$ to 36° lat N) and the total duration ΔT (here: 17 years), the choices of cell size Δl and time slice duration Δt represent the only free

parameters in this analysis. This is in stark contrast to other methods dealing with the quantitative description of seismicity patterns.

In summary, the time slice observables $\mathbf{S}_i$ (see also Appendix) result as a function of the parameters

$$\{\Delta X, \Delta Y, \Delta T, \Delta l, \Delta t, t_0, [\Delta Z, \Delta M, \ldots]\}, \quad (9)$$

where t_0 denotes a possible time shift of the first time slice against the absolute beginning of the observation. ΔZ and ΔM are optional restrictions to a certain depth or magnitude band. Further constraints are possible depending on the purpose of the analysis. Most parameters may simply be chosen according to the purpose of the analysis (e.g. short- vs. long-term), but numerical contraints also exist in that, for example, the choice of Δl and Δt is a compromise between resolution and a sufficient population of the grid cells. Δt also restricts the ability to detect short-term transient changes or periodic signals (Gao et al. (2000) present an example for annually modulated seismicity rates). The time shift t_0 may become crucial in the case of fluctuations whose lifetime or onset should not be spread over adjacent time slices or if one wants a time window to end prior to the occurence of a large event in order to detect possible precursors (see Tiampo et al. (2000) for such an application).

After the discretisation according to Eq. 9 has been decided, the data is projected into Cartesian coordinates and a grid is formed by determining the cumulative number of events per cell, thus obtaining the spatial distribution of earthquake rates for each time interval. Due to the fact that the results of the PCA are area-weighted, it is important to use an equal area projection before the grid is formed (Eastman et al., 1994). In the present study, Δl was chosen to be 5 km, according to the arguments in Sect. 3, thus resulting in a grid of 112×90 cells. Δt was set to one year. None of the optional constraints were applied. It should be noted, however, that despite the neglect of event magnitude, the largest part of the event count is contributed by small events, as they are potentially more frequent than larger events, as stated by the Gutenberg-Richter law (Gutenberg and Richter, 1954). Thus, effectively a study of microseismicity patterns is conducted.

"Empty" cells in the raster data, i.e. the inherent sparsity of epicentre distributions due to fractal clustering (e.g. Goltz, 1997), pose no problem for the PCA. If an empty cell becomes populated at a later time for example, this represents real change; if a cell remains permanently empty, this is equivalent to "no change" and the region will automatically show up in the first component. Remember that effects showing up in a particular component will have been removed from all subsequent components.

As has been stated in Sect. 2.2, the stationary constituent will only show up in the first component if it contributes to the dominant portion of overall variance. Due to the fractal nature of earthquake (frequency) distribution in space and time (e.g. Hirabayashi et al., 1992; Goltz, 1997), i.e. the existence of singular values, this requirement is not fulfilled for raw seismicity data. The occurrence of large events with

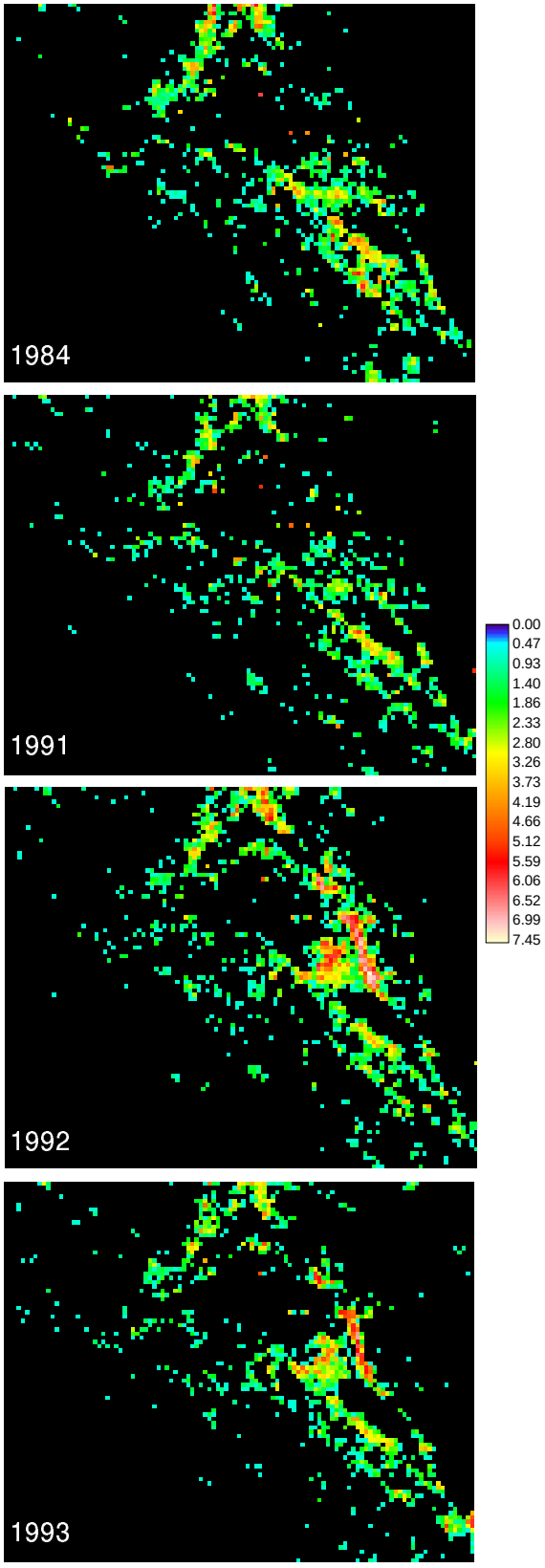

Fig. 3. Examples of log-transformed time slices $\mathbf{S}_i$ representing seismicity rates of 1984 (quiet), and 1991 to 1993, bracketing the Landers event in June 1992.

numerous aftershocks produces too much variance. A statistically feasible way out is to log-transform the data. Taking

$$S_{ij} \rightarrow \log(1 + S_{ij}), \tag{10}$$

with S_{ij} representing the elements of $\mathbf{S}_i$, additionally avoids having to manually deal with unpopulated areas. In addition to statistical reasons, log-transforming the data may also be prompted geophysically since the well-known effects of the clustering of seismicity are all of the power-law type, including the Omori-law $R(t) \approx 1/(t_E + t)^{\rho}$ for the decay in the rate of aftershocks after the main event at time t_E (Omori, 1894).

Standardised PCA (see Sect. 2.1, Eq. 3) must be used when the original observables possess different physical units, for example, but it is also preferable in the case of change detection (e.g. Eastman and Fulk (1993) and references therein), especially when the spatial patterns are to be emphasised (Wallace and Gutzler, 1981).

Figure 3 gives examples of time slices $\mathbf{S}_i$, with the log-transformed $\mathbf{S}_{1984}$ representative of a seismically quiet year, and $\mathbf{S}_{1991}$ through $\mathbf{S}_{1993}$ bracketing the $M = 7.3$ Landers earthquake on 28 June 1992. The colour-coding of all slices has been scaled to the range of $\mathbf{S}_{1992}$ for easier comparison.

5 Results

Figure 4 is a plot of the eigenvalues λ_i as obtained by solving Eq. 6. Such a plot is also called a scree plot, in the analogy to the accumulations of broken rock on steep slopes. The values have been expressed as percent total variance explained (see Sect. 2.1) according to

$$\lambda_i \rightarrow \lambda_i \times 100 / \sum_{k=1}^{17} \lambda_k. \tag{11}$$

The rapid decline in the size of the eigenvalues indicates a high degree of correlation between time slices and warrants that statistically useful results will be obtained by the transformation. Component $\mathbf{C}_1$ will have captured $\approx 61.5\%$ of the total information contained in the spatio-temporal seismicity patterns observed, $\mathbf{C}_2$ will have captured about 10.7%, $\mathbf{C}_3$ will have captured about 4.5%, etc. Examples of standardised components resulting from Eq. 4 are given in Fig. 5. As they result from a linear combination of earthquake rates occuring at different times (all time slices enter all components, weighted by the eigenvectors $\boldsymbol{\epsilon}_i$), they cannot simply be interpreted as "separated" earthquake rates in all cases. Spatio-temporally coherent patterns (see Sect. 2.2), however, allow for this interpretation. In addition, due to the coordinate rotation, negative values occur. While this problem could be easily corrected without an effect on the transformation by shifting the origin of the component space (Richards, 1986), it is the relative values within each component that define detail. Therefore, the output images in Fig. 5 were not post-processed except for setting the no-change areas of the change components to black. The components $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$ and $\mathbf{C}_6$ are shown, of which the latter explains about 2.6% of the total information. The principal component $\mathbf{C}_1$, which should represent stationary seismicity, is indeed very similar to $\mathbf{S}_{1984}$ (Fig. 3). None of the larger event sequences mentioned in Sect. 3 can be singled out. The maximum positive change seen in $\mathbf{C}_2$, the first component (and most dramatic) of change, can easily be associated with the occurence of the Landers earthquake (see Figs. 2 and 3). The effects due to the $M = 7.1$ Hector Mine event on 19 October 1999, and the $M = 6.7$ Northridge event on 17 January 1994, can also be identified to some extent. $\mathbf{C}_3$, however, the second component (and less dramatic) of change, very clearly singles out the change in seismcity rate due to the Hector Mine event. $\mathbf{C}_6$ as an example of a higher component shows subtle, spatially distributed change which cannot easily at first be associated with any individual cluster.

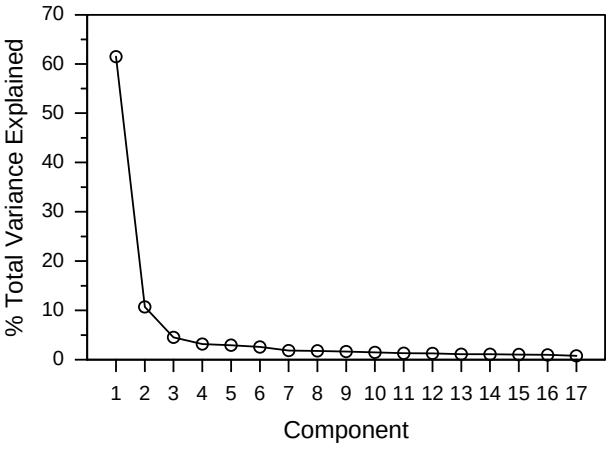

Fig. 4. Eigenvalues λ_i of the correlation matrix of yearly seismicity rates $\mathbf{S}_i$ as obtained from Eq. 6. The values have been expressed in terms of the percentage of total variance explained (see text). The rapid decline in the size of λ_i indicates a high correlation between time slices and warrants that statistically useful results will be obtained by a principal components transformation.

So far, the patterns decomposed by the transformation have only been identified by their spatial location. Figure 6 delivers the second crucial ingredient for a spatio-temporal analysis, the loadings graph introduced in Sect. 2.2. To aid interpretation, the lower part of the figure shows a plot of earthquake magnitude ($M \geq 4.0$) versus time. The persistent high correlation of $\mathbf{C}_1$ with seismicity patterns throughout all years ($\bar{r} = 0.78$) nicely confirms that the principal component indeed captures activity patterns common to all years, i.e. background seismicity. It also shows that the major element of variability in (log-transformed) earthquake rate data is that which occurs spatially. The latter is also expected due to the standardisation of $\mathbf{S}_i$. Component 2 shows a strong, negative correlation with seismicity rates up to $\mathbf{S}_{1991}$. From 1991 to 1992, a sudden reversal occurs and the correlation stays constantly positive until the end of observation in 2000. This behaviour not only confirms that $\mathbf{C}_2$ indeed captures the change due to the Landers earthquake, but also that it de-

scribes the unusually strong persistence of aftershock activity of this event. The loadings graph of $\mathbf{C}_3$ is somewhat similar to that of $\mathbf{C}_2$ in that the correlation jumps from slightly negative to significantly positive during and after the Hector Mine event. $\mathbf{C}_6$ shows gradual change. Namely, r increases monotonously from -0.37 in 1987 to +0.38 in 1991. Afterwards, correlation drops rapidly. Taken together with the component image, i.e. the distributed nature or spatial incoherence of this change, one might be led to interpret this component as a gradual change in seismic activity throughout the Southern California area, possibly a precursor to the Landers event. On the other hand, the $M = 5.8$ Sierra Madre earthquake happened on 28 June 1991 at the location of the maximum in $\mathbf{C}_6$. 1991 is also the time of maximum positive correlation of $\mathbf{C}_6$ with any time slice. The fairly small aftershock sequence of this event would explain the rapid declire in r. A more localised study would clarify this occurence. The example demonstrates that interesting information is contained in higher components which are usually disregarded in standard applications of PCA. It should be noted, in this regard, that the covariance and correlation matrices are measures of global variability; localised anomalies may, therefore, show up in high components, not necessarily in a low component. In addition, large magnitude fluctuations of small spatial extent are equivalent to small size fluctuations of large spatial extent, i.e. it is equally probable that they show up in higher components.

While not shown here, temporal randomisation of the earthquake catalogue (see Tiampo et al., 2000) leads to the disappearance of all coherence in the components and their loadings. This is similar to results obtained from raw seismicity rates. Randomisation of $\mathbf{S}_i$ such that the shape of patterns is preserved, i.e. keeping the binary pattern of activity, does not have such a pronounced effect (see Sect. 4 for reasons).

6 Conclusions and outlook

Spatio-temporal standardised principal components analysis has been successfully applied to earthquake rate evolution in the Southern California area. It was possible to separate seismicity into a pattern of background activity and patterns of change. Dominant earthquake events were unambiguously identified as coherent structures in the respective PCA components. Furthermore, the example of the Sierra Madre event (component 6) illustrates that comparatively subtle effects may be isolated as well. The gradual pre-seismic increase in loading of this component hints at possible precursory qualities and warrants further research in this direction. PCA as detailed here thus may be regarded as the first method of abstraction and quantification of significant spatio-temporal change effects in seismicity. So far, characteristic seismicity patterns had to be described qualitatively and decomposition of the space-time coupled earthquake process was arbitrary. The current approach possesses no free parameters except in

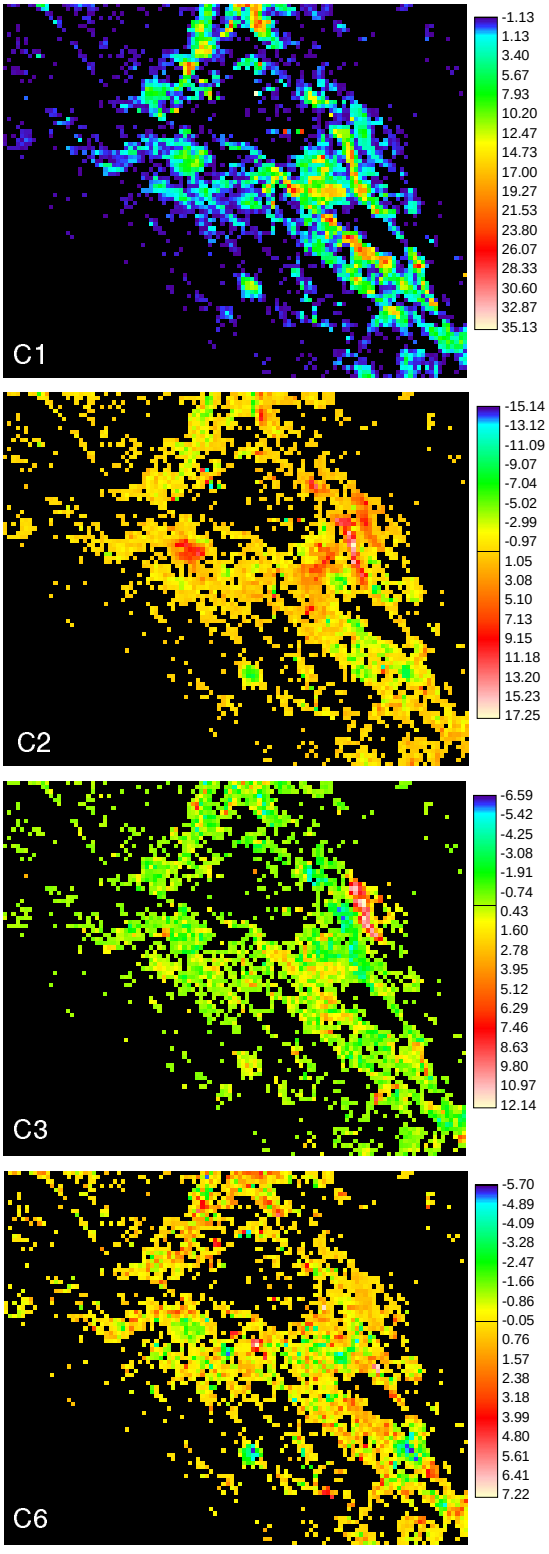

Fig. 5. Examples of components $\mathbf{C}_i$, the new variables after the principal components transformation has been applied to the log-transformed earthquake rates $\mathbf{S}_i$. The principal component $\mathbf{C}_1$, which captures background seismicity, and some higher components representing change are shown. No-change areas have been manually blackened. See text for discussion of patterns.

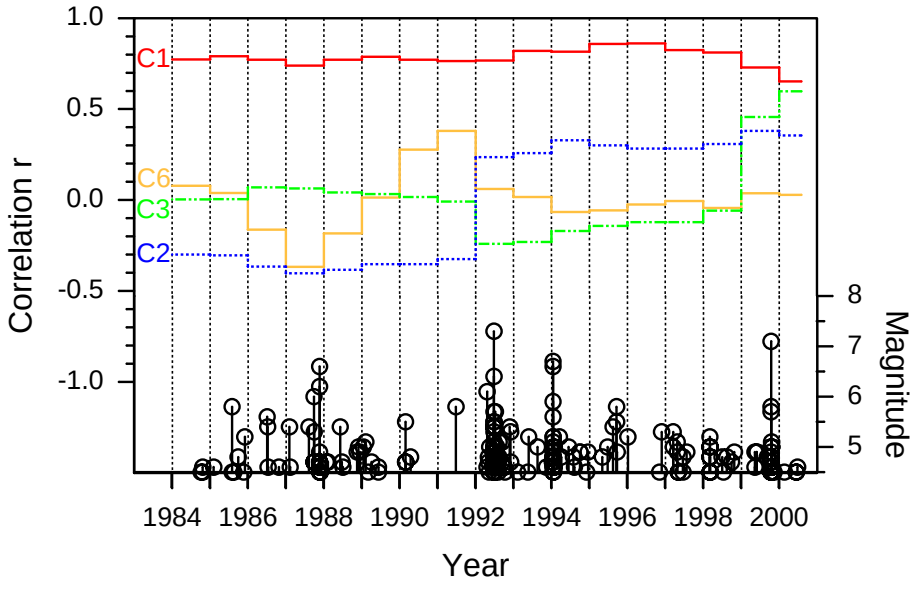

Fig. 6. Loadings graphs for components $\mathbf{C}_1$ through $\mathbf{C}_3$ and $\mathbf{C}_6$, giving the degree of correlation between these components and all time slices $\mathbf{S}_i$. These graphs allow for the determination of the onset and duration (i.e. temporal coherence) of the change patterns captured in the respective components. Earthquake magnitude versus time has been added to aid interpretation. The consistent high correlation of $\mathbf{C}_1$ confirms that this component indeed captures background seismicity. See text for a discussion of the change component loadings.

the preparation of the data for analysis. Those parameters, however, are not critical.

The approach of this paper opens up new possibilities with respect to several applications: for earthquake declustering, given the objective decomposition of seismicity patterns, for the detection of periodically or otherwise (antropogenically) induced seismicity, by studying the spectral contents of the loadings graph, and for the removal of effects due to the seismological network (see Eastman and Fulk (1993) for an example of the removal of sensor effects by PCA). The method should also prove valuable in the comparative study of seismogenesis in tectonically different settings as the spatio-temporal evolution and dependance of event sequences can be quantified.

The recent developments in the application of the critical point concept (e.g. Sornette, 2000, for a comprehensive general introduction, including applications to earthquakes) to earthquake dynamics (e.g. Jaumé and Sykes, 1999; Hainzl and Zöller, 2001) crucially depend on the choice of the critical region. The mechanism of critical point dynamics predicts a long-range correlated stress field with exponentially accelerating cumulative seismic energy/moment release (e.g. Bowman et al., 1998) and diverging correlation length (e.g. Zöller et al., 2001) in this region. All tests of this hypothesis either define the critical region arbitrarily to some extent or require a priori knowledge of the fault (segment) to rupture and/or determine it's size as a result of an optimisation with respect to what one is seeking. PCA based on earthquake rates might not only help to identify large coherent patterns which are indicative of the critical region (by incorporating magnitude thresholds), but may of course also be applied directly to cumulative energy [Benioff strain, (e.g. Bufe et al., 1994)] release or maximum magnitude.

Also within the context of earthquake prediction and seismic hazard assessment lies the important question of the physical interpretation of PCA components. While the first component may always be regarded as background activity, interpretation of the higher components is not so simple. Component patterns are constructed to be optimal in the sense of variance and covariance (as in joint variance), not in the sense of coherence (physical connections) or maximum correlation (e.g. Chen and Harr, 1993). While the issue of coherence can be overcome by utilising standardised PCA, there is thus no guarantee that the components tell us something about the structure of an underlying continuous process from which $\mathbf{S}_i$ is sampled. In fact, only physical modes which act independently and with orthogonal patterns (normal modes) can be reproduced in the components (North, 1984). Von Storch and Navarra (1995) discuss these issues in the context of climatology (where PCA components are termed empirical orthogonal functions (EOFs)).

Tiampo et al. (2000) present results of an eigenanalysis of Southern Californian seismicity from 1980 to 1991. The dominant emerging coherent patterns identified in the present work (i.e. Landers, Northridge, Hector Mine) are also identified indirectly by these authors. Under the assumption that the components do represent physical modes (see also the work by Rundle et al., 2000), Tiampo et al. (2000) regard the eigenvectors of the correlation matrix as the square root of a probability which is directly related to the probability of the occurrence of a future earthquake. Correlation of areas of increased probability with the later occurence of large earthquakes yields some cases of success, but the fact that increased probabilities appear and vanish without occurrence of a major event as well as the occurrence of large earthquakes without prior increased probabilities may be taken as an argument that in fact, seismicity cannot be decomposed completely with the constraint of orthogonality. Rather, it appears that some earthquake sequences are amenable to orthogonal decomposition, while others are not. Qualitatively,

the most striking result from Tiampo et al. (2000) is the emergence of coherent structures around the later epicentre of, for example, Landers despite the fact that the data analysed ended in 1991. It is regrettable, however, that the authors did not take a long time series approach and that the loadings information was subsequently not obtained.

Dropping the constraint of orthogonality by allowing an oblique transformation of the original data (oblique PCA) as well as generalising the notion of dependence between variables might improve the results of seismicity eigenanalysis. Independent Components Analysis (ICA), for example, goes beyond PCA by finding a linear transformation that renders the components independent (in the sense of minimum mutual information) rather than just uncorrelated (e.g. Comon, 1994).

In addition to addressing the points mentioned above, future plans include a realisation of a significant number of case studies in different areas, including the explicit search for precursory patterns with an assessment of their statistical significance and the variation of the parameters relevant in the preparation of $\mathbf{S}_i$. By restricting the area of interest to a smaller one than the one shown in this paper might not only facilitate the analysis of the results, but also allow for an extension to three-dimensional observables by incorporating earthquake depth information. Finally, the issue of delineation and quantification of areas of significant change in the component images has not yet been addressed in this paper.

Appendix Using a vector to describe multi-dimensional data

In digital satellite remote sensing, input data for the the PCA consists of several rasterised two-dimensional images, i.e. the data is multi-dimensional. The (brightness) value of each raster cell or pixel represents the measured value at the spatial location of that pixel. It is convenient to construct a vector space with as many dimensions as there are images carrying information for each pixel. One may then simply plot the complete information for each pixel from all images as a point in this vector space. In other words, the components of the pixel vectors are the individual measurements in each image.

The above example from satellite remote sensing is directly applicable to the data analysed in this paper. Earthquake rates in a grid correspond to "pixel values", which are temporal windows (time slices) of earthquake activity to "images". The vector $\boldsymbol{x}$ from Eq. 1 thus represents the data as a whole; $\boldsymbol{x}_i$ (Eq. 2) represents the set of earthquake counts for each grid cell at all times. Referring to the comments about Eq. 2 in Sect. 2.1, it is $m = 2$, $n = 10\,080$ and $l = 17$ in the case presented here. The spatial distributions of earthquake rates during a particular time window are designated by $\mathbf{S}_i$ (Sect. 4). $\mathbf{S}_i$, as well as the components $\mathbf{C}_i$ (Sect. 2.1), have been denoted as matrices here, due to the two-dimensional nature of epicentre distributions.

Acknowledgements. The earthquake data was obtained from the Southern California Earthquake Data Center (SCEDC) which processes data from the Southern California Seismic Network (SCSN) which is operated jointly by the Seismological Laboratory at Caltech and the U. S. Geological Survey, Pasadena. Figure 2 was produced using GMT (Wessel and Smith, 1991). A travel grant from the Deutsche Forschungsgemeinschaft enabled me to attend the 31st International Geological Congress to present and discuss initial results. Many thanks to W. Rabbel for useful discussions. This publication is contribution no. 3 of the Sonderforschungsbereich 574 "Volatiles and Fluids in Subduction Zones: Climate Feedback and Trigger Mechanisms for Natural Disasters" at the University of Kiel.

References

Bowman, D. D., Oullion, G., Sammis, C. G., Sornette, A., and Sornette, D.: An observational test of the critical earthquake concept, J. Geophys. Res., 103, 24359–24372, 1998.

Brehm, D. J. and Braile, L. W.: Intermediate-term earthquake prediction using precursory events in the New Madrid Seismic Zone, Bull. Seism. Soc. Am., 88, 564–580, 1998.

Bufe, C. G. and Varnes, D. J.: Predictive modeling of the seismic cycle of the greater San Francisco Bay region, J. Geophys. Res., 98, 9871–9883, 1993.

Bufe, C. G., Nishenko, S. P. and Varnes, D. J.: Seismicity trends and potential for for large earthquakes in the Alaska-Aleutian region, Pageoph, 142, 83–99, 1994.

Byrne, E. F., Crapper, P. F., and Mayo, K. K.: Monitoring landcover change by principal components analysis of multitemporal Landsat data, Remote Sensing of Environment, 10, 175–184, 1980.

Chen, J. M. and Harr, P. A.: Interpretation of extended empirical orthogonal function (EEOF) analysis, Mon. Wea. Rev., 121, 2631–2636, 1993.

Comon, P.: Independent component analysis: A new concept?, Signal Processing, 36, 287–314, 1994.

Dodge, D. A., Beroza, G. C., and Ellsworth, W. L.: Detailed observations of California foreshock sequences: Implications for the earthquake initiation process, J. Geophys. Res., 101, 22371–22392, 1996.

Eastman, J. R. and Fulk, M.: Long sequence time series evaluation using standardized principal components, Photogrammetric Engineering & Remote Sensing, 8, 1307–1312, 1993.

Eastman, J. R., McKendry, J., and Fulk, M.: Change and Time Series Analysis, 2nd edition, UNITAR Explorations in GIS Technology, UNITAR, Geneva, 1994.

Freire, S. L. M. and Ulrych, T. J.: Application of singular value decomposition to vertical seismic profiling, Geophysics, 53, 778–785, 1988.

Fröhlich, C.: Aftershocks and temporal clustering of deep earthquakes, J. Geophys. Res., 92, 13944–13956, 1987.

Gao, S. S., Silver, P. G., Linde, A. T., and Sacks, I. S.: Annual modulation of triggered seismicity following the 1992 Landers earthquake in California, Nature, 406, 500–504, 2000.

Gershenfeld, N.: The Nature of Mathematical Modeling, Cambridge Univ. Press, Cambridge, 1999.

Goltz, C.: Fractal and Chaotic Properties of Earthquakes, Springer, Berlin, 1997.

Golub, G. H. and Van Loan, C. F.: Matrix Computations, 3rd edition, John Hopkins University Press, Baltimore, MD, 1996.

Gutenberg, B. and Richter, C. F.: Seismicity of the Earth and Associated Phenomena, 2nd edition, Princeton Univ. Press, Princeton, 1954.

Haberman, R. E.: Precursory seismic patterns: stalking the mature seismic gap, in: Earthquake Prediction: An International Review, AGU Monograph, AGU, Washington, 29–42, 1981.

Hainzl, S. and Zöller, G.: The role of disorder and stress concentration in nonconservative fault systems, Physica A, 294, 67–84, 2001.

Hirabayashi, T., Ito, K., and Yoshii, T., Multifractal analysis of earthquakes, Pageoph, 138, 591–610, 1992.

Hotelling, H.: Analysis of a complex of statistical variables with principal components, J. Educ. Psych., 24, 417–441, 1933.

Howarth, P. J. and Boasson, E.: Landsat digital enhancements for change detection in urban environments, Remote Sensing of Environment, 13, 149–1160, 1983.

Jackson, J. E.: A User's Guide to Principal Components, John Wiley and Sons, Inc., New York, 1991.

Jaumé, S. C. and Sykes, L. R.: Evolving towards a critical point: A review of accelerating seismic moment/energy release prior to large and great earthquakes, Pageoph, 155, 279–306, 1999.

Johnston, R. J.: Multivariate Statistical Analysis in Geography, Longman, New York, 1980.

Lillesand, T. M. and Kiefer, R. W.: Remote Sensing and Image Interpretation, 3rd edition, John Wiley & Sons, Inc., New York, 1994.

Loève, M.: Probability Theory. D. van Nostrand Co., Princton, NJ, 1955.

Mogi, K.: Some features of recent seismic activity in and near Japan, (2) Activity before and after large earthquakes Bull. Earthquake Res. Inst., Tokyo Univ., 47, 395–417, 1969.

North, G. R.: Empirical orthogonal functions and normal modes, J. Atmos. Sci., 41, 879–887, 1984.

Omori, F.: On the aftershocks of earthquakes, J. Coll. Sci. Imp. Univ. Tokyo, 7, 111–200, 1894.

Reasenberg, P. A.: Second-order moment of Central California seismicity, 1969–1982, J. Geophys. Res., 90, 5479–5495, 1985.

Richards, J. A. and Milne, A. K.: Thematic mapping from multitemporal image data using the principal components analysis, Remote Sensing of Environment, 16, 35–46, 1983.

Richards, J. A.: Remote Sensing Digital Image Analysis. Springer, Berlin, 1986.

Rundle, J. B.: A physical model for earthquakes: 2. Application to southern California, J. Geophys. Res., 93, 6255–6274, 1988.

Rundle, J. B., Klein, W., Tiampo, K., and Gross, S.: Linear pattern dynamics in nonlinear threshold systems, Phys. Rev. E, 61, 2418–2431, 2000.

Singh, A.: Digital change detection techniques using remotely sensed data, Int. J. Remote Sensing, 6, 989–1003, 1989.

Singh, A. and Harrison, A.: Standardized principal components, Int. J. Remote Sensing, 6, 883–896, 1985.

Swan, F. H., Schwartz, D. P., and Cluff, L. S.: Recurrence of moderate to large magnitude earthquakes produced by surface faulting on the Wasatch fault zone, Utah, Bull. Seism. Soc. Am., 70, 1431–1462, 1980.

Tiampo, K. F., Rundle, J. B., McGinnis, S., and Gross, S. J.: Observation of systematic variations in non-local seismicity patterns from Southern California, in: GeoComplexity and the Physics of Earthquakes, AGU, Washington, 211–218, 2000.

Sornette, D.: Critical Phenomena in Natural Sciences, Chaos, Fractals, Selforganization and Disorder: Concepts and Tools, Springer, Berlin, 2000.

von Storch, H. and Navarra, A.: (Eds), Analysis of Climate Variability, Springer, Berlin, 1995.

Wallace, J. M. and Gutzler, D. S.: Teleconnection in the geopotential height field during the northern hemisphere winter, Mon. Wea. Rev., 109, 784–812, 1981.

Wessel, P. and Smith, W. H. F.: Free software helps map and display data, Eos Trans. AGU, 72, 441, 445–446, 1991.

Wyss, M., Shimazaki, K., and Urabe, T.: Quantitative mapping of a precursory seismic quiescence to the Izu-Oshima, 1990 (M6. 5) earthquake, Japan, Geophys. Jour. Int., 127, 735–743, 1996.

Wyss, M., Shimazaki, K., and Ito, A.: (Eds), Seismicity Patterns, their Statistical Significance and Physical Meaning, Birkhäuser, Basel, 1999.

Wyss, M. and Toya, Y.: Is background seismicity produced at a stationary Poissonian rate?, Bull. Seism. Soc. Am., 90, 1174–1187, 2000.

Zhuang, J., Ogata, Y., and Vere-Jones, D.: Stochastic declustering of space-time earthquake occurrences, submitted to Journal of American Statistical Society, 2001.

Zöller, G., Hainzl, S., and Kurths, J.: Observation of growing correlation length as an indicator for critical point behavior prior to large earthquakes, J. Geophys. Res., 106, 2167–2176, 2001.