Using Eigenpattern Analysis to Constrain Numerical Fault Models - Applications to Southern California

K. F. Tiampo[(1)], J. B. Rundle[(2)], S. McGinnis[(3)], W. Klein[(4)], and Y. Ben-Zion[(5)]

(1) CIRES, University of Colorado, Boulder, CO 80309, USA (email: kristy@cires.colorado.edu; phone: +1 303-492-4779).
(2) Dept. of Physics, Colorado Center for Chaos & Complexity, CIRES, University of Colorado, Boulder, CO, 80309, USA, and Distinguished Visiting Scientist, Jet Propulsion Laboratory, Pasadena, CA, 91125, USA (email: rundle@cires.colorado.edu; phone: +1 303-492-5642).
(3) CIRES, University of Colorado, Boulder, CO 80309, USA (email: sethmc@turcotte.colorado.edu; phone: +1 303-492-4779).
(4) Dept. of Physics, Boston University, Boston, MA 02215, USA and Center for Nonlinear Science, Los Alamos National Laboratory, Los Alamos, NM 87545, USA (email: klein@buphy.edu).
(5) Dept. of Earth Sciences, University of Southern California, Boulder, CO (email: benzion@terra.usc.edu ; phone: +1 (213) 740-6734).

Abstract

Earthquake fault systems are now thought to be an example of a complex nonlinear system (Bak, 1987; Rundle, 1995). The spatial and temporal system complexity of this system translates into a similar complexity in the surface expression of the underlying physics, including deformation and seismicity. Our research suggests that a new pattern dynamic methodology can be used to define a unique, finite set of seismicity or deformation patterns for a given fault system (Tiampo et al., 2002). Similar in nature to the empirical orthogonal functions historically employed in the analysis of atmospheric and oceanographic phenomena (Preisendorfer, 1988), the method derives the eigenvalues and eigenstates from the diagonalization of the correlation matrix using a Karhunen-Loeve expansion (Fukunaga, 1990, Rundle, et al., 2000, TIampo et al., 2002). This Karhunen-Loeve expansion (KLE) technique may be used to help determine the important modes in both time and space for southern California seismicity as well as deformation (GPS) data. These modes potentially include such time dependent signals as plate velocities, viscoelasticity, and seasonal effects. This can be used to better model geophysical signals of interest such as coseismic deformation, viscoelastic

effects, and creep. These, in turn, can be used for both model verification in large-scale numerical simulations of southern California and error analysis of remote sensing techniques such as InSar.

Introduction

Data assimilation is the process by which observational data is incorporated into models to set their parameters, and to tune them in real time as new data becomes available. The result of the data assimilation process is a model that is maximally consistent with the observed data, producing a model that is useful in ensemble forecasting. With the idea that the state of the model follows an evolutionary path through state space as time progresses, and that observations can be used to periodically adjust model parameters, so that the model path is as close as possible to the path represented by the observed system, a *cost function*, or *fitness function*, defines the misfit between the model path through time and the path represented by the observed data. Model tuning occurs by computation of a gradient vector in state space, which specifies the direction in which corrections to the model must be applied in order to return the model evolution towards the path defined by the observations.

A major drawback to all these methods is that although geophysical dynamics can sometimes be considered quasi-linear, in general, they are not. However, earthquake dynamics are characterized by time intervals over which evolution of state occurs smoothly, but then sudden large jumps to a new state occur that produce large rearrangements in system state. Problems of this type have motivated recent searches for methods to assimilate data into models that have highly nonlinear, stochastic dynamics. At the present time, data assimilation into models such as Virtual_California is carried out via a *static* assimilation algorithm rather than a *dynamic* algorithm. One alternative, a genetic algorithm (GA), is a nonlinear adaptation procedure based on the mechanics of natural selection and genetics. A GA offers the advantage of being independent of the choice of starting model, and not requiring the linearization of the source function. For a geophysical problem, the GA evolves a vector of model parameters that optimizes the cost (fitness) function, producing fitter solutions in each new generation. The cost function is defined quantifying the fit of the simulation data to the historic data through time. The potential data types include earthquake occurrence time, location, and moment release, as well as surface deformation data obtained from InSAR, GPS, and other methods.

Karhunen-Loeve Expansion (KLE) Analysis

Pattern evolution and prediction in nonlinear systems is complicated by nonlinear interactions and noise, but understanding such patterns, which are simply the surface expression of the underlying dynamics, is critical to understanding and perhaps characterizing the physics which control the system. Rundle, et al., 2000, proposed a method of decomposing the complex spatial and temporal patterns that are the surface expression of the obscure dynamics of the physical system, into the orthonormal pattern eigenstates. This decomposition implicitly assumes that one is dealing with a process that is both Markov and stationary in time. The procedure involves constructing a correlation operator, $C(x_i,x_j)$, for the sites that contains the spatial relationship of slip events over time. $C(x_i,x_j)$ is decomposed into the orthonormal spatial eigenmodes for the nonlinear threshold system, e_j, and their associated time series, $a_j(t)$. These spatial-temporal pattern states can be used to reconstruct the primary modes of the system, with or without noise, and quantify their relative magnitude and importance. In addition, these primary modes can be used to characterize the underlying dynamics and the physical parameters such as stress levels and interactions that control the observable patterns of events. We propose to apply this technique to SCIGN data in order to determine the principal modes of deformation for the southern California fault system.

Similar to the empirical orthogonal function (EOF) technique developed by Preisendorfer, 1988, for the atmospheric sciences, the Karhunen-Loeve expansion is obtained from the p time series that record the deformation history at particular locations in space. Each time series, $y(x_s,t_i) = y_i^s$, $s = 1, \ldots p$, consists of n time steps, $i = 1, \ldots n$. The goal is to construct a time series for each of a large number of locations for a given short period of time. If, for example, the time interval was decimated into units of days, the result could be a time series of 365 time steps for every year of data, with values of deformation for that location at each time step. These time series are incorporated into a matrix, $\boldsymbol{T}$, consisting of time series of the same measurement for p different locations, i.e.

$$T = [\bar{y}_1, \bar{y}_2, \ldots \bar{y}_p] = \begin{bmatrix} y_1^1 & y_1^2 & \cdots & y_1^p \\ y_2^1 & y_2^2 & \cdots & y_2^p \\ \vdots & \vdots & \ddots & \vdots \\ y_n^1 & y_n^2 & \cdots & y_n^p \end{bmatrix}. \tag{1}$$

For analysis of SCIGN data, the values in the matrix $\boldsymbol{T}$ will consist of either strain or deformation measurements, horizontal or vertical. The covariance matrix, $S(x_i,x_j)$, for these events is formed by multiplying $\boldsymbol{T}$ by $\boldsymbol{T^T}$, where S is a p x p real, symmetric matrix. The covariance matrix, $S(x_i,x_j)$, is converted to a correlation operator, $C(x_i,x_j)$, by dividing each element of $S(x_i,x_j)$, by the variance of each time series, $y(x_i,t)$ and $y(x_j,t)$,

$$\sigma_p = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_k^p)^2}, \tag{2}$$

and

$$C = \begin{bmatrix} \frac{s_{11}}{\sigma_1\sigma_1} & \frac{s_{12}}{\sigma_1\sigma_2} & \cdots & \frac{s_{1p}}{\sigma_1\sigma_p} \\ \frac{s_{21}}{\sigma_2\sigma_1} & \frac{s_{22}}{\sigma_2\sigma_2} & \cdots & \frac{s_{2p}}{\sigma_2\sigma_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{s_{p1}}{\sigma_p\sigma_1} & \frac{s_{p2}}{\sigma_p\sigma_2} & \cdots & \frac{s_{pp}}{\sigma_p\sigma_p} \end{bmatrix}. \qquad (3)$$

This equal-time correlation operator, $C(x_i,x_j)$, is decomposed into its eigenvalues and eigenvectors in two parts. The first employs the trireduction technique to reduce the matrix ***C*** to a symmetric tridiagonal matrix, using a Householder reduction. The second part employs a ql algorithm to find the eigenvalues, λ_j^2, and eigenvectors, e_j, of the tridiagonal matrix (Press, et al., 1992). These eigenvectors, or eigenstates, are orthonormal basis vectors arranged in order of decreasing variance that reflect the spatial relationship of events in time. If one divides the corresponding eigenvalues, λ_j^2, by the sum of the eigenvalues, the result is that percent of the variance accounted for by that particular mode. We then reconstruct the time series associated with each location for each eigenstate by projecting the initial data back onto these basis vectors in what is called a principal component analysis (PCA) (Preisendorfer, 1988). These time dependent expansion coefficients, $a_j(t)$, which represent temporal eigenvectors, are reconstructed by multiplying the original data matrix by the eigenvectors, i.e.

$$a_j(t_i) = \vec{e}^{\,T} \cdot T = \sum_{s=1}^{p} e_j y_i^s, \qquad (4)$$

where $j, s = 1, \ldots p$ and $i = 1, \ldots n$. This eigenstate decomposition technique produces the orthonormal spatial eigenmodes for this nonlinear threshold system, e_j, and the associated principal component time series, $a_j(t)$. These principal component time series represent the signal associated with each particular eigenmode over time. For purposes of clarity, the spatial eigenvectors are designated KLE modes and the associated time series Principal Component (PC) vectors.

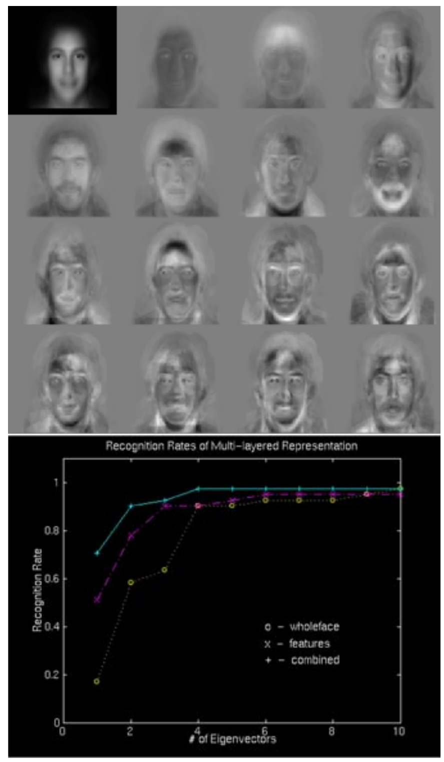

Figure 1: Sixteen standard eigenfaces from PC analysis of 128 face samples. Recognition rate as a function of the number of eigenfaces used for reconstruction (http://www-white.media.mit.edu/vismod)

The KL expansion, a linear decomposition technique in which a dynamical system is decomposed into a complete set of orthonormal subspaces, has been applied to a number of other complex nonlinear systems over the last fifty years, including the ocean-atmosphere interface, turbulence, meteorology, biometrics, statistics, and geophysics (Preisendorfer, 1988; Savage, 1988; Penland, 1989; Vautard and Ghil, 1989; Fukunaga, 1990; Penland and Sardeshmukh, 1995; Tiampo et al., 2002). Applied over sufficiently long time periods, a KL decomposition reveals the underlying correlations in the spatial pattern of the seismicity (Tiampo et al., 2002). Shown in Figure 1 are the results of a biometrics KL analysis, in which the decomposition is applied to a facial recognition system (Moghaddam et al., 1998).

Results and Discussion

Rundle, 2000, details successful application of these techniques to computer simulations of the complex fault systems. In particular, the method is extended to include an unequal-time correlation operator that can be used to forecast events in time; in much the same manner as EOF analysis is used to predict El Nino events in meteorology (Garcia and Penland, 1989; Penland, 1989; Preisendorfer, 1988). A more pertinent example to illustrate its application to SCIGN data can be found in the analysis of actual seismicity data for southern California. The data used is from the entire Caltech catalogue, obtained from the SCEC database, from 1932 through June of 1998. Relevant information consists

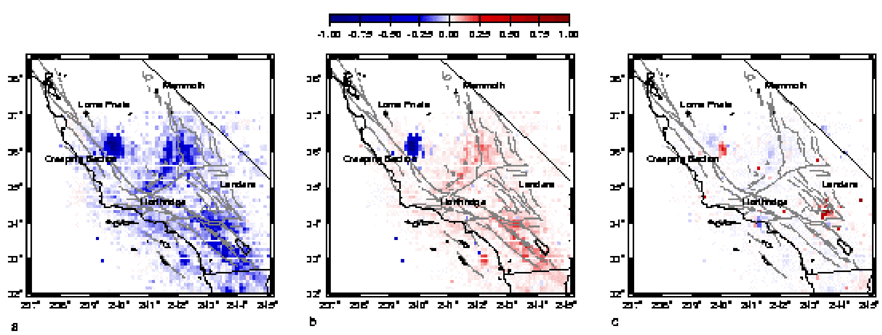

Figure 2: a) First KL mode, 1932-1991; b) Second KL eigenpattern and c) eighth KL eigenpattern. Note the appearance of correlations associated with the 1992 Landers event, not included in this data set.

of location in latitude and longitude, and the time that the event occurred. Each time step is given an initial value of 1.0, if one or more events occur in that time period. This is done for each location time series, after which the mean for each time series is removed from the data. A Karhunen-Loeve expansion (KLE) analysis was then performed on the entire data set (Tiampo, et al, 2002). Shown in Figure 2 are the results of a KL analysis for historic seismicity data from southern California for the period 1932 through 1991 (Tiampo et al., 2002).

The results in Figure 2 demonstrate the viability of applying this new pattern dynamics methodology to historical seismicity data. Actual data can be decomposed into its

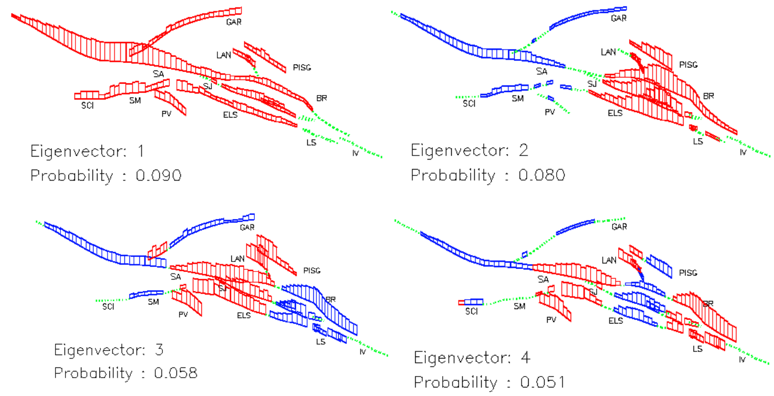

Figure 3: First four KL modes, Virtual California.

orthonormal basis functions, and the eigenstates derived from these analyses can be used to reconstruct the primary modes, either with or without the associated noise. Information about the primary modes of this pattern analysis can be used to model the underlying dynamics of the system; in particular the stress fields associated with the fault interactions. Figure 3 shows the first two KL decomposition modes for Virtual California, illustrating the success of the technique on the numerical simulations, and the possibilities for data assimilation.

Application of this equal-time correlation operator technique to coherent GPS data, specifically the SCIGN data array of southern California, can be used to characterize the underlying dynamics and the physical parameters such as stress levels and interactions which control the observable patterns of events. Modeling of the potential geophysical sources of the deformation can be accomplished in a variety of ways, most notably by employing an updated, more detailed version of the southern California fault patch model developed in the mid-1980s.

For each of the existing SCIGN stations, approximately 200 at this time, the appropriate strain or deformation time series was be identified, and a KLE analysis performed. For example, a potentially rewarding avenue could include the separation of vertical motions from horizontal motions, leading to the identification of modes whose surface expression has a large vertical component, such as viscoelastic response or seasonal hydrologic patterns. Another expected result are modes consisting of localized strain interactions between particular stations, signaling small-scale features previously unidentified in the data, such as creep events or blind thrust faults. In addition, there are several modes that one would expect to see and which could be better modeled without the interaction of additional modes and noise, as this method will allow for their separation. A partial list would include plate motions, coseismic deformation, groundwater or oil well extraction, creep events or `slow' earthquakes, blind thrust faults, viscoelastic response, local variations in strain rate, and seasonal hydrologic cycles.

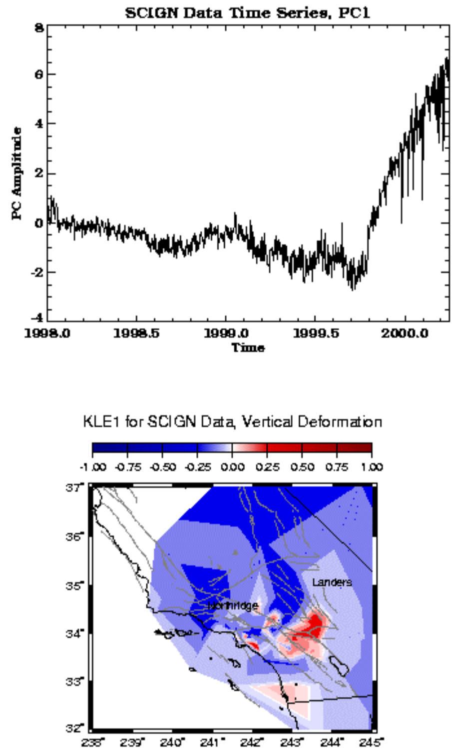

Figure 4: First KLE mode, SCIGN data. Top - PCA time series. Bottom – Spatial eigenmode.

A very simple analysis has been done as an example on vertical motions for stations, with data obtained directly from the SCIGN web site. Results for the first KLE mode is shown in Figure 4. Shown the time series for the first KL mode for the vertical data after 1998. Note the correlation through the Mojave desert which incorporates both the area of the Landers sequence of 1992 and the Hector Mine event of 1999. This corresponds to a jump in the associated time series, also shown in the attached figure, at the time of the Hector Mine event in the fall of 1999. Unexpected results include the correlations just inshore of the 1933 Long Beach earthquake, and the correlated increase in vertical motion near the location of the 1899 Cajon Pass event.

Finallly, Figure 5 shows both the GPS and the InSar deformation for one simulated earthquake from the Virtual California computer model. Note that the resolution is sufficient to incorporate into an inversion model for the purpose of data assimilation.

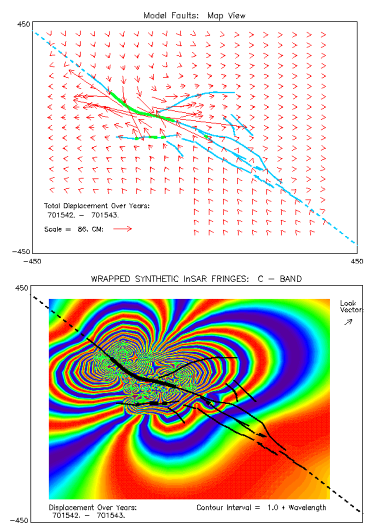

Figure 5: GPS type observations (left) and InSar type observations (left), corresponding to a simulated earthquake event.

Acknowledgments

Research by KFT was funded by NGT5-30025 (NASA). Research by JBR was funded by USDOE/OBES grant DE-FG03-95ER14499 (theory), and by NASA grant NAG5-5168 (simulations). Research by SM was funded by NASA Fellowship ESS-97-0110 to CIRES. Research by WK was supported by USDOE/OBES grant DE-FG02-95ER14498 and W-7405-ENG-6 at LANL. WK would also like to acknowledge the hospitality and support of CNLS at LANL.

References

[1]Bak, P., C. Tang, and K. Weisenfield. Self-organized criticality: An explanation of the 1/f noise. PRL, **59**, 1987.

[2]Fukunaga, K. Introduction to Statistical Pattern Recognition, Academic Press, N.Y., 1970.

[3]Garcia, A., and C. Penland. Fluctuating hydrodynamics and principal oscillation pattern analysis. J. Stat. Phys., **64**, 1991.

[4]Moghaddam, B., Wahid, W., and Pentland, A., 1998. Beyond eigenfaces: probabilistic matching for face recognition. Third IEEE Intl. Conf. on Automatic Face & Gesture Recognition, 1-6.

[5]Penland, C. Random forcing and forecasting using principal oscillation pattern analysis, Monthly Weather Review, **117**, 1989.

[6]Penland, C., and P. D. Sardeshmukh. The optimal growth of tropical sea surface temperature anomalies. J. Climate, **8**, 1999.

[7]Preisendorfer, R.W. Principle Component Analysis in Meteorology and Oceanography, ed. C. D. Mobley, Develop. Atm. Sci. 17 (Elsevier, Amsterdam, 1988).

[8]Press, W., S. Teukolosky, W. Vettering, and B. Flannery. Numerical Recipes in C: The Art of Scientific Computing, Cambridge University, Cambridge, 1992.

[9]Rundle, J.B. and W. Klein. New ideas about the physics of earthquakes. Rev. Geophys., **33**, Part 1, Suppl. S, 1995.

[10]Rundle, J.B., W. Klein, K.F. Tiampo, and S. Gross. Linear pattern dynamics of nonlinear threshold systems. Phys. Rev. E, **61**, 2000.

[11]Savage, J.C. Principal component analysis of geodetically measured deformation in Long Valley Caldera, eastern California, 1983-1987, JGR, **93**, 1988.

[12]The SCIGN Project Report to NSF, SCEC, October 1998.

[13]Tiampo, K.F., J.B. Rundle, S. McGinnis, S.J. Gross, and W. Klein. Observation of systematic variations in non-local seismicity patterns from southern California, *Geocomplexity and the Physics of Earthquakes*, AGU Monograph, 2000.

[14]Tiampo, K.F., J. B. Rundle, S. J. Gross, W. Klein, Eigenpatterns in southern California, JGR, accepted for publication, 2002.

[15]Vautard, R., and M. Ghil. Singular spectrum analysis in nonlinear dynamics, with applications to paleodynamic time series. Physica D, **35**, 1989.