

**Pattern Dynamics in Southern California  
Seismicity**

by

**Kristy French Tiampo**

B.S.C.E., Tufts University, 1983

M.S.C.E., Stanford University, 1984

A thesis submitted to the  
Faculty of the Graduate School of the  
University of Colorado in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
Geophysics Graduate Program  
Department of Geological Sciences  
2000

This thesis entitled:  
Pattern Dynamics in Southern California Seismicity  
written by Kristy French Tiampo  
has been approved for the Department of Geological Sciences

---

John B. Rundle

---

Douglas S. Robertson

Date \_\_\_\_\_

The final copy of this thesis has been examined by the signatories, and we find  
that both the content and the form meet acceptable presentation standards of  
scholarly work in the above mentioned discipline.

Tiampo, Kristy French (Ph.D., Geophysics)

Pattern Dynamics in Southern California Seismicity

Thesis directed by Professor John B. Rundle

While a variety of innovative, nonlinear analysis techniques are applied to measured data in other areas of physics, they rarely are employed in the field of solid earth geophysics. Yet a wide variety of geophysical data already exists which could be studied using these methods. The historical earthquake record, for example, while not complete, spans hundreds to thousands of years of human history. As a result, earthquakes in seismically active regions are known to be associated with a variety of spatial and temporal patterns. Examples include precursory quiescence or enhanced activity, aftershocks, characteristic earthquakes, and seismic clustering. A spatially coherent, uniformly high level of stress on a fault is a necessary precondition for the occurrence of a large earthquake. Recently, several groups have found that spatial coherence in the stress field is reflected in a similar coherence in the seismic activity. Here I employ a Karhunen-Loeve decomposition method to systematically quantify the spatial and temporal characteristics of patterns in historic seismicity in southern California based upon these underlying correlations. This Karhunen-Loeve decomposition expands the seismicity into its orthonormal eigenvectors and eigenvalues, isolating the important normal mode patterns. The demonstrated results strongly support the view that seismic activity is highly correlated across many space and time scales within large volumes of the earth's crust.

I then identify systematic space-time variations in the seismicity from southern California using a new technique. This procedure is based upon the idea that seismic activity corresponds geometrically to the rotation of a pattern state vector in the high-dimensional correlation space spanned by the eigenvectors of a correlation operator. Using this new technique it is possible to isolate emergent regions of coherent, correlated seismicity. Analysis of data taken only up to December 31, 1991 reveals that the appearance of these correlated regions is often associated with the future occurrence of major earthquakes in the same areas. These major earthquakes include the 1992 Landers, the 1994 Northridge, and the 1999 Hector Mine events.

The space-time patterns in the seismic activity, identified by these two phase dynamical applications, directly reflect the existence of stress correlations. The success of both the Karhunen-Loeve decomposition on the historic seismicity in southern California, and the identification of anomalous seismicity or precursory quiescence prior to large earthquakes, supports recent work that suggests that correlated structure exists in the underlying stress and strain fields, a necessary precondition for the occurrence of large earthquakes.

To My Family  
Jim, Matt, Steven, and Carmen

## Acknowledgements

"It is good to have an end to journey towards; but it is the journey that matters in the end." U.K. LeGuin

There are so many people who have helped me over the last few years that it seems impossible to thank them all. However, I am going to try, once in my life, to tell them all how much they have meant to me.

First, it is my pleasure to thank my advisor, John Rundle, for his support, inspiration and patience. The work I've been fortunate enough to be a part of for these last five years would not have been possible without him - he always knew the right path to take, and had faith enough in me to let me travel it.

I would like to thank Susan Avery, Roger Bilham, Lang Farmer, Kristine Larson, Doug Robertson, Anne Sheehan, and John Wahr for serving on my committee throughout my graduate work at CU. Their advice, instruction and criticism made both the work and the student better.

I would also like to thank Jose Fernandez for his willingness to include me in a number of projects related to both our work. His patience and advice were invaluable; the experience expanded both my abilities and my confidence.

My fellow graduate students, colleagues, and friends, Marian Anghel, Joydeep Bhattacharyya, Eric Blough, Peter Furey, Hersh Gilbert, Michelle Hofton, Jorge Simoes de Sa Martins, Paul Vincent, Ting-To Yu, and others, with their advice and willingness to listen, made this journey both lighter and more enjoyable. I must also thank the computing staff at CIRES for their part in producing this work, and for being patient with my, at times, frenzied requests for help. In particular, I want to thank Susanna Gross, Seth McGinnis, and Seth Veitzer, for if they don't know it already, their assistance and friendship have proven invaluable at more than one point along the way.

Itziar, Sandra and Eugenia cared for my children, in love and security - without which this journey would have been neither right nor possible. Itziar, in particular, I owe a huge debt - her wholehearted devotion and outstanding character gave me the security I needed at a critical time, the start of this journey.

So many other friends and family provided help and support, I can only name a few. Clem and Pilar, Janet and David, Mike and Louise, Roe and Sharon, Ivan, Sylvia, Jill, Stephanie, Rhonda, and others. I appreciated not just their

support, but their interest in something so important to me. No journey is made without good friends - I count myself lucky in you.

My family, from its oldest members to its youngest, always believed in me. I would like to especially honor the following:

My parents, Larry and Cynthia French, provided me with the values, education, and drive without which I could not succeed. Thank you for your generosity, your patience, and your love - this journey was as much yours as mine.

My grandparents, Bill and Ted Crutch, who believed I could do anything, even the things I couldn't. You were with me every step of the way, I know, I just wish I could thank you in person.

My sister, Kim, who proved to me a long time ago that you can find your dreams, and still have the good grace and sense of humor to live with them once you do. I always wanted to grow up to be just like you - today I'm one step closer.

Mary Jo, Henrik, Indi, and Moss, without whom this journey might have been possible, but would never have been as easy, or, more importantly, been filled with as much love and laughter.

Our children, Matt, Steven and Carmen, deserve more thanks than I can ever say. Without their unfailing belief that this was not only the right thing for me, but for all of us, I could never have done this. I am more proud of you than I am of any other accomplishment - your hard work and teamwork, your patience and intelligence. I can only wish that you let me travel many more journeys with you, and that I can help you as much as you have helped me.

Finally, but most importantly, my husband Jim. If this was a journey, you have been the beginning, the middle and the end. The beginning, because it was your enthusiasm and persistence which pushed me out the door and started me off on my first faltering steps. The end, because it was your stubbornness and determination which dragged me through those last painful months, despite my whining. And in the middle, there was your integrity and judgement. I knew you couldn't lie to me, and you had faith in me, believed I could do this even in those dark moments when I no longer did. And I knew you must be right. Thank you. May all our journeys be together.

## Contents

### Chapter

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Motivation . . . . .	3
1.3	Historic Applications of EOF Techniques . . . . .	5
1.4	Applications to Seismicity . . . . .	15
<b>2</b>	<b>KL Decomposition - Background &amp; Theory</b>	<b>19</b>
2.1	Introduction . . . . .	19
2.2	Threshold Systems . . . . .	20
2.3	Earthquake Fault Systems . . . . .	24
2.4	KL Theory . . . . .	28
<b>3</b>	<b>KL Decomposition of Southern California Seismicity</b>	<b>38</b>
3.1	Introduction . . . . .	38
3.2	KL Matrix Decomposition Methodology . . . . .	39
3.3	Seismicity Data . . . . .	41
3.4	Results . . . . .	43
3.4.1	Time Period: 1932 through June, 1998. . . . .	43
3.4.2	Time Period: 1932 through August, 1999. . . . .	52
3.4.3	Time Period: 1932 through December, 1991. . . . .	63

3.4.4	Time Period: 1932 through December, 1978. . . . .	74
3.4.5	Time Period: 1812 through 1994, $M \geq 6.0$ . . . . .	80
3.5	Precursory Modes . . . . .	93
<b>4</b>	<b>Dynamics - Theory</b>	<b>101</b>
<b>5</b>	<b>Pattern Dynamics Applied to Southern California Seismicity</b>	<b>107</b>
5.1	Introduction . . . . .	107
5.2	Data . . . . .	108
5.3	Results . . . . .	109
5.3.1	The Phase Dynamical Probability Change, $\Delta P$ . . . . .	109
5.3.2	Phase Angle Rotation, $\Delta \hat{S}$ . . . . .	112
5.3.3	Cumulative probability . . . . .	114
5.4	Random Catalogs . . . . .	116
5.5	Likelihood Tests . . . . .	116
5.6	Discussion . . . . .	122
<b>6</b>	<b>Summary and Future Work</b>	<b>124</b>
6.1	Summary . . . . .	124
6.2	Future Work . . . . .	125
6.2.1	KL Decomposition . . . . .	125
6.2.2	PDPC Method . . . . .	127
<b>Bibliography</b>		<b>128</b>
<b>Appendix</b>		
<b>A</b>	<b>Genetic Algorithm Applications to Geophysical Inverse Problems</b>	<b>141</b>
A.1	Introduction . . . . .	141

A.2 Long Valley . . . . .	142
A.2.1 Geology . . . . .	142
A.2.2 Deformation . . . . .	144
A.2.3 Modeling . . . . .	147
A.2.4 Data Sources . . . . .	148
A.3 Inversion Methodology . . . . .	148
A.3.1 The Genetic Algorithm . . . . .	148
A.3.2 The Fitness Function – the forward problem. . . . .	155
A.4 Spherical vs. Ellipsoidal Functions . . . . .	156
A.4.1 Spherical Source Model . . . . .	157
A.4.2 Ellipsoidal Source Model . . . . .	160
A.5 Synthetic Tests and Sensitivity Analyses . . . . .	165
A.5.1 Sensitivity Tests . . . . .	166
A.6 Inversion Results . . . . .	171
A.7 Conclusions . . . . .	175
A.8 Future work . . . . .	178
<b>B Publications &amp; Grants</b>	<b>179</b>
<b>C Seismicity Catalogs for Southern California</b>	<b>182</b>
C.1 SCEC Database, 1932 - 1999 . . . . .	182
C.2 Catalog of Large Events, M > 6, 1812-1992. . . . .	190
C.3 Catalog of Large Events, M > 5, 1933 - 1994. . . . .	193
<b>D KL Decomposition Programs</b>	<b>199</b>
D.1 Data Sorting Programs . . . . .	199
D.2 KL Decomposition . . . . .	208
D.3 Ouput programs . . . . .	228

**E** Pattern Dynamics (PDPC) Programs 239

**F** Genetic Algorithm Programs 256

## Tables

Table

## Figures

### Figure

1.1	Seismicity for southern California, 1932-1991, normalized to the maximum number of events for the period. . . . .	4
1.2	EOF modes 1, 10, and 20 for SST and thermocline height ( <b>h</b> ) [Goswami et al., 1997]. . . . .	8
1.3	Principal component time series for two years of TOPEX data. The first two modes, a) and b), display an annual cycle, while the third and fourth modes are interannual oscillations [Hendricks et al., 1996]. .	9
1.4	Prediction of future SST anomaly (dark line) and actual SST (light line) for $\tau = 3, 5, 7$ , and 9 months. The dashed line shown is the prediction given by a univariate AR process [Penland and Magonian, 1993]. . . . .	10
1.5	Sixteen standard eigenfaces from PC analysis of 128 face sample space [ <a href="http://www-white.media.mit.edu/vismod/demos/facerec/basic.html">http://www-white.media.mit.edu/vismod/demos/facerec/basic.html</a> ]. . . . .	12
1.6	Typical results for a similarity search using the MIT Photobook program [ <a href="http://www-white.media.mit.edu/vismod/demos/facerec/basic.html">http://www-white.media.mit.edu/vismod/demos/facerec/basic.html</a> ]. . . . .	13

1.7	Recognition rate as a function of the number of eigenfaces used in the reconstruction, again from the MIT Photobook program [ <a href="http://www-white.media.mit.edu/vismod/demos/facerec/basic.html">http://www-white.media.mit.edu/vismod/demos/facerec/basic.html</a> ].	14
1.8	First three eigenmodes, summed, for uplift at various leveling profiles, relative to 1982 datum, Long Valley, California [Savage, 1988].	16
1.9	Spatial (top) and temporal (bottom) factors for the two new modes from the recombination of the first two principal modes, along Highway 395, Long Valley, California. Error bars represent one standard deviation either side of the plotted point [Savage, 1988]. . . . .	17
2.1	Schematic diagram of threshold systems. Note the unobservable state variables $s(\mathbf{x}_i, t)$ and the observable seismicity pattern variables $\psi(\mathbf{x}_i, t)$ [Rundle et al., 2000 (1)]. . . . .	22
2.2	Fault patch model for southern California. Shown at the top is the time-distance plot of simulated seismic activity. The $N = 80$ fault segments (abbreviations in text) are plotted end to end, north to the left. In the center is shown the difference between static and dynamic friction coefficients, $\mu_s - \mu_d$ , as a function of distance along the faults. At the bottom is shown a map view of the fault systems superimposed on a map of California [Rundle, 1988]. . . . .	27
2.3	First eigenvector, $\phi_1$ , for the space-time evolution of the fault patch model shown above. a) $\phi_1$ is normalized for display purposes so that its maximum value = 1.0. b) Segments shown in bold are for $\phi_1(\mathbf{x}_i) > 0.05$ , segments with values of $\phi_1(\mathbf{x}_i) < -0.05$ are shown in dotted lines. All others are shown with a thin line [Rundle et al. 2000]. . . . .	29
2.4	Second eigenvector, $\phi_2$ , as in Fig. 2.3 [Rundle et al., 2000 (1)]. . .	30



3.6 Nineteenth KLE mode for southern California seismicity, 1932-1998. a) PC time series for nineteenth KLE mode; and b) nineteenth KLE mode, normalized to maximum. . . . .	51
3.7 First through fourth KLE modes for southern California seismicity, 1932-1998, $M > 3.0$ . a)First KLE mode; b) second KLE mode; c) third KLE mode; and d) fourth KLE mode, all normalized to the maximum. . . . .	53
3.8 Seventh through tenth KLE modes for southern California seismicity, 1932-1998, $M > 3.0$ a) Seventh KLE mode; b) eighth KLE mode c) ninth KLE mode; and d) tenth KLE mode, all normalized to the maximum. . . . .	54
3.9 Eigenvalues, $\lambda_i^2$ , normalized by the sum of all eigenvalues, versus mode for the KLE decomposition of the entire southern California data set, 1932-1999. . . . .	55
3.10 KLE mode 500, southern California seismicity 1932-1998. . . . .	56
3.11 First two KLE modes for southern California seismicity, 1932-1999. a) PC time series for first KLE mode; b) first KLE mode, normalized to maximum; c) PC time series for second KLE mode; and d) second KLE mode, also normalized to the maximum. . . . .	57
3.12 Third and fourth KLE modes for southern California seismicity, 1932-1999. a) PC time series for third KLE mode; b) third KLE mode, normalized to maximum; c) PC time series for fourth KLE mode; and d) fourth KLE mode, also normalized to the maximum. . . . .	59
3.13 Fifth and sixth KLE modes for southern California seismicity, 1932-1999. a) PC time series for fifth KLE mode; b) fifth KLE mode, normalized to maximum; c) PC time series for sixth KLE mode; and d) sixth KLE mode, also normalized to the maximum. . . . .	60

3.14 Seventh and eighth KLE modes for southern California seismicity, 1932-1998. a) PC time series for seventh KLE mode; b) seventh KLE mode, normalized to maximum; c) PC time series for eighth KLE mode; and d) eighth KLE mode, also normalized to the maximum. . . . .	61
3.15 Ninth and tenth KLE modes for southern California seismicity, 1932-1999. a) PC time series for ninth KLE mode; b) ninth KLE mode, normalized to maximum; c) PC time series for tenth KLE mode; and d) tenth KLE mode, also normalized to the maximum. . . . .	62
3.16 First two KLE modes for southern California seismicity, 1932-1999, with a magnitude cut of 3.0. a) PC time series for first KLE mode; b) first KLE mode, normalized to maximum; c) PC time series for second KLE mode; and d) second KLE mode, also normalized to the maximum. . . . .	64
3.17 Third and fourth KLE modes for southern California seismicity, 1932-1999, with a magnitude cut of 3.0. a) PC time series for third KLE mode; b) third KLE mode, normalized to maximum; c) PC time series for fourth KLE mode; and d) fourth KLE mode, also normalized to the maximum. . . . .	65
3.18 Fifth and sixth KLE modes for southern California seismicity, 1932-1999, with a magnitude cut of 3.0. a) PC time series for fifth KLE mode; b) fifth KLE mode, normalized to maximum; c) PC time series for sixth KLE mode; and d) sixth KLE mode, also normalized to the maximum. . . . .	66

3.19 Seventh and eighth KLE modes for southern California seismicity, 1932-1999, with a magnitude cut of 3.0. a) PC time series for seventh KLE mode; b) seventh KLE mode, normalized to maximum; c) PC time series for eighth KLE mode; and d) eighth KLE mode, also normalized to the maximum. . . . .	67
3.20 Ninth and tenth KLE modes for southern California seismicity, 1932-1999, with a magnitude cut of 3.0. a) PC time series for ninth KLE mode; b) ninth KLE mode, normalized to maximum; c) PC time series for tenth KLE mode; and d) tenth KLE mode, also normalized to the maximum. . . . .	68
3.21 Eigenvalues, $\lambda_i^2$ , normalized by the sum of all eigenvalues, versus mode for the KLE decomposition of the entire southern California data set, 1932-1991. . . . .	70
3.22 First two KLE modes for southern California seismicity, 1932-1991. a) PC time series for first KLE mode; b) first KLE mode, normalized to maximum; c) PC time series for second KLE mode; and d) second KLE mode, also normalized to the maximum. . . . .	71
3.23 The third and fourth KLE modes for southern California seismicity, 1932-1991. a) PC time series for third KLE mode; b) third KLE mode, normalized to maximum; c) PC time series for fourth KLE mode; and d) fourth KLE mode, also normalized to the maximum. . . . .	72
3.24 Fifth and sixth KLE modes for southern California seismicity, 1932-1991. a) PC time series for fifth KLE mode; b) fifth KLE mode, normalized to maximum; c) PC time series for sixth KLE mode; and d) sixth KLE mode, also normalized to the maximum. . . . .	73

3.25 Seventh and eighth KLE modes for southern California seismicity, 1932-1991. a) PC time series for seventh KLE mode; b) seventh KLE mode, normalized to maximum; c) PC time series for eighth KLE mode; and d) eighth KLE mode, also normalized to the max- imum. . . . .	75
3.26 Ninth and tenth KLE modes for southern California seismicity, 1932-1991. a) PC time series for ninth KLE mode; b) ninth KLE mode, normalized to maximum; c) PC time series for tenth KLE mode; and d) tenth KLE mode, also normalized to the maximum. . . . .	76
3.27 First two KLE modes for southern California seismicity, 1932-1991, $M > 3.0$ . a) First KLE mode, normalized to maximum; b) second KLE mode, also normalized to the maximum. . . . .	77
3.28 Third through sixth KLE modes for southern California seismicity, 1932-1991, $M > 3.0$ . a) Third KLE mode; b) fourth KLE mode; c) fifth KLE mode; and d) sixth KLE mode, all normalized to the maximum. . . . .	78
3.29 Seventh through tenth KLE modes for southern California seismic- ity, 1932-1991, $M > 3.0$ . a) Seventh KLE mode; b) eighth KLE mode; c) ninth KLE mode; and d) tenth KLE mode, all normalized to the maximum. . . . .	79
3.30 Eigenvalues, $\lambda_i^2$ , normalized by the sum of all eigenvalues, versus mode for the KLE decomposition of the entire southern California data set, 1932-1978. . . . .	81
3.31 First two KLE modes for southern California seismicity, 1932-1978. a) PC time series for first KLE mode; b) first KLE mode, normal- ized to maximum; c) PC time series for second KLE mode; and d) second KLE mode, also normalized to the maximum. . . . .	82

3.32 Third and fourth KLE modes for southern California seismicity, 1932-1978. a) PC time series for third KLE mode; b) third KLE mode, normalized to maximum; c) PC time series for fourth KLE mode; and d) fourth KLE mode, also normalized to the maximum.	83
3.33 Seventh and ninth KLE modes for southern California seismicity, 1932-1978. a) PC time series for seventh KLE mode; b) seventh KLE mode, normalized to maximum; c) PC time series for ninth KLE mode; and d) ninth KLE mode, also normalized to the maximum. . . . .	84
3.34 First two KLE modes for southern California seismicity, 1932-1978, $M > 3.0$ . a) First KLE mode and b) second KLE mode, normalized to maximum. . . . .	85
3.35 Fifth through seventh KLE modes for southern California seismicity, 1932-1978, $M > 3.0$ . a) Fifth KLE mode; b) sixth KLE mode; c) seventh KLE mode; and d) eighth KLE mode, all normalized to the maximum. . . . .	86
3.36 First and second KLE modes for southern California seismicity, 1812 - 1994, $M > 6.0$ . a) PC time series to first KLE mode; b) first KLE mode; c) PC time series to second KLE mode; and d) second KLE mode, all normalized to the maximum. . . . .	88
3.37 Third and fourth KLE modes for southern California seismicity, 1812 - 1994, $M > 6.0$ . a) PC time series to third KLE mode; b) third KLE mode; c) PC time series to fourth KLE mode; and d) fourth KLE mode, all normalized to the maximum. . . . .	89

3.38 Fifth and sixth KLE modes for southern California seismicity, 1812 - 1994, $M > 6.0$ . a) PC time series to fifth KLE mode; b) fifth KLE mode; c) PC time series to sixth KLE mode; and d) sixth KLE mode, all normalized to the maximum. . . . .	90
3.39 Ninth and tenth KLE modes for southern California seismicity, 1812 - 1994, $M > 6.0$ . a) PC time series to ninth KLE mode; b) ninth KLE mode; c) PC time series to tenth KLE mode; and d) tenth KLE mode, all normalized to the maximum. . . . .	91
3.40 Major earthquakes at Parkfield, California, since 1850 [from Turcotte, 1997]. . . . .	92
3.41 $\alpha_i$ , plotted for the five years prior to the 1992 Landers earthquake. Values are normalized to the first mode. . . . .	95
3.42 $\alpha_i$ , plotted for the five years prior to the 1979 Imperial Valley earthquake. Values are normalized to the first mode. . . . .	97
3.43 $\psi^{2-100}$ , summed over the years 1971 to 1972. Values are normalized to the first mode. . . . .	98
3.44 $\psi^{2-100}$ , summed over the years 1971 to 1975. Values are normalized to the first mode. . . . .	99
3.45 $\psi^{2-100}$ , summed over the years 1971 to 1979. Values are normalized to the first mode. . . . .	100
4.1 Cumulative number of earthquake, $N$ , with magnitude greater than $m$ for each year between 1980 and 1994 for southern California. a) 1980-1984; b) 1985-1989; c) 1990-1994. The solid straight line is the Gutenberg-Richter relation with $b = 1.05$ and $a = 2.06 \times 10^5 \text{ yr}^{-1}$ [Turcotte, 1997]. . . . .	103

5.1	Maps of $\Delta P$ for the time intervals a) 1986-1975, b) 1991-1984, c) 1991-1988 and d) 1991-1975. In constructing these maps, no data is used from time periods after December 31, 1991. Inverted triangles represent events that occurred during the indicated time periods, with three sizes corresponding to magnitudes of: $5 < M < 6$ , $6 <$ $M < 7$ , and $M > 7$ . Filled circles represent events that occur after January 1, 1992. Three increasing circle sizes again correspond to the same magnitude ranges as for the inverted triangles. . . . .	111
5.2	Maps of $\Delta P$ for the time intervals shown. The plot in the top right-hand corner shows the events which occur between 1991 and 1999. Again, no data is used from time periods after December 31, 1991. Inverted triangles represent events that occurred during the indicated time periods. The size of the symbols corresponds to magnitudes as in Figure 5.1 above. . . . .	113
5.3	Map of $\Delta \hat{S}(\mathbf{x}_i, 1978, 1991)$ , normalized to the maximum absolute value. Red indicates an increase in $\Delta \hat{S}$ , blue a decrease in $\Delta \hat{S}$ . . .	115
5.4	$ \Delta \hat{S}(\mathbf{x}_i, t_1, t_2) ^2$ plotted for the time period 1970 to 1990, at the lo- cation of the historic Parkfield event. . . . .	117
5.5	Map of $\Delta P$ for one random catalog for the time period 1991-1978. Inverted triangles represent events that occurred during the indi- cated time periods. The sizes of both circles and triangles corre- sponds to magnitudes as in Figure 5.1 above. . . . .	118

5.6 Histogram of likelihood values for one hundred random catalogs of southern California seismicity, time period 1991-1978. In addition, the likelihood value for the actual catalog, and the same time period, as shown in Figure 5.2 above, is superimposed on this plot as a dashed line. The likelihood value for the normalized southern California seismicity of 1.1 is plotted as a dash-dotted line. . . . .	121
A.1 Long Valley caldera, eastern California. . . . .	143
A.2 Two-color EDM & leveling lines at Long Valley caldera (adapted from LVO web site, <a href="http://quake.wr.usgs.gov/">http://quake.wr.usgs.gov/</a> QUAKES/ geodetic/ twocolor /longvalley.html). . . . .	146
A.3 Schematic of genetic algorithm inversion program. . . . .	151
A.4 Schematic view of GA operations, a) crossover and b) mutation. .	153
A.5 Volcanic source models for Long Valley caldera. a) Spherical Mogi source and b) finite, prolate ellipsoid. . . . .	158
A.6 Coordinate system for ellipsoidal model. . . . .	163
A.7 Sensitivity of fit to Mogi source parameters. . . . .	168
A.8 Sensitivity of fit to ellipsoidal parameters. . . . .	169
A.9 Variation in fitness with a) volume, b) length of semimajor axis, c) depth, and d) theta. . . . .	172
A.10 Location of Mogi point sources as determined by GA inversion. . .	173
A.11 Location of finite ellipsoid sources as determined by GA inversion.	
a) Source at 9.6 km: $\phi = 179^\circ$ , $\theta = 109^\circ$ , $V=0.037 \text{ km}^3$ , $a = 2000$ m.	
b) Source at 11.8 km: $\phi = 105^\circ$ , $\theta = 1^\circ$ , $V=0.002 \text{ km}^3$ , $a = 290$ m.	
	174

# Chapter 1

## Introduction

### 1.1 Background

The historical earthquake record, while not complete, spans hundreds to thousands of years of human history. As a result, it has long been recognized that earthquake mainshocks occur at quasi periodic intervals and that, for some parts of the world, average recurrence intervals are well defined [Kanamori, 1981]. In addition, both temporal and spatial clustering is evident in the data, exemplified by such phenomena as foreshocks, aftershocks, seismic gaps, or mainshock triggering, with the result that none of these observed patterns are compatible with a Poisson probability function [Kagan and Jackson, 1992; Savage, 1993; Grant and Sieh, 1994; Dieterich, 1994; Rundle and Klein, 1995; Jones and Hauksson, 1997; Turcotte, 1997].

In the 1990s, two major developments in seismic research greatly added to the understanding of earthquake fault systems. Stress transfer interaction studies by Harris and Simpson, 1992, Jaume and Sykes, 1992, Stein et al., 1992, King et al., 1994, Gomberg, 1996, and Stein, 1999, to name a few, have shown that fault interactions are in large part controlled by the stress state of the underlying geophysical medium. On the other hand, a combination of theoretical analysis and numerical simulations established the link between earthquake fault networks and the physics of critical point systems. Earthquake fault systems have been shown to

be critical point systems, characterized by nonlinear dynamics, large correlation lengths, and the Gutenberg-Richter scaling law [Bak and Tang, 1989; Rundle, 1989; Pacheco et al., 1992; Ben-Zion et al., 1993; Romanowicz and Rundle, 1993; Rundle, 1993; Saleur et al., 1995; Rundle and Klein, 1995; Rundle et al., 1995; Klein et al., 1996; Sammis et al., 1996; Eneva and Ben-Zion, 1997; Fisher et al., 1997; Rundle et al., 1997; Huang et al., 1998; Ferguson et al., 1999; Jaume and Sykes, 1999; Rundle et al., 1999]. However, direct analysis of the actual earthquake fault system remains incomplete due to the inherent difficulty in sampling the continuum and its physical parameters [Richter, 1958; Kanamori, 1981; Turcotte, 1991; Geller et al., 1997; Wyss, 1997; Sykes et al., 1999].

Much of the recent geophysical research associated with earthquakes themselves has centered on investigating the spatial and temporal patterns that exist in local and regional seismicity data [Kanamori, 1981]. Notable examples include characteristic earthquakes [Swan et al., 1980; Ellsworth and Cole, 1997], repeating earthquakes [Bakun et al., 1986; Marone et al., 1995], seismic gaps [Haberman 1981; House et al., 1981; Kagan and Jackson, 1992; Wyss and Wiemer, 1999], well-defined recurrence intervals [Bakun and McEvilly, 1984; Lyzenga et al., 1991; Savage, 1993], Mogi donuts [Mogi, 1969; Mogi, 1977], temporal clustering [Frohlich, 1987; Dodge et al., 1996; Jones and Hauksson, 1997; Rundle et al., 1997], ‘slow’ earthquakes [Linde et al., 1996; McGuire et al., 1996; Kerr, 1998], precursory quiescence [Yamashita and Knopoff, 1989; Wyss et al., 1996; Wyss et al., 2000; Kato et al., 1997], aftershock sequences [Nanjo et al., 1998; Gross and Kisslinger, 1994], earthquake triggering over large distances [Hill et al., 1993; King et al., 1994; Hill et al., 1995; Stark and Davis, 1996; Pollitz and Sacks, 1997], and time-to-failure analyses [Bufe and Varnes, 1993; Bowman et al., 1998; Brehm and Braile, 1999]. Although much of this work represents important efforts to describe these characteristic patterns using empirical probability density functions,

none of these observations or methodologies systematically identifies all possible seismicity patterns. The quantification of all possible space-time patterns would seem to be a necessary first step in the process of identifying which patterns are precursory to large events, leading to the possible development of new approaches in forecast methodology. Yet, as can be seen in Figure 1.1, a plot of relative southern California seismicity during the time period 1932-1991, the identification and quantification of these patterns is no easy matter.

## 1.2 Motivation

Karhunen-Loeve (KL) decomposition, a linear decomposition technique in which a dynamical system is decomposed into a complete set of orthonormal subspaces, have 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 even geophysics [Hotelling, 1933; Preisendorfer, 1988; Savage, 1988; Penland, 1989; Vautard and Ghil, 1989; Fukunaga, 1990; Garcia and Penland, 1991; Penland and Magorian, 1993; Penland and Sardeshmukh, 1995; Holmes et al., 1996]. The notable success of this method in analyzing the ocean-atmosphere interface, a nonlinear system whose underlying physics is governed by the Navier-Stokes equation, suggested its application to the analysis of the earthquake fault system [North, 1984; Hasselman, 1988; Preisendorfer, 1988; Penland and Magorian, 1993; Penland and Sardeshmukh, 1996].

It is important to recognize that the KL decomposition, while most often applied to nonlinear processes, is a linear procedure. It therefore produces linear spaces, or eigenpatterns, in which the dynamical nonlinear system is defined. In other words, while the decomposition process produces a set of complete set of linear operators, with the inherent strengths of linearity, they do not capture the underlying nonlinear process [Holmes et al., 1996]. However, we can think of

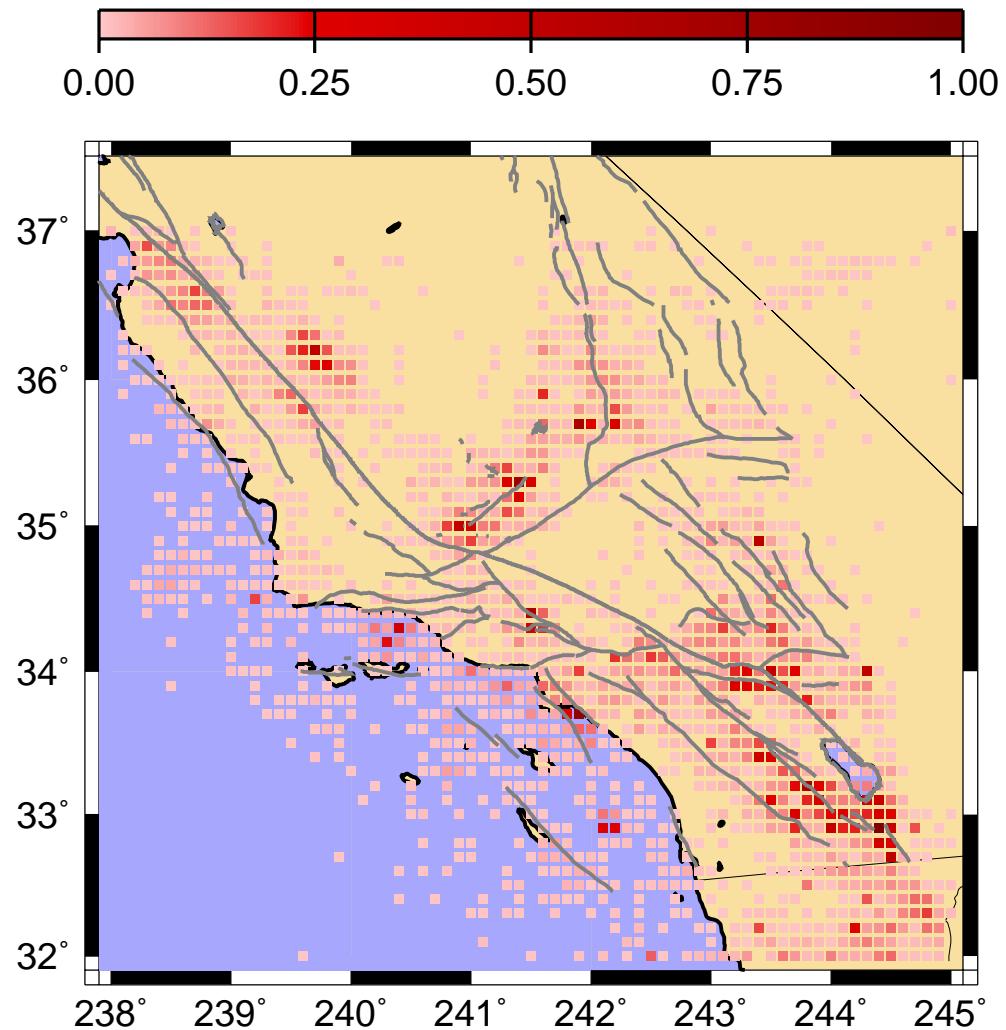


Figure 1.1: Seismicity for southern California, 1932-1991, normalized to the maximum number of events for the period.

the set of orthonormal basis functions produced by the KL decomposition as a description of the pattern phase space, in which the basis modes themselves are attractors, that capture information about the basins of attraction and invariant sets which they comprise [Holmes et al., 1996].

### **1.3 Historic Applications of EOF Techniques**

Empirical orthogonal functions (EOFs), initially introduced as principal components applied to the field of multivariate statistics during the 1930s by Hotelling, were subsequently applied to the meteorological sciences by Lorenz in the 1940s [Hotelling, 1933; North, 1984; Penland, 1989]. Although in the study of stochastic fields these techniques have come to be known as Karhunen-Loeve basis functions, by far the most widespread application and development of their capabilities and properties is in the atmospheric sciences, where EOFs have been used to represent observed data, to develop predictors, or to physically interpret the data [North, 1984].

EOFs, orthogonal eigenfunctions calculated from a set of observed data time series, are statistically independent spatial modes, the normal modes of the system of interest. The associated eigenvalues represent the percent of the variance in the data accounted for by that particular eigenmode, and the principal components, which are the time series for each eigenmode reconstructed from the actual data, are statistically stationary representations of the eigenmodes signal in time [North, 1984; Preisendorfer, 1988; Penland, 1989].

The foundation for the theory of EOF analysis is that geophysical data can be thought of as pattern realizations of random fields generated by a stochastic process, and that each measurement time step is just another unique visualization of that process. Various second moment statistics can be constructed from this data which describe the connections between individual points on that map,

but the one used for most EOF analyses is the covariance statistic. The data matrix is constructed from the observation time series at each measurement location. The covariance matrix then is formed by multiplying the data matrix with its transpose, and subsequently is decomposed into a unique set of independent, orthonormal eigenvectors and eigenvalues [Aubrey and Emery, 1983; North, 1984; Preisendorfer, 1988; Savage, 1988; Penland, 1989]. These EOF modes then are ordered from largest variance to smallest variance, based upon their eigenvalues [North, 1984; Preisendorfer, 1988].

Applications of EOF decomposition techniques diversified in recent years to include model verification, mode shaping, and forecasting, and even to include rare examples in the field of solid earth geophysics. For example, North, 1984, Preisendorfer, 1988, Vautard and Ghil, 1989, and Navarra, 1993, all generalized the method to a variety of linearized meteorological problems, exploring the various constraints imposed by boundary conditions, the response to forcing, selection rules, sampling realizations and associated errors. Aside from effects associated with data completeness, accuracy of the EOF decomposition is governed primarily by the limits of the linear decomposition of nonlinear systems and the assumption of time stationarity. However, even in an extremely nonlinear, noise-forced system such as the ocean-atmosphere interface, the technique has been used successfully in various forms and applications [North, 1984; Hasselman, 1988; Penland and Magorian, 1993; Lee and Cornillon, 1995; Hendricks et al., 1996].

Goswami et al., 1997, studied the dependence on scale and the source of seasonality in their coupled ocean-atmosphere model, and the effects of small perturbations on the subsequent error growth in their model. Figure 1.2 shows EOF modes 1, 10 and 20 for sea surface temperature (SST) and thermocline height (**h**). While there is no one-to-one correspondence between EOF index and spatial scale, i.e., the  $(n + 1)$  mode is not necessarily of a smaller spatial scale than the

*n*th mode, EOF1 certainly is a large scale basis function, while EOF10 is a much smaller scale [Goswami et al., 1997].

Lee and Cornillon, 1995, studied the patterns of temporal variation in the Gulf Stream, while Hendricks et al., 1996, used EOF techniques to analyze annual and interannual variability in TOPEX/POSEIDON data. Figure 1.3 shows the principal component time series associated with the first four EOF modes determined from only two years of TOPEX data. Modes one and two contain a large annual signal, while modes three and four are interannual oscillations [Hendricks et al., 1996].

Numerous researchers in the atmospheric/oceanographic sciences expanded the use of EOF for forecasting future events. This method involves calculating the four-dimensional EOFs obtained by Fourier transforming the data, so that a complete set of normal modes is obtained for each frequency as well, revealing those eigenmodes which are frequency dependent [North, 1984]. Given various appellations such as POP (principal oscillation patterns) [Penland, 1989], PIP (principal interaction patterns) [Hasselmann, 1988], dynamic EOF analysis [Chang and Mak, 1993], or an unequal time operator [Broomhead and King, 1986; Vautard and Ghil, 1989], these techniques are based upon the development of complex EOFs which can be used to project the data signal forward in time, normally to include an analysis of the noise signal effects [Hasselmann, 1988; Preisendorfer, 1988; Penland, 1989; Vautard and Ghil, 1989; Chang and Mak, 1993; Goswami et al., 1997]. Figure 1.4 shows the prediction of Nino 3 using ten complex normal modes in a POP analysis for various lead times. While the predictions degrade over time periods which include a large anomaly, the results support the implicit assumption in EOF analysis that the dynamics can be analyzed as if linear over appropriate time periods [Penland and Magorian, 1993].

Other researchers used EOF analysis techniques to decompose the data sig-

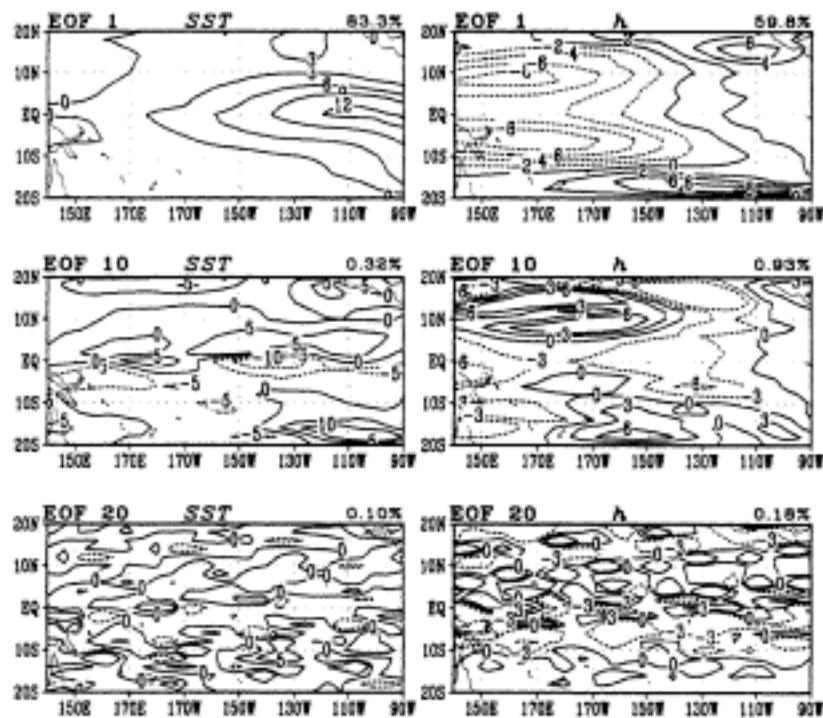


Figure 1.2: EOF modes 1, 10, and 20 for SST and thermocline height ( $h$ ) [Goswami et al., 1997].

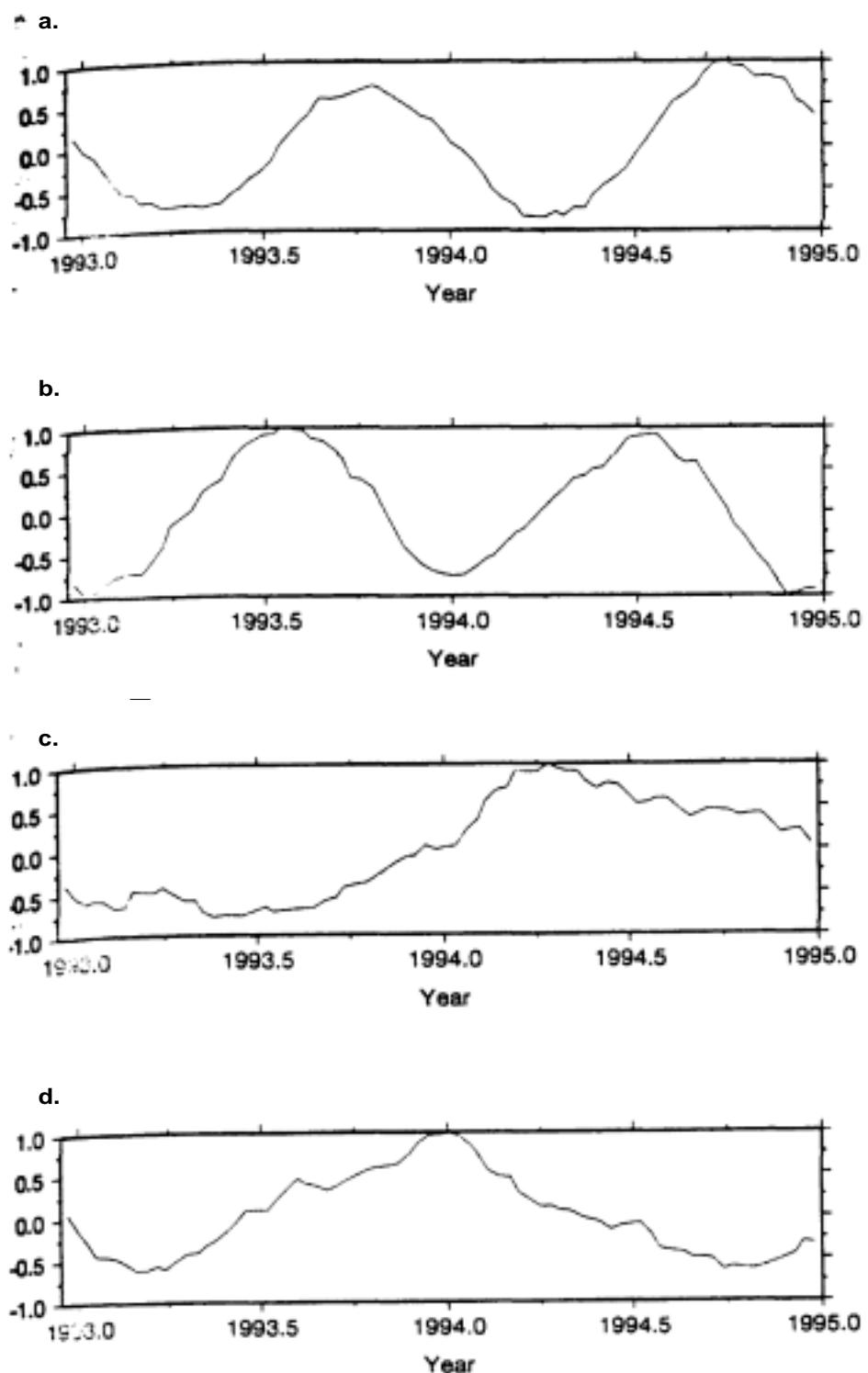


Figure 1.3: Principal component time series for two years of TOPEX data. The first two modes, a) and b), display an annual cycle, while the third and fourth modes are interannual oscillations [Hendricks et al., 1996].

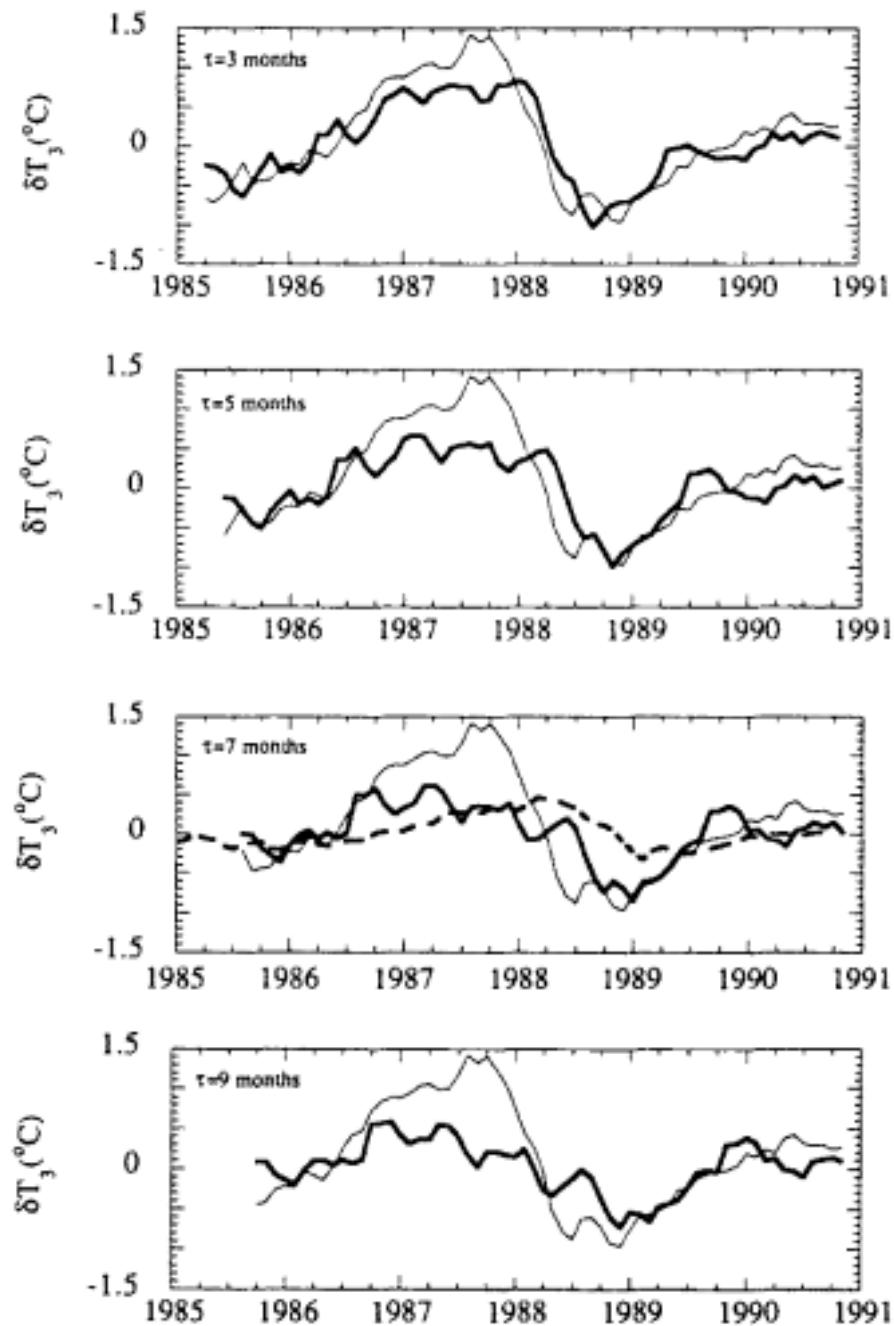


Figure 1.4: Prediction of future SST anomaly (dark line) and actual SST (light line) for  $\tau = 3, 5, 7$ , and 9 months. The dashed line shown is the prediction given by a univariate AR process [Penland and Magorian, 1993].

nal into its predominant modes, keeping those modes which describe the largest portion of the variance, and removing those modes which are of little to no interest, assumed to represent noise modes. In the field of biometrics, facial recognitions systems are being developed which use principal component analysis to create eigentemplates for both pattern recognition and noise reduction. From an initial database of several hundred faces, eigenvectors are produced which are reformed into a set of eigenfaces, from a smaller subset of the eigenvectors [<http://www-white.media.mit.edu/vismod/demos/facerec/basic.html>; <http://www.cam-orl.co.uk/facesataglance.html>; Moghaddam et al., 1998; [www.cc.gatech.edu/classes/cs7322\\_97\\_spring/participants/Gonzalez](http://www.cc.gatech.edu/classes/cs7322_97_spring/participants/Gonzalez)]. Figure 1.5 shows one set of standard eigenfaces decomposed from 128 representative faces. The method then compares the best matches for the subject test. The reconstruction error, or distance-from-face-space (DFFS), for each potential match is computed using the residual variance from the principal component analysis. This DFFS is used to judge the best fit solution. Figure 1.6 is a set of typical results for a similarity search in Photobook, an interactive demonstration from MIT available on the internet [<http://www-white.media.mit.edu/vismod/demos/facerec/basic.html>]. This particular database contains 7562 images of approximately 3000 people, with varying expressions, headgear, facial hair, and lighting. The recognition rate, depending on whether the detection is enhanced based on particular facial features, varies from 95 to 98 percent asymptotically (Figure 1.7) [<http://www-white.media.mit.edu/vismod/demos/facerec/basic.html>].

EOF techniques also have been applied to solid earth data, although not on a widespread basis. Aubrey and Emery [Aubrey and Emery, 1983; Aubrey and Emery, 1986], for example, used eigenfunction analysis to study crustal deformation using tide gauge data, while Savage analyzed the deformation at Long Valley caldera, California, using principal component analysis to determine the primary

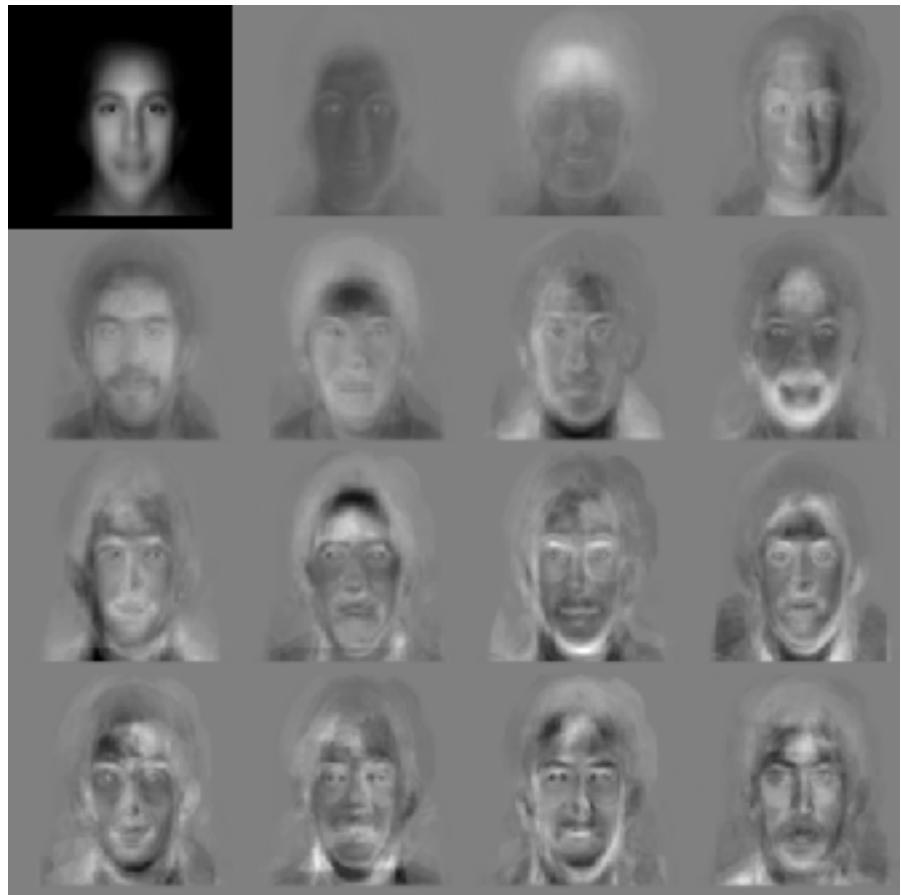


Figure 1.5: Sixteen standard eigenfaces from PC analysis of 128 face sample space [<http://www-white.media.mit.edu/vismod/demos/facerec/basic.html>].

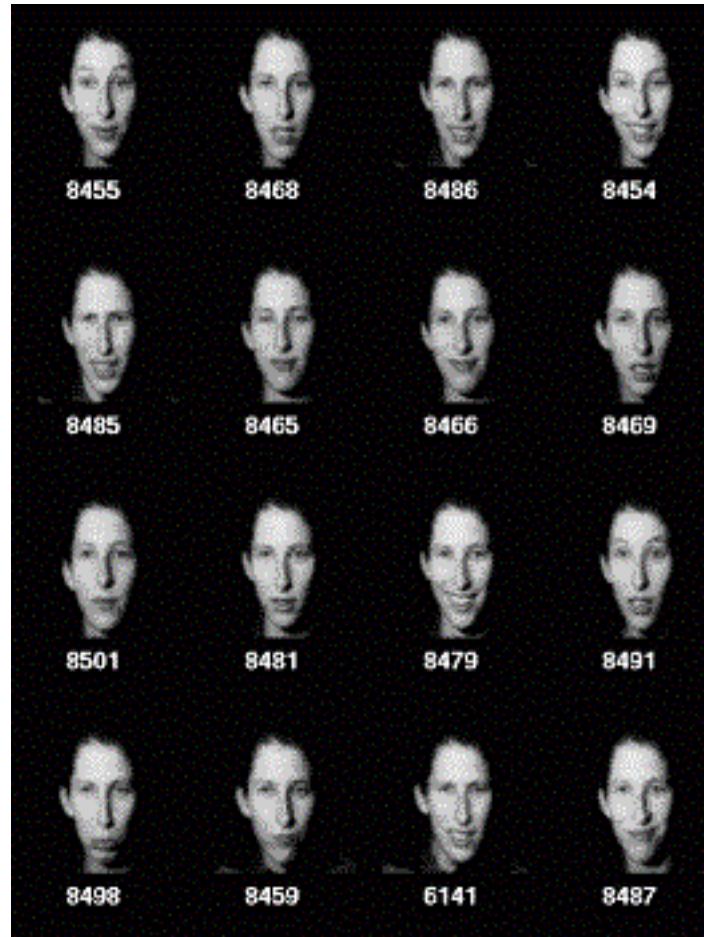


Figure 1.6: Typical results for a similarity search using the MIT Photobook program [<http://www-white.media.mit.edu/vismod/demos/facerec/basic.html>].

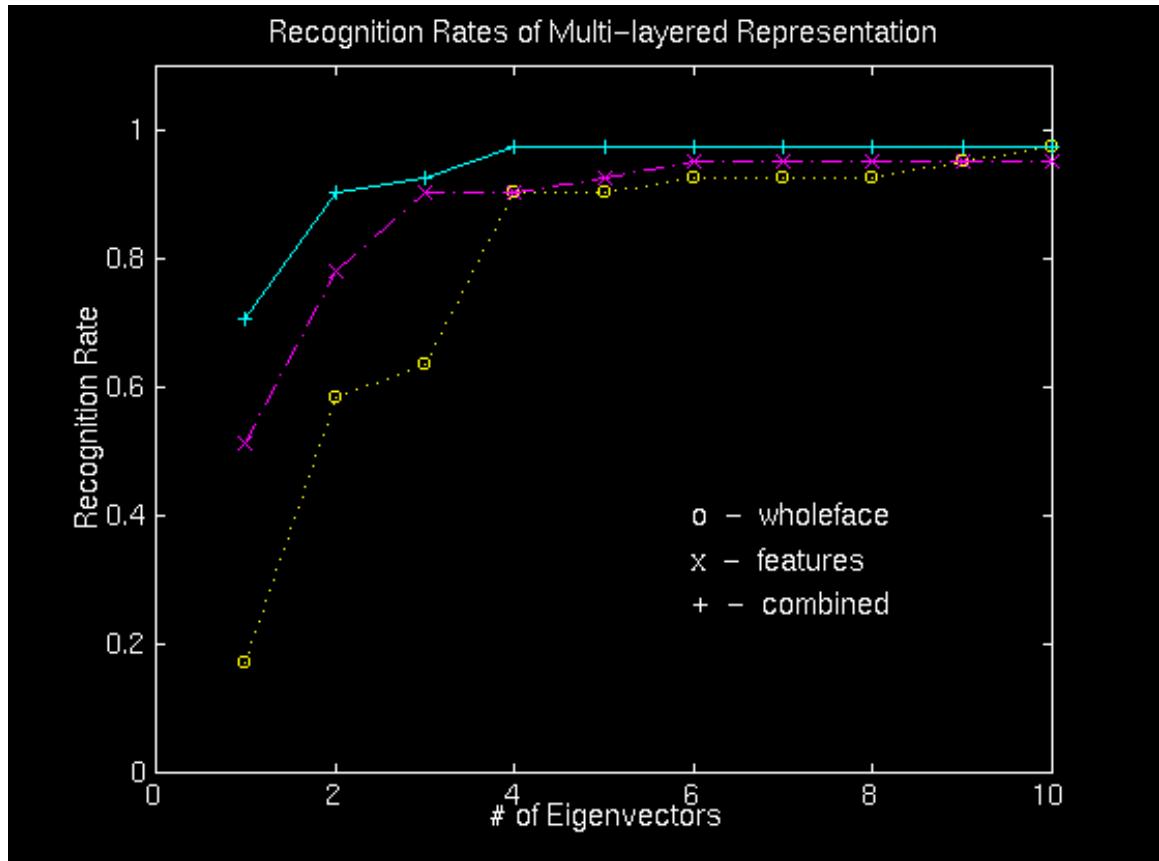


Figure 1.7: Recognition rate as a function of the number of eigenfaces used in the reconstruction, again from the MIT Photobook program [<http://www-white.media.mit.edu/vismod/demos/facerec/basic.html>].

modes of deformation between 1983 and 1987 [Savage, 1988]. Figure 1.8 shows plots of various summations for the first three eigenmodes of deformation along Highway 395, for various time periods. The first eigenmode represents uplift from the magma chamber inflation under Long Valley, while the second is likely due to a systematic leveling error, and the third is correlated with the topography. One interesting application of this eigenfunction expansion is that Savage recombines the first two modes to eliminate a suspicious tilt in the data, resulting in the profile shown in Figure 1.9. This illustrates a powerful tool of eigenfunction expansion techniques - orthonormal, independent eigenvectors can be recombined to represent various parts of the signal of interest for analysis or forecasting purposes [Savage, 1988].

On the other hand, Douma et al. used the EOF decomposition to analyze the solutions to various seismic tomography models of the earth to see which modes accounted for the smallest variance - i.e. which modes were reproduced almost exactly by every model - in order to determine which features of the earth were well captured by all models, and conversely, which features were difficult to characterize properly [Douma et al., 1996].

#### **1.4 Applications to Seismicity**

In their recent paper, Rundle et al., 2000 (1), describe the application of a method developed to analyze the results of large scale numerical simulations of southern California. This pattern dynamics technique, which is based upon the decomposition of a correlation matrix in a manner similar to EOF techniques, but is more closely aligned with the Karhunen-Loeve (KL) expansion that is its theoretical foundation, defines the seismicity in terms of a complex phase angle. The KL decomposition then identifies spatial eigenvalue patterns, or eigenpatterns, in slip time histories for a southern California fault patch model which

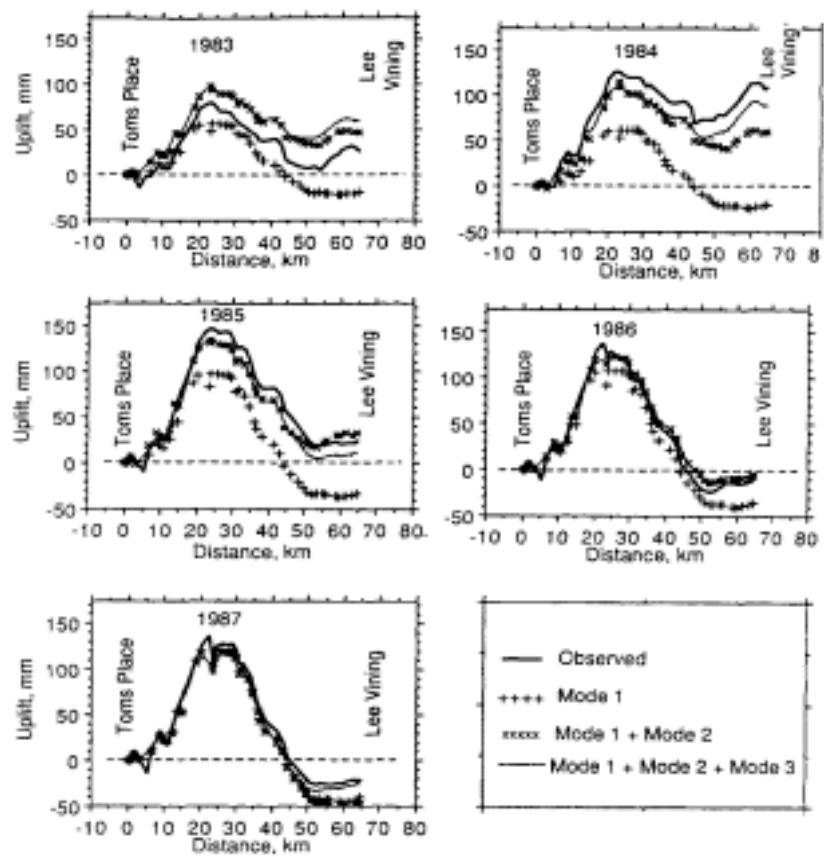


Figure 1.8: First three eigenmodes, summed, for uplift at various leveling profiles, relative to 1982 datum, Long Valley, California [Savage, 1988].

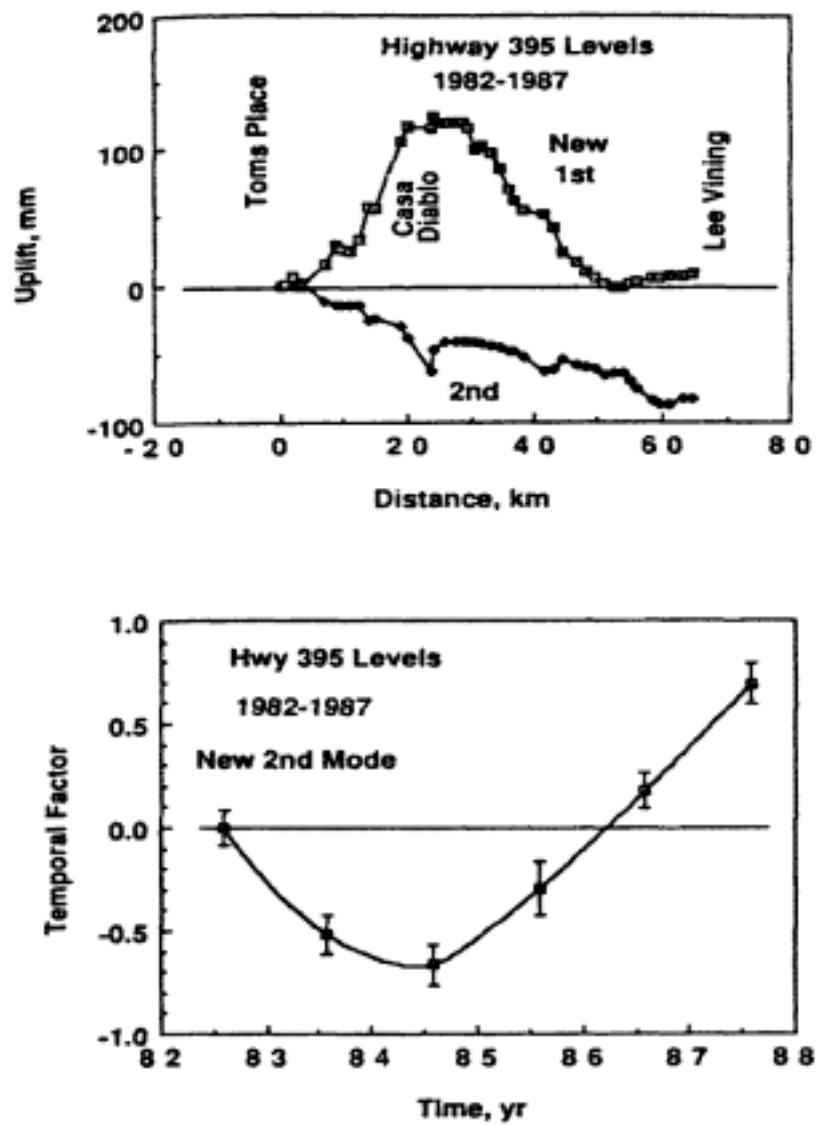


Figure 1.9: Spatial (top) and temporal (bottom) factors for the two new modes from the recombination of the first two principal modes, along Highway 395, Long Valley, California. Error bars represent one standard deviation either side of the plotted point [Savage, 1988].

spans thousands of time steps.

Here I apply this pattern dynamics method [Tiampo et al., 1998; Rundle et al., 1999; Tiampo et al., 1999; Rundle et al., 2000 (1)] to the analysis of observed seismicity data from southern California in order to identify the basis patterns for all possible space-time seismicity configurations. These basis states represent a complete, orthonormal set of eigenvectors and associated eigenvalues that are obtained from the diagonalization of the correlation operators computed for the regional historic seismicity data. Next, I expand on this approach to seismicity as a phase angle and apply it to the historic data in order to identify the precursory modes and the cumulative change in probability of an event at a particular location in southern California.

## Chapter 2

### KL Decomposition - Background & Theory

#### 2.1 Introduction

Earthquake fault systems are an example of a complex nonlinear system [Rundle and Klein, 1995; Fisher et al., 1997]. Interactions among a spatial network of fault segments are mediated by means of a potential that allows stresses to be redistributed to other segments following slip on any particular segment. For faults embedded in a linear elastic host, this potential is a stress Green's function whose exact form is calculated from the equations of linear elasticity, once the geometry of the fault system is specified. A persistent driving force, arising from plate tectonic motions, increases stress on the fault segments. Once the stresses reach a threshold characterizing the limit of stability of the fault, a sudden slip event results. The slipping segment can also trigger slip at other locations on the fault surface whose stress levels are near the failure threshold as the event begins. In this manner, earthquakes result from the interactions and nonlinear nature of the stress thresholds [Fisher et al., 1997; Rundle et al., 1997; Rundle et al., 1999].

Because the external fault medium is elastic, the interactions between different patches and different faults is also elastic over short time scales. This fundamental elastic interaction results in the formulation of a mean field regime for the earthquake system [Rundle and Klein, 1995; Rundle et al., 1997; Rundle et al., 1999]. A mean field system is one in which fluctuations are suppressed for all but

the longest wavelengths. This mean field behavior occurs in elastic systems due to the inverse-cube nature of the stress Green's function (proportional to  $1/|x - x'|^3$ ) and controls the elastic stress interactions between the defects, or fault patches, in the solid medium [Klein et al., 1996; Ferguson et al., 1999]. These long-range interactions lead to an averaging of stress over the system, damping the effects of short wavelength details [Eshelby, 1957; Yeomans, 1992]. Longer spatial and temporal wavelength effects become increasingly important, and the correlation lengths of these mean field systems become increasingly larger as they approach a critical point, in association with power law scaling similar to the Gutenberg-Richter relation [Yeomans, 1992; Rundle et al., 1995; Klein et al., 1996; Rundle et al., 1999]. Finally, such mean field lattice systems are nonequilibrium systems that can be treated as equilibrium systems as they settle into a metastable equilibrium state. In these systems, the time averaged elastic energy of the system fluctuates around a constant value for some long period of time. These periods are punctuated by major events which reorder the system before it settles into another metastable well around a new mean energy state [Rundle et al., 1995; Klein et al., 1997; Rundle et al., 1999; Ferguson et al., 1999].

## 2.2 Threshold Systems

From the references noted above, earthquake fault systems are examples of driven nonlinear threshold systems, comprised of interacting spatial networks of statistically identical, nonlinear units that are subjected to a persistent driving force [Scholz, 1990; Rundle et al., 1995; Fisher et al., 1997]. Numerous examples of such systems exist, including neural networks [Hertz et al., 1991; Herz and Hopfield, 1995], sandpiles [Bak et al., 1987], and superconductors [Fisher, 1985], of which earthquakes are but another example. Such systems are composed of cells which fire, or fail, when the driving force causes the force or potential,  $\sigma(\mathbf{x}, t)$

on a cell at location  $\mathbf{x}$  and time  $t$  to reach a predefined threshold value  $\sigma^F$ . The behavior of these systems is determined by parameters such as threshold values, residual stresses, quenched disorder and noise [Bak et al., 1987; Rundle et al., 1995; Fisher et al., 1997]. Complex spatial and temporal firing patterns result which are difficult to analyze deterministically [Nijhout et al., 1997]. In the case of an earthquake fault system, the driving force is tectonic plate motion, and the internal potential is the stress on each fault cell or patch. The firing, or failure of each patch results in an increase in the internal state variable  $s(\mathbf{x},t)$  and a decrease in the cell potential to some residual value  $\sigma^R$ . The interactions between the cells, or fault patches, may be excitatory, bringing another closer to failure, or inhibitory, in which the failure of one cell can move neighboring cells further from failure. The spatial and temporal firing patterns,  $\psi(\mathbf{x},t)$ , of these driven threshold systems are complex and often difficult to understand and interpret from a deterministic perspective, as these patterns are emergent processes that develop from the obscure underlying structures, parameters, and dynamics of a multidimensional nonlinear system [Nijhout et al., 1997].

Analysis of a number of these driven threshold systems is complicated by the fact that the underlying dynamics and the state variables which control the physics of the system,  $s(\mathbf{x},t)$ , are unknown and difficult to observe. The earthquake fault system is no exception. While it is not only probable, but essential, that space-time patterns and correlations exist in the variables and interactions which control earthquake dynamics,  $s(\mathbf{x},t)$  and  $\sigma(\mathbf{x},t)$ , from which the observable surface patterns and correlations,  $\psi(\mathbf{x},t)$ , arise, those patterns are difficult or impossible to observe within the earth [Scholz, 1990; Turcotte 1997].

The schematic shown in Figure 2.1 illustrates the physical problem. As the state variable  $s(\mathbf{x},t)$  at a particular location  $\mathbf{x}$  evolves in time under the deterministic dynamics  $\mathbf{D}_t$  to a value  $s(\mathbf{x},t + \Delta t)$ , the force or potential  $\sigma(\mathbf{x},t)$

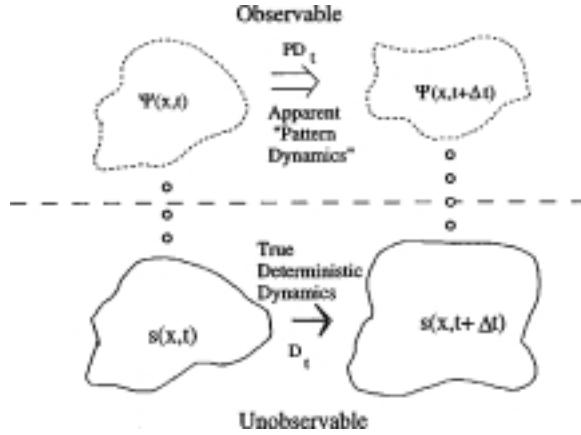


Figure 2.1: Schematic diagram of threshold systems. Note the unobservable state variables  $s(\mathbf{x}_i, t)$  and the observable seismicity pattern variables  $\psi(\mathbf{x}_i, t)$  [Rundle et al., 2000 (1)].

also evolves to  $\sigma(\mathbf{x}, t + \Delta t)$ . While the values of  $s(\mathbf{x}, t)$  and  $\sigma(\mathbf{x}, t)$ , along with the specifics of  $\mathbf{D}_t$ , are hidden from view below the dashed line of Figure 2.1, the firing patterns of  $\psi(\mathbf{x}, t)$  are observable. For example, while it is not possible to measure all the cellular potentials of the neurons of the human brain and the physical and chemical parameters which control the temporal evolution of its potentials and currents, it is possible to observe the firing patterns of neural cells [Gray, 1997; Shatz, 1997]. In the earth, there is no means at present to measure the stress and strain at every point in an earthquake fault system, or the constitutive parameters which characterize this heterogeneous medium and its dynamics. However, the seismicity which is the surface expression of its firing activity can be located in both space and time with considerable accuracy [Bakun and McEvilly, 1984; Sieh et al., 1989; Hill et al., 1990]. For example, the firing activity,  $\psi(\mathbf{x}, t)$ , can be represented as a set of time series at all positions  $\mathbf{x}$ , where  $\psi(\mathbf{x}, t) = 1$  if an event occurs in the time interval between  $t$  and  $t + \Delta t$ , and  $\psi(\mathbf{x}, t) = 0$  otherwise.

There are two possible approaches to characterizing the evolution of these systems. In the first, the observation of  $\psi(\mathbf{x}, t)$  is used to infer the behavior and

properties of  $s(\mathbf{x},t)$ ,  $\sigma(\mathbf{x},t)$  and  $\mathbf{D}_t$ .  $s(\mathbf{x},t + \Delta t)$  and  $\psi(\mathbf{x},t + \Delta t)$  are calculated from these inferred models, and used for their validation. Unfortunately, it is not clear that  $\psi(\mathbf{x},t)$ ,  $s(\mathbf{x},t)$ , and  $\sigma(\mathbf{x},t)$  are uniquely related in an earthquake fault system. There may be a large number of states  $s(\mathbf{x},t)$ , and just as large a number of associated potentials  $\sigma(\mathbf{x},t)$ , that produce a given pattern  $\psi(\mathbf{x},t)$ , and vice versa. An alternative approach to characterizing the system is to apply eigenvector decomposition techniques based on Karhunen-Loeve expansion (KLE) methods in order to construct a pattern dynamics operator,  $P(\mathbf{D}_t)$ . The firing activity pattern state  $\psi(\mathbf{x},t)$  evolves under the operation of  $P(\mathbf{D}_t)$  into a future pattern state  $\psi(\mathbf{x},t + \Delta t)$ , all of which are readily observable and quantifiable.  $P(\mathbf{D}_t)$  is constructed without any knowledge or assumptions about  $\mathbf{D}_t$ , using only information about the pattern states  $\psi(\mathbf{x},t)$  [Rundle et al., 2000 (1)].

In contrast to the underlying dynamics  $\mathbf{D}_t$ , which are probably strongly nonlinear in the case of earthquake fault systems, the pattern operator  $P(\mathbf{D}_t)$  is assumed to be linear over small time intervals  $\Delta t$ . Mean-field threshold systems are characterized by long-range interactions between sites  $\mathbf{x}$  and  $\mathbf{x}'$ , and can be treated as equilibrium systems with dynamics that are statistically stationary over long time intervals [Rundle et al., 1995; Klein et al., 1997; Egolf, 2000]. Large events that reorder the entire system occur at rare intervals throughout these time periods. The larger the interaction range, the longer the equilibrium-like time periods become, until, in the limit of infinite range interactions, the system is ergodic. The dynamical variables of the system can therefore be written as a sum of complex exponentials, a pattern dynamics approach which characterizes the system completely [Klein et al., 1996; Klein et al., 1997; Ferguson et al., 1999].

### 2.3 Earthquake Fault Systems

As discussed in Chapter 1, KL expansion methods have been used to describe and forecast other multidimensional, nonlinear systems [Hotelling, 1933; Preisendorfer, 1988; Penland, 1989; Fukunaga, 1990; Penland and Magorian, 1993; Nijhout et al., 1997]. While the underlying dynamics of such systems are frequently unknown, significant progress has been made in recent years in analyzing the space-time patterns of one particular nonlinear system, the ocean-atmosphere interface. These methods are based upon the empirical orthogonal function (EOF) analysis techniques developed for use in the atmospheric sciences [Preisendorfer, 1988].

The Karhunen-Loeve approach, the theoretical basis for EOF techniques, represents these space-time patterns as a set of eigenvectors,  $e_n(\mathbf{x}_i)$  of an equal-time correlation function, their associated time series,  $a_n(t)$ , and eigenfrequencies,  $\lambda_n$ .  $n = 1 \dots N$ , where  $N$  is the total number of locations. The eigenvectors  $e_n(\mathbf{x}_i)$  provide information about the spatial correlations of the patterns; the time series characterizes each eigenvectors temporal pattern; the eigenfrequencies  $\lambda_n$  provides information about how often they occur in the data. In a statistically stationary process, as in mean-field systems, the eigenvectors and eigenfrequencies will be independent of time [Rundle et al., 1995; Klein et al., 1997; Egolf, 2000]. The  $e_n(\mathbf{x}_i)$  and  $\lambda_n$ , in conjunction with the information contained in the most recent firing activity,  $\psi_{obs}(\mathbf{x},t)$ , can then be used to construct an equal-time correlation operator and a reconstructed pattern state,  $\psi_R(\mathbf{x},t)$ , from which a probability density function  $P(\mathbf{x},t)$  can be computed [Rundle et al., 2000 (1)]. After, a complex, linear correlation operator for the state and force variables  $s(\mathbf{x},t)$  and  $\sigma(\mathbf{x},t)$  is constructed in order to extrapolate future system behavior such as the El Nino southern oscillation [North, 1984; Hasselman, 1988; Preisendorfer, 1988; Penland,

1989; Penland and Sardeshmukh, 1995].

Several important differences distinguish the ocean/atmosphere interface from the earthquake system, however. It is not a threshold system in the sense described above, and, more importantly, the underlying dynamics are well understood and the variables that control its physics are frequently measurable and quantifiable. The variables  $s(\mathbf{x},t)$  and  $\sigma(\mathbf{x},t)$  which define the deterministic system, such as temperature, pressure, and wind speed, are directly observable within a reasonable sampling density, and the subsequent evolution is controlled by the Navier-Stokes equation. For earthquakes, such an operator acting on  $s(\mathbf{x},t)$  and  $\sigma(\mathbf{x},t)$  would be ineffective, as the field variables of stress and displacement on the fault segments are effectively invisible. However, the firing activity,  $\psi(\mathbf{x}, t)$ , is both observable and potentially quantifiable.

Average recurrence intervals for earthquake mainshocks are defined for a number of well studied faults and fault segments throughout the world, based on the observation that, over time spans of hundreds or thousands of years, earthquakes occur at semiperiodic intervals [Scholz, 1990]. One simple method for constructing a probability density for future earthquake mainshocks (without regional foreshocks or aftershocks), is based on the average frequency of events. If the average frequency of a mainshock on segment  $\mathbf{x}_i$  is  $\omega_i$ , where

$$\omega_i = \pi \frac{[\text{number of events on } \mathbf{x}_i]}{\text{time interval}} = \pi\nu_i, \quad (2.1)$$

Then, the probability,  $P(\mathbf{x}_i, t)$ , of a mainshock on  $\mathbf{x}_i$  is

$$P(\mathbf{x}_i, t) = Z^{-1} \{ \cos[\omega_i(t - t_0)] \}^2, \quad (2.2)$$

where  $t_0$  is the time since the last event, and  $Z$  is a constant of normalization [Rundle et al., 2000 [1]]. This, for example, is one interpretation of the approach

taken to forecasting the Parkfield event, which was expected to occur with a probability near 1.0 in the early 1990s. However, this and other methods have had intermittent success. As a result, dynamical computer simulations have been developed in order to study the complex nature of the earthquake fault system [Kossobokov and Carlson, 1995; Sammis et al., 1996; Eneva and Ben-Zion, 1997; Gross and Rundle, 1998; Bowman et al., 1998; Brehm and Braile, 1999; Jaume and Sykes, 1999].

Spatial patterns which seemed to recur at regular intervals in time during numerical simulations which ran for many thousand of time steps, prompted the investigation of the theoretical, mathematical basis for these patterns. The particular dynamical fault system model developed by Rundle in the late-1980s [Rundle, 1988] includes all the major faults in southern California at a relatively crude scale, as shown in Figure 2.2. In this cellular-automaton type model, 80 fault segments are each driven at an individual long-term velocity loading rate of a few cm/yr [Rundle, 1988; Rundle et al., 2000 (2)].

The *i*th segment slips when the shear stress  $\sigma_{ss,i}(t) = \mu_s \sigma_{n,i}(t)$ , where  $\mu_s$  is then coefficient of static friction, and  $\sigma_{n,i}(t)$  is the normal stress on the fault patch. At failure, slip occurs that is sufficient to reduce the shear stress to  $\sigma_{sd,i}(t) = \mu_d \sigma_{n,i}(t)$ , where  $\mu_d$  is the coefficient of dynamic friction. The difference in frictional coefficients,  $\mu_s - \mu_d$ , is tuned so that realistic sequences occur on the Big Bend and southern sections of the San Andreas (SA) fault, at intervals of  $\sim 150$  years. Additional tuning is performed to produce realistic sequences on the Imperial (I), San Jacinto (SJ), Garlock (G), and Elsinore (E) faults as well. The segments consist of vertical strike-slip faults which extend to a depth of 18 kms, in an elastic plate of 30 kms. Realistic elastic and viscoelastic interactions are included by means of the stress Green's function. Fault segment interactions permit stress transfer so that one slipping segment can trigger slip on another segment during

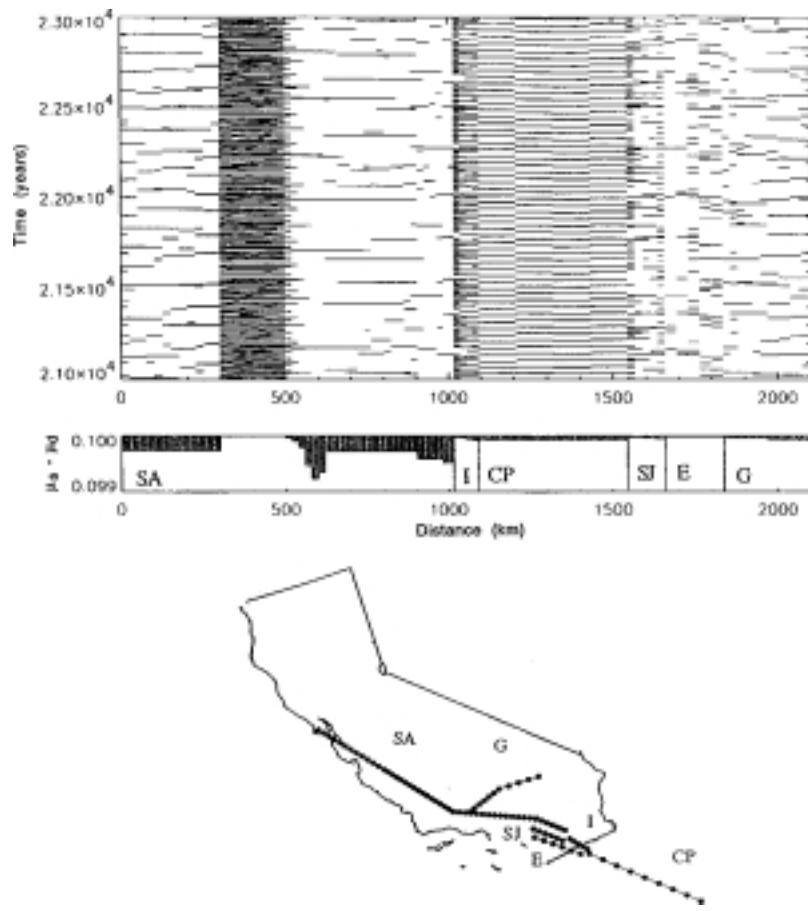


Figure 2.2: Fault patch model for southern California. Shown at the top is the time-distance plot of simulated seismic activity. The  $N = 80$  fault segments (abbreviations in text) are plotted end to end, north to the left. In the center is shown the difference between static and dynamic friction coefficients,  $\mu_s - \mu_d$ , as a function of distance along the faults. At the bottom is shown a map view of the fault systems superimposed on a map of California [Rundle, 1988].

the same event. The elastic plate is welded to a Maxwell viscoelastic halfspace in which the shear stress relaxes via viscoelastic flow [Rundle, 1988].

The success of KL decomposition techniques in identifying normal modes in such fields as the atmospheric sciences and biometrics, and the appearance of recurrent spatial and temporal patterns in the numerical simulations of realistic earthquake fault systems prompted development of a similar method for application to seismicity data [Rundle, 1988; Rundle et al., 2000 (1)]. The application of the KL decomposition, or expansion, to these numerical simulations produced eigenpatterns in space which are similar to those seen on the southern California earthquake system. Figure 2.3b, below, is the first spatial eigenvector,  $\phi_1$ , with a recurrence period of 804 years. Figure 2.3a is a plot of magnitude versus distance along the fault. This eigenvector shows a striking similarity to events along the Big Bend portion of the San Andreas fault. Figure 2.4 is the second eigenvector, with a period of 680 years. Figure 2.5 is the third eigenvector, with a period of 597 years. The results from this analysis demonstrates that the very complicated space-time firing patterns which arise from a complicated dynamical model, with time spans of many thousands of years, can be effectively decomposed into their relatively simple normal modes using KL expansion (KLE) methods [Rundle et al., 2000 (1)].

## 2.4 KL Theory

The cells of a threshold system such as an earthquake fault patch network interact over time and space, causing the development of space-time correlations and patterns of correlations [Ouchi, 1993; Nijhout, 1997]. Correlations in the state variables  $s(\mathbf{x},t)$  and  $\sigma(\mathbf{x},t)$  lead to correlations in the firing activity patterns  $\psi(\mathbf{x},t)$ . As shown by Rundle et al., 2000 (1), and detailed below, it is possible to define a variety of equal time correlation operators in order to quantify these

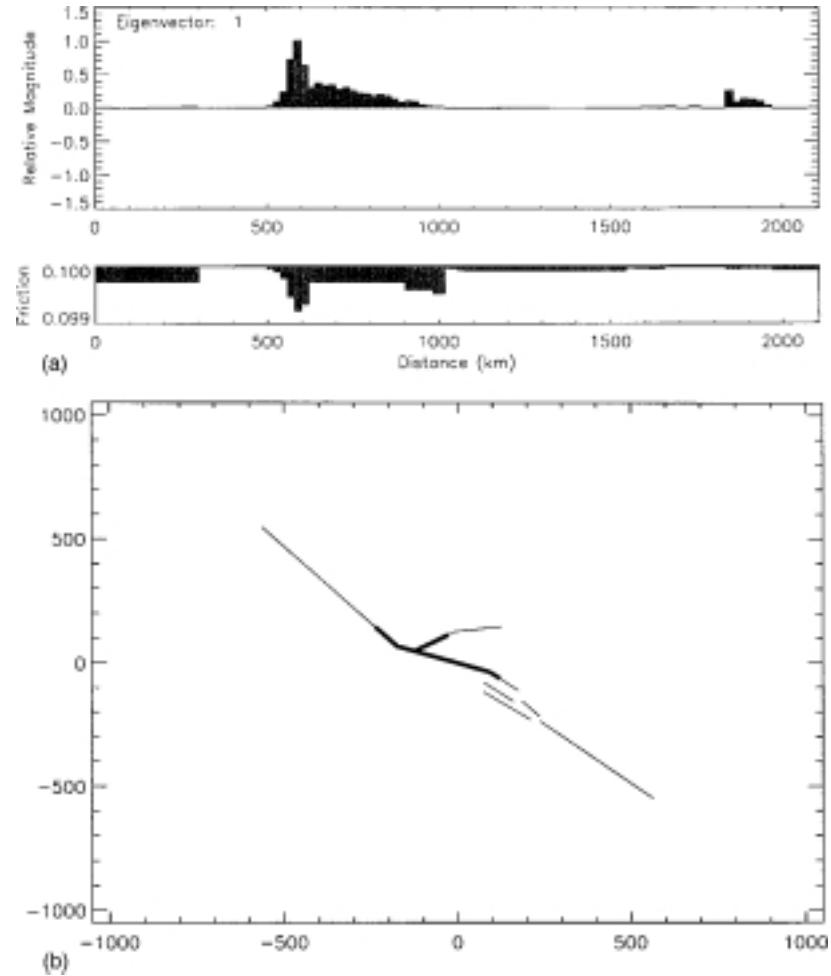


Figure 2.3: First eigenvector,  $\phi_1$ , for the space-time evolution of the fault patch model shown above. a)  $\phi_1$  is normalized for display purposes so that its maximum value = 1.0. b) Segments shown in bold are for  $\phi_1(\mathbf{x}_i) > 0.05$ , segments with values of  $\phi_1(\mathbf{x}_i) < -0.05$  are shown in dotted lines. All others are shown with a thin line [Rundle et al. 2000].

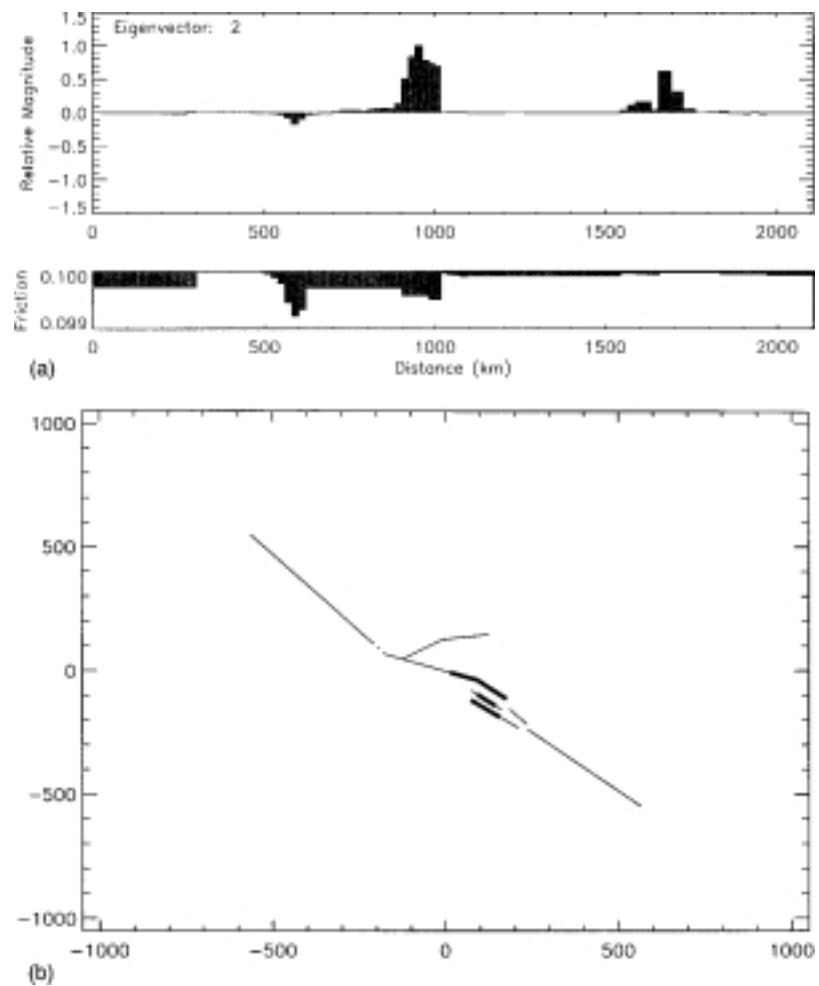


Figure 2.4: Second eigenvector,  $\phi_2$ , as in Fig. 2.3 [Rundle et al., 2000 (1)].

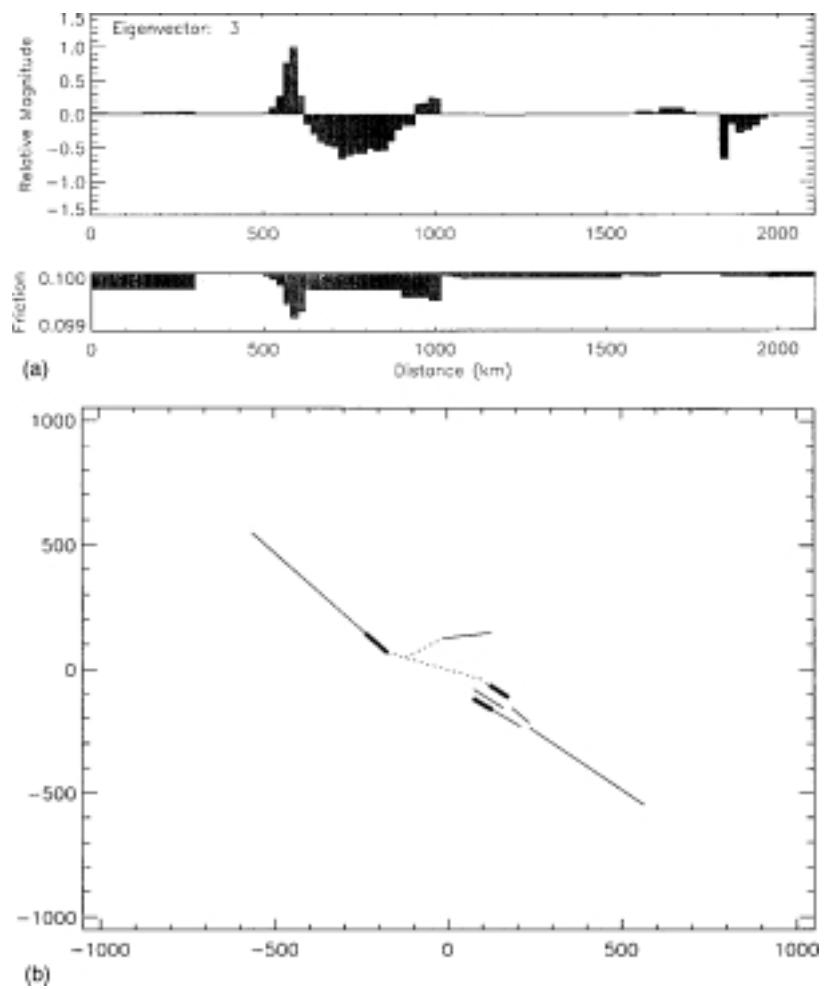


Figure 2.5: Third eigenvector,  $\phi_3$ , as in Fig. 2.3 [Rundle et al., 2000 (1)].

correlations. The most obvious is the static correlation operator,  $C(\mathbf{x}_i, \mathbf{x}_j)$ , based upon the univariate, mean zero time series of activity at each site,  $y(\mathbf{x}_i, t)$ . The eigenvalues of  $C(\mathbf{x}_i, \mathbf{x}_j)$  are units of relative probability [Vicsek, 1989; Benjamin and Cornell, 1970]. Another possibility is the rate correlation operator,  $K(\mathbf{x}_i, \mathbf{x}_j)$ , obtained using the first derivative with respect to time of  $y(\mathbf{x}_i, t)$ . The eigenvalues of  $K(\mathbf{x}_i, \mathbf{x}_j)$  are related to the squares of the frequencies,  $(\omega_n/2\pi)^2$ . Or a correlation operator could be defined using a time series  $w(\mathbf{x}_i, t)$  whose elements are not based upon unit events of ones and zeros, but upon the number of events occurring in the time interval  $t$  and  $t + \Delta t$ , and whose eigenvalues are the squares of relative frequencies. Another variation might employ a correlation operator computed from a time series  $z(\mathbf{x}_i, t)$  whose entries are the slip on segment  $\mathbf{x}_i$  in the time interval between  $t$  and  $t + \Delta t$  and whose eigenvalues are in units of slip squared. The following deals only with the first two correlation operators,  $C(\mathbf{x}_i, \mathbf{x}_j)$  and  $K(\mathbf{x}_i, \mathbf{x}_j)$  [Rundle et al., 2000 (1)].

As noted above, it is possible to define a univariate time series based upon the seismic activity at a location  $\mathbf{x}_i$ ,  $\psi(\mathbf{x}_i, t)$ , where  $\psi(\mathbf{x}_i, t) = 1$  if an event occurs in the time interval between  $t$  and  $t + \Delta t$ , and  $\psi(\mathbf{x}_i, t) = 0$  otherwise. The mean of each time series,  $\rho(\mathbf{x}_i)$ , is

$$\rho(\mathbf{x}_i) = \frac{1}{T} \sum_{all\ times} \psi(\mathbf{x}_i, t) \quad (2.3)$$

where  $T$  is the total number of time steps. The univariate, mean zero time series for each location is then

$$y(\mathbf{x}_i, t) = \frac{(\psi(\mathbf{x}_i, t) - \rho(\mathbf{x}_i))}{\sigma_i}, \quad (2.4)$$

where  $\sigma_i$  is the standard deviation of the time series  $y(\mathbf{x}_i, t)$ , such that

$$\sigma_i^2 = \frac{1}{T} \sum_{\text{all times}} (\psi(\mathbf{x}_i, t) - \rho(\mathbf{x}_i))^2. \quad (2.5)$$

The static correlation operator,  $C(\mathbf{x}_i, \mathbf{x}_j)$ , is then

$$C(\mathbf{x}_i, \mathbf{x}_j) \equiv \frac{1}{T} \int_0^T y(\mathbf{x}_i, t) y(\mathbf{x}_j, t) dt. \quad (2.6)$$

The rate correlation matrix,  $K(\mathbf{x}_i, \mathbf{x}_j)$ , is

$$K(\mathbf{x}_i, \mathbf{x}_j) \equiv -\frac{1}{T} \int_0^T \frac{\partial y(\mathbf{x}_i, t)}{\partial t} \frac{\partial y(\mathbf{x}_j, t)}{\partial t} dt. \quad (2.7)$$

As the elements of the time series  $\psi(\mathbf{x}_i, t)$  and  $y(\mathbf{x}_i, t)$  consist of impulse functions, the computation of the derivatives of  $y(\mathbf{x}_i, t)$  for  $K(\mathbf{x}_i, \mathbf{x}_j)$  is difficult. However, the computation of  $C(\mathbf{x}_i, \mathbf{x}_j)$  is simple. The alternative approach to computing  $K(\mathbf{x}_i, \mathbf{x}_j)$  is based on the Fourier transform of  $y(\mathbf{x}_i, t)$ ,

$$y(\mathbf{x}_i, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{y}(\mathbf{x}_i, \varpi_i) e^{i\varpi_i t} d\varpi_i. \quad (2.8)$$

Then,

$$C(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{2\pi T} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^T e^{i(\varpi_i + \varpi_j)t} \hat{y}(\mathbf{x}_i, \varpi_i) \hat{y}(\mathbf{x}_j, \varpi_j) d\varpi_i d\varpi_j dt. \quad (2.9)$$

If  $P_*(\mathbf{x}_i, \mathbf{x}_j, \varpi_i, \varpi_j)$  is the joint probability density function of the two variables  $\varpi_i$  and  $\varpi_j$ ,

$$P_*(\mathbf{x}_i, \mathbf{x}_j) = (\hat{y}(\mathbf{x}_i, \varpi_i) \hat{y}(\mathbf{x}_j, \varpi_j)) \times \left\{ \frac{1}{2\pi T} \int_0^T e^{i(\varpi_i + \varpi_j)t} dt \right\}, \quad (2.10)$$

then

$$C(\mathbf{x}_i, \mathbf{x}_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_*(\mathbf{x}_i, \mathbf{x}_j, \varpi_i, \varpi_j) d\varpi_i d\varpi_j, \quad (2.11)$$

and

$$K(\mathbf{x}_i, \mathbf{x}_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_*(\mathbf{x}_i, \mathbf{x}_j, \varpi_i, \varpi_j) \varpi_i \varpi_j d\varpi_i d\varpi_j. \quad (2.12)$$

The covariance  $\langle (\varpi_i - \langle \varpi_i \rangle)(\varpi_j - \langle \varpi_j \rangle) \rangle$  is then

$$\langle (\varpi_i - \langle \varpi_i \rangle)(\varpi_j - \langle \varpi_j \rangle) \rangle = \langle \varpi_i \varpi_j \rangle - \langle \varpi_i \rangle \langle \varpi_j \rangle, \quad (2.13)$$

where

$$\langle \varpi_i \varpi_j \rangle = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_*(\mathbf{x}_i, \mathbf{x}_j, \varpi_i, \varpi_j) \varpi_i \varpi_j d\varpi_i d\varpi_j}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_*(\mathbf{x}_i, \mathbf{x}_j, \varpi_i, \varpi_j) d\varpi_i d\varpi_j} = \frac{K(\mathbf{x}_i, \mathbf{x}_j)}{C(\mathbf{x}_i, \mathbf{x}_j)}. \quad (2.14)$$

If the joint spectra of the time series are tightly peaked around the average frequencies  $\langle \varpi_i \rangle$  and  $\langle \varpi_j \rangle$ , then

$$\left| \frac{\langle (\varpi_i - \langle \varpi_i \rangle)(\varpi_j - \langle \varpi_j \rangle) \rangle}{\langle \varpi_i \rangle \langle \varpi_j \rangle} \right| \ll 1. \quad (2.15)$$

and

$$\frac{K(\mathbf{x}_i, \mathbf{x}_j)}{C(\mathbf{x}_i, \mathbf{x}_j)} \approx \langle \varpi_i \rangle \langle \varpi_j \rangle. \quad (2.16)$$

The frequencies  $\langle \varpi_i \rangle$  are observable, they are simply, as in Equation 2.1 above,

$$\omega_i = \pi \nu_i,$$

with the result that

$$K(\mathbf{x}_i, \mathbf{x}_j) \approx C(\mathbf{x}_i, \mathbf{x}_j) \langle \varpi_i \rangle \langle \varpi_j \rangle. \quad (2.17)$$

Once  $C(\mathbf{x}_i, \mathbf{x}_j)$  is found from Equation 2.6, it can be diagonalized to obtain the normal modes of the patterns, the eigenvalues,  $\lambda_n$  and  $e_n(\mathbf{x}_i)$ . The associated time series,  $a_n(t)$ , are the time dependent expansion coefficients for each

eigenvector, and are reconstructed by multiplying the original data matrix by the eigenvectors, i.e.

$$a_n(t) = \sum_{n=1}^N y(\mathbf{x}_i, t) e_n(\mathbf{x}_i). \quad (2.18)$$

Multiplying the KLE value for a particular box by the accompanying PC amplitudes will give the time series for that particular location accounted for by that particular mode. Summing those values for all the modes for that particular location will return the original time series value for that particular location, either a 0 or a 1. In effect, the eigenmodes at each location, squared, are the percent of the total signal accounted for by those particular spatial correlations, or the probability of an event occurring at one site given that an event occurs at another site [Rundle et al., 2000 (1)]. The eigenvectors provide detailed information about the spatial correlations between the  $N$  sites.

The calculation of  $K(\mathbf{x}_i, \mathbf{x}_j)$  follows from  $C(\mathbf{x}_i, \mathbf{x}_j)$  using Equation 2.17. Its eigenvalues,  $\varpi_n$ , and eigenvectors,  $\phi_n(\mathbf{x}_i)$ , are obtained by diagonalization of  $K(\mathbf{x}_i, \mathbf{x}_j)$ . The eigenvectors again provide information about the specific patterns of spatial correlation between the sites, and the eigenvalues, in this case eigenfrequencies, provide information about how often they recur.

The complex-valued, reconstructed pattern state,  $\psi_R(\mathbf{x}_i, t)$ , computed from the  $N$  eigenvectors and eigenfrequencies of the rate correlation operator is

$$\psi_R(\mathbf{x}_i, t) = \sum_{i=1}^N \alpha_i e^{-i\omega t} \phi_i(\mathbf{x}_i). \quad (2.19)$$

The real, observable part of  $\psi_{obs}(\mathbf{x}_i, t)$  is

$$\psi_{obs}(\mathbf{x}_i, t) = \frac{1}{\sqrt{2}} \{ \psi_R(\mathbf{x}_i, t) + \psi_R^*(\mathbf{x}_i, t) \} = \sum_{n=1}^N \beta_n \phi_n(\mathbf{x}_i) \cos(\omega_n t). \quad (2.20)$$

The constants  $\beta_n$  can be determined by fitting  $\psi_R(\mathbf{x}_i, t)$  to the time of the most recent slip event,  $t_{mi}$ , on segment  $\mathbf{x}_i$ ,

$$\psi_{obs}(\mathbf{x}_i, t_{0i}) = (-1)^m, \quad (2.21)$$

where  $m$  represents the index of each event at site  $\mathbf{x}_i$ , with the first event beginning at  $m = 0$ .

As the correlation operator,  $K(\mathbf{x}_i, \mathbf{x}_j)$ , is directly related to a probability function [Vicsek, 1989],  $\psi_R(\mathbf{x}_i, t)$  is the square root of a probability, and the probability amplitude,  $P(\mathbf{x}_i, t + \Delta t)$ , is

$$P(\mathbf{x}_i, t + \Delta t) = |\psi_R(\mathbf{x}_i, t + \Delta t)|^2 = \{\psi_{obs}(\mathbf{x}_i, t + \Delta t)\}^2. \quad (2.22)$$

Equations 2.20 and 2.21 lead to a set of  $N \times N$  set of independent equations for  $\beta_n$  at each time step, which can be solved using standard methods. The entire process is repeated iteratively in order to calculate  $P(\mathbf{x}_i, t + 2\Delta t)$ , and so forth, taking care to always use the most recent activity data to calculate  $\beta_n$ .

This forecasting technique was applied to the model and decomposition shown in Figures 2.2 through 2.5, and the resulting probability calculated according to Equation 2.22. Figure 2.6 is a superposition of model events from the actual model evolution on a contour plot of the probabilities,  $P(\mathbf{x}_i, t)$ , for a 500 year model time period. It can be seen that, in many cases, there is reasonable agreement between probabilities and both time and location of event, although the false alarm rate is high [Rundle et al., 2000 (1)].

Building on these methods for analyzing nonlinear threshold systems, space-time seismicity patterns can be identified in both numerical simulations of realistic earthquake models for southern California and actual historic seismicity records [Tiampo et al., 1998; Rundle et al., 2000(1)]. Here I apply this pattern dynamics technique to historic seismicity data from southern California in order to study the eigenvectors, or eigenpatterns, associated with moderate to large earthquakes.

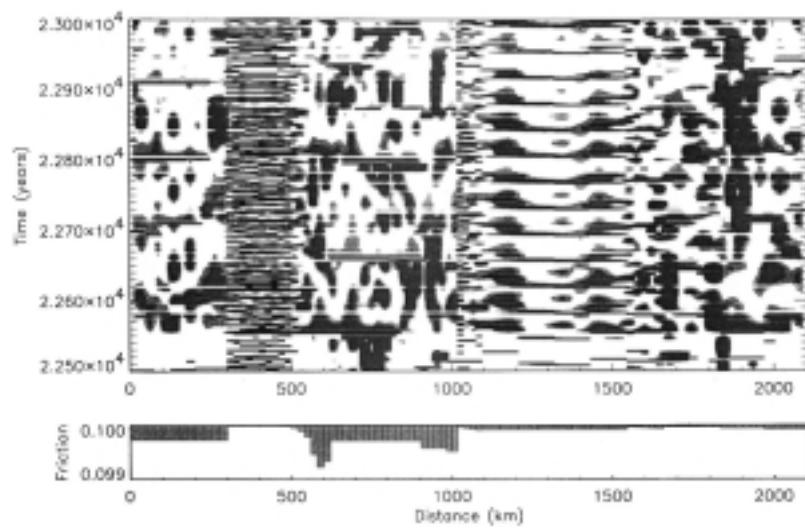


Figure 2.6: Superposition of events (solid horizontal lines) for the simulation time year 22500 to year 23000 shown at the top, with probabilities  $P(\mathbf{x}_i, t)$  calculated according to equation 2.22 above. Light shaded regions are for  $P(\mathbf{x}_i, t) > 0.3333$ ; dark regions are for  $P(\mathbf{x}_i, t) > 0.5$  [Rundle et al., 2000 (1)].

## Chapter 3

### KL Decomposition of Southern California Seismicity

#### 3.1 Introduction

Karhunen-Loeve decomposition, or expansion, methods can be used to define a unique, complete pattern basis set for a given dynamical system [Holmes et al., 1996]. For driven threshold systems, an adaptation of these KLE methods can be employed to characterize both the space-time patterns of threshold transitions, i.e. “firings”, as well as the underlying, usually unobservable Markov variables that define the dynamics [Tiampo et al., 1998; Rundle et al., 2000 (1)]. In either case, the patterns are defined by the eigenstates and eigenvalues of one of an appropriately constructed family of correlation operators. In the method of Rundle et al. 2000 (1), analysis of the time evolution of the patterns can be facilitated, and perhaps forecast, by reconstructing a pattern state, using the basis eigenvectors and eigenvalues obtained from the space-time correlation operator for that system. While the Karhunen-Loeve expansion (KLE) decomposes a nonlinear system into linear spaces, it also implicitly assumes that the underlying process is both Markov and stationary in time. These KLE basis functions are the physical expression of the underlying correlation or pattern dynamics of the system [Fukunaga, 1970; Holmes et al., 1996].

Here I apply this Pattern Dynamics method [Tiampo et al., 1998; Rundle et al., 2000 (1)] to the analysis of observed seismicity data from southern California in

order to identify basis patterns for all possible space-time seismicity configurations. These basis states are a complete, orthonormal set of eigenvectors and associated eigenvalues that are obtained from the diagonalization of the correlation operators computed for the regional historic seismicity data.

### 3.2 KL Matrix Decomposition Methodology

Rundle et al. 2000 (1), extended the standard KLE methods to include the construction of pattern states that can be used to forecast events in time, in much the same manner as EOF analysis is used to predict El Niño events in meteorology [Preisendorfer, 1988; Penland, 1989; Garcia and Penland, 1991]. This 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,  $a_j(t)$ .

The KL expansion is obtained from the  $p$  time series that record the deformation history at particular locations in space, where each time series consists of  $n$  time steps,  $i = 1, \dots, n$ ,  $y(x_s, t_i) = y_i^s$ , and  $s = 1, \dots, p$ . The goal is to construct a time series for each of a large number of locations that records, for a given short period of time, whether an earthquake occurred at that location (value = 1) or did not occur (value = 0). If, for example, the time interval was decimated into units of 0.001 years, approximately eight hour time segments, the result would be a time series of 1000 time steps for every year of data, with either a zero or a one at each time step. These time series are incorporated into a matrix,  $T$ , consisting

of time series of the same measurement for  $p$  different locations, i.e.

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

$T$  is therefore an  $n \times p$  matrix of real values [Fukunaga, 1970].

The covariance matrix,  $S(x_i, x_j)$ , for these events is formed by multiplying  $T$  by  $T^T$ , where  $S$  is a  $p \times p$  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  $y(x_j, t)$ , as follows.

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

$$C = \begin{bmatrix} \frac{s_{11}}{\sigma_1 \sigma_1} & \frac{s_{12}}{\sigma_1 \sigma_2} & \dots & \frac{s_{1p}}{\sigma_1 \sigma_p} \\ \frac{s_{21}}{\sigma_2 \sigma_1} & \frac{s_{22}}{\sigma_2 \sigma_2} & \dots & \frac{s_{2p}}{\sigma_2 \sigma_p} \\ \vdots & \vdots & \dots & \vdots \\ \frac{s_{p1}}{\sigma_p \sigma_1} & \frac{s_{p2}}{\sigma_p \sigma_2} & \dots & \frac{s_{pp}}{\sigma_p \sigma_p} \end{bmatrix}$$

This equal-time correlation operator,  $C(x_i, x_j)$ , is decomposed into its eigenvalues and eigenvectors in two parts. The first employs the tri reduction technique to reduce the matrix  $C$  to a symmetric tridiagonal matrix, using a Householder reduction. The second part uses a ql algorithm to find the eigenvalues,  $\lambda_j^2$ , and eigenvectors,  $e_j$  of the tridiagonal matrix [Press et al., 1992]. These eigenstates thus represent the orthonormal basis vectors arranged in order of decreasing correlation, and reflect the relative importance of the various modes over the time interval of interest. Dividing the corresponding eigenvalues,  $\lambda_j^2$ , by the sum of the eigenvalues, yields that percent of the correlation accounted for by that particular mode. The associated orthonormal time series can be reconstructed by projecting

the initial data set onto these basis vectors [Preisendorfer, 1988; Holmes et al., 1996]. The time dependent expansion coefficients,  $a_j(t)$ , which represent temporal eigenvectors, are reconstructed by multiplying the original data matrix by the eigenvectors, i.e.

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

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

### 3.3 Seismicity Data

The primary seismicity data set for southern California employed in this analysis is the entire Caltech catalog from 1932 through August of 1999, obtained from the SCEC database, with all blast events specifically removed from the catalog [[http://www.scec.org.](http://www.scec.org/)]. A small segment of the catalog, for year 1978, is shown in Appendix C, Section C.1. Relevant data includes location, in latitude and longitude, and the time the event occurred. Seismic events between  $-115^\circ$  and  $-123^\circ$  longitude and  $32^\circ$  and  $37^\circ$  latitude were selected, and all quality events were acquired. Separate analyses were performed for the entire data set, consisting of all events of magnitude greater than or equal to 0.0, and on another data set in which only those events of magnitude greater than or equal to 3.0 were included in the binning process described below. The time periods evaluated were from 1932 to 1978, 1932 to 1991, and 1932 to June of 1998. In all cases, the seismicity was then binned into squares of  $0.1^\circ$  latitude and  $0.1^\circ$  longitude to a side, and a time series constructed for each location square, boxes of approximately 11 km to a side. Each time step is given an initial value of 1.0 if one or more events occurs

in that time period, or a value of 0.0 otherwise. Subsequently, the mean for each time series is removed from the data.

For the time period 1932 to 1998, the time interval first was decimated into units of 0.001 years, approximately eight hour time segments, resulting in a total of 66,500 time steps. This particular analysis includes only those locations in which at least one event occurs over the entire time series. In addition, the total area used was slightly smaller. A subset of the data which included only those seismic events between  $-115^{\circ}$  and  $-121^{\circ}$  longitude and  $32.5^{\circ}$  and  $36.5^{\circ}$  latitude was selected, and only events of quality A or B were accepted. The number of locations affected by the seismicity,  $p$ , therefore varies with the cutoff and ranges from approximately 800 to 1350.

A slightly longer time period, 1932 through August of 1999, was also analyzed using the entire data set, including the larger areal extent and events of all quality. The time interval for this second decomposition was increased to one day, so that the total number of time steps is approximately 24333. In addition, all locations from the entire database, and all quality events, were included, even those where no event occurred for more than 67 years. The number of location time series therefore remains constant at 3162.

For the time period 1932 to 1991, the time interval for the analysis was again one day, so that the total number of time steps is approximately 21535. Again, all locations, and all quality events, were included, even those where no event occurred for the entire 59 years. Again, the number of location time series therefore remains constant at 3162.

The same conditions were held for the time period 1932 to 1978, so that the number of locations,  $p$ , remains at 3162, while the total number of time steps is approximately 16,790.

In addition, the database compiled by Deng and Sykes, 1996, for large earth-

quakes in southern California, was decomposed as well. This database, attached in Appendix C, Section C.2, covers the time period 1812 to 1994, with events of magnitude greater than 6.0 only. Section C.3 is a list of all events greater than magnitude 5.0, 1933 through 1995, for comparison in the analyses below [<http://www.gps.caltech.edu/~jishu>].

### 3.4 Results

#### 3.4.1 Time Period: 1932 through June, 1998.

For the time period 1932 to 1998, the time interval first was decimated into units of 0.001 years, for a total of 66,500 time steps. Additionally, only those locations in which at least one event occurs over the time entire time series were included, and the spatial extent ranged from  $-115^{\circ}$  to  $-121^{\circ}$  longitude and  $32.5^{\circ}$  to  $36.5^{\circ}$  latitude, amounting to approximately 1350 sites.

Figure 3.1 is a plot of the eigenvalues,  $\lambda_i^2$ , versus mode for the KLE decomposition of the entire southern California data set, normalized to the sum of all the eigenvalues. It should be noted that while the first five modes explain the greatest correlations in the data, they only account for approximately 12.5 percent of the total, and there is little difference between the remaining modes. This is most likely due to the incompleteness of the data set relative to the actual lifetime of southern California seismicity. As a result of the small variance between the lower eigenvalues, their contribution to the total signal may continue to be significant. Figure 3.2 shows the first two KLE modes (b,d) for the analysis performed on the entire data set (no cutoff or summation), and their associated principal component (PC) time series (a,c). The KLE results have been normalized by their absolute maximum or minimum value. The color scale represents the correlation between locations - red values can be thought of as positively correlated, or “on”, while

blue values are anticorrelated, or “off”. The ordinate of the PC plot reflects that portion of the amplitude in the data which can be accounted for by that particular mode. Multiplying the KLE value for a particular box by the accompanying PC amplitudes will give the time series for that particular location accounted for by that particular mode. Note that the first mode, Figure 3.2b, is the expression of better detection limits over time, the ability to detect more smaller events with the development of more sensitive equipment.

Figure 3.2d shows the second KLE mode. Here the region surrounding the 1992 Landers event is “on” (red) whereas the rest of the southern San Andreas fault system is “off” (blue), including the San Jacinto fault. The Northridge, San Fernando and Coalinga thrust events are also “off”, as well as the eastern and western sides of the Sierra Nevada fault zone. One can detect the Big Bear and Joshua Tree events clearly. Visible in the PC time series (3.2c) are the 1971 San Fernando earthquake, the 1979 Imperial Valley earthquake, the 1984 Coalinga earthquake, the 1986 North Palm Springs event, the 1987 Superstition Hills event, and the 1994 Northridge earthquake.

Figure 3.3 shows the third (b) and fourth (d) KLE modes for the entire data set. In Figure 3.3b, the Northridge and San Fernando events are “on”. The main Landers shock is “on”, but the Joshua Tree and Big Bear portions of that sequence are “off”. In addition, the Coalinga and North Palm Springs events are also “off”.

The fourth KLE mode displays a spike in the decomposition, probably an artifact of the matrix decomposition process.

The fifth KLE mode in Figure 3.4 suggests that the Coalinga event of 1984 is anticorrelated with the 1979 Imperial Valley and the 1987 Superstition Hills events, which are correlated with each other. In addition, the Whittier Narrows earthquake of 1987 can be seen at  $242^{\circ}$  longitude,  $34^{\circ}$  latitude.

The seventh through tenth KLE modes, Figure 3.5, are increasingly noisier,

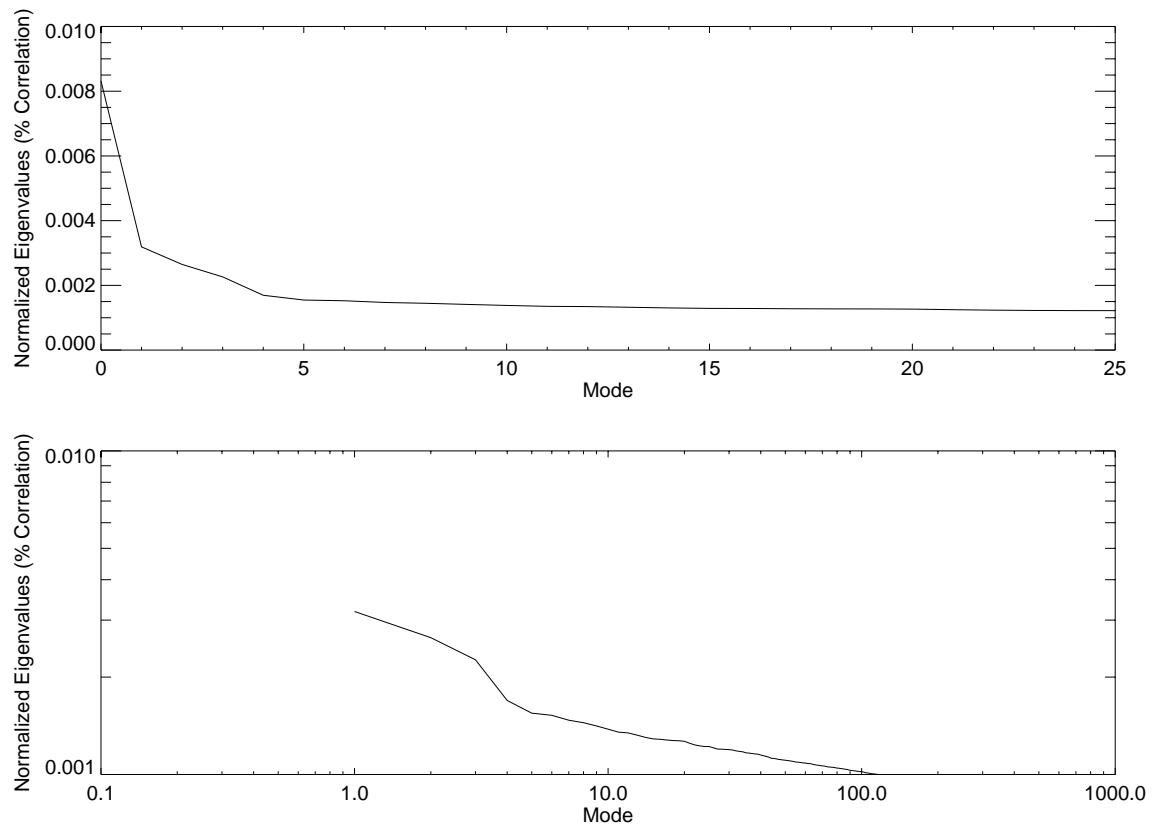


Figure 3.1: Eigenvalues,  $\lambda_i^2$ , normalized by the sum of all eigenvalues, versus mode for the KLE decomposition of the entire southern California data set, 1932-1998.

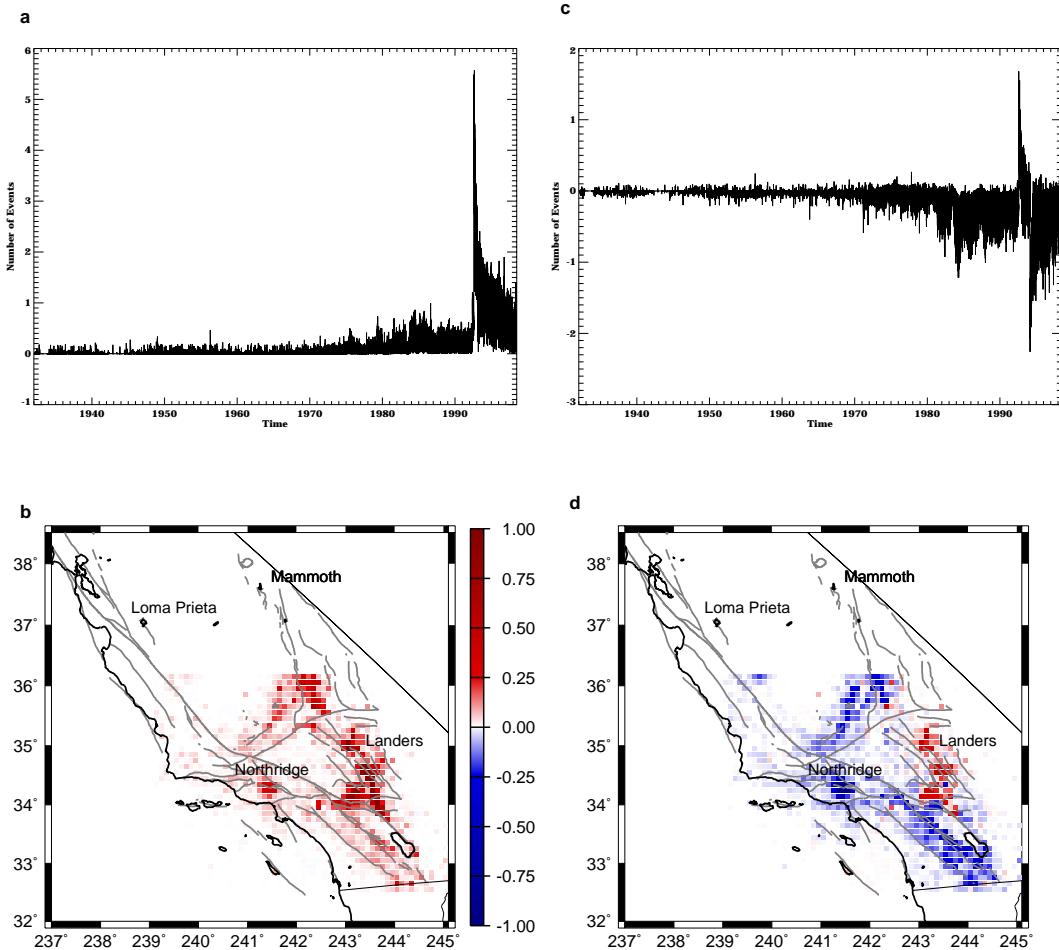


Figure 3.2: First two KLE modes for southern California seismicity, 1932-1998. a) PC time series for first KLE mode; b) first KLE mode, normalized to maximum; c) PC time series for second KLE mode; and d) second KLE mode, also normalized to the maximum.

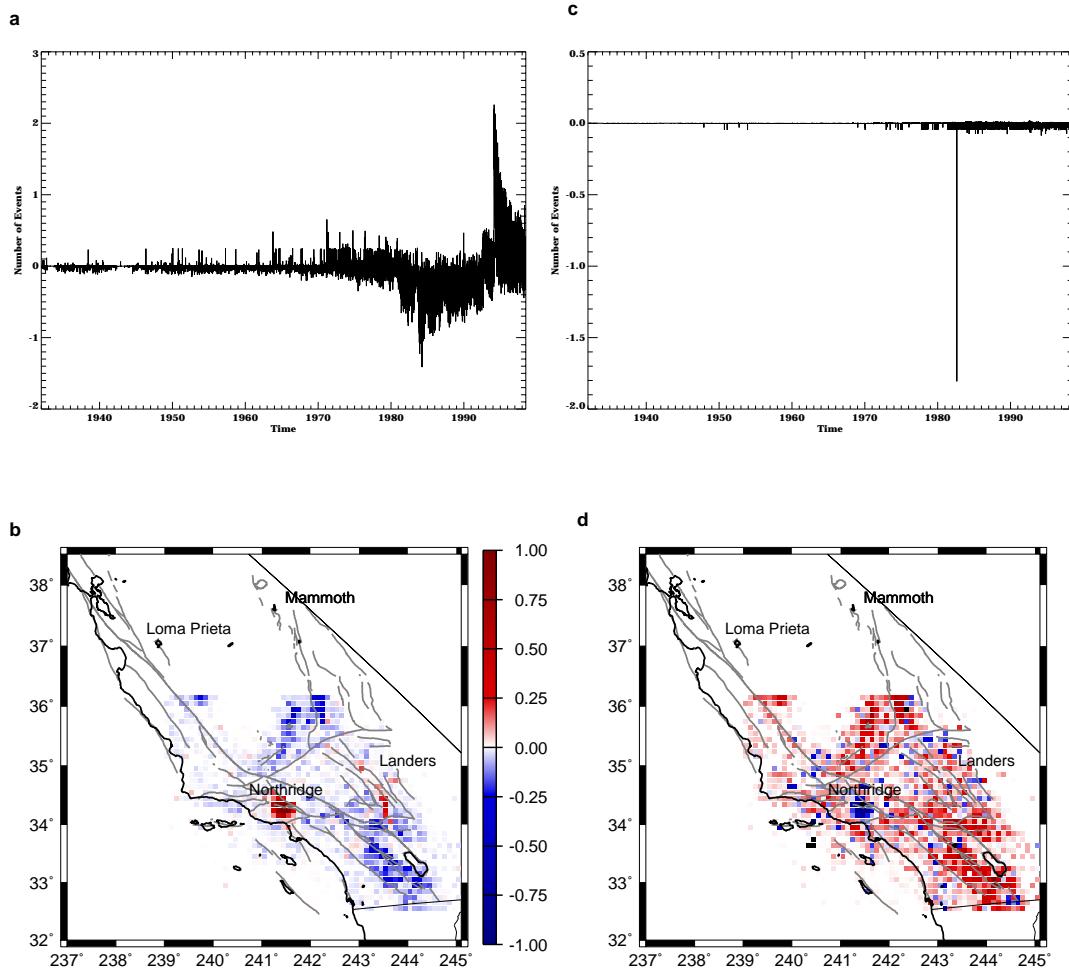


Figure 3.3: Third and fourth KLE modes for southern California seismicity, 1932–1998. a) PC time series for third KLE mode; b) third KLE mode, normalized to maximum; c) PC time series for fourth KLE mode; and d) fourth KLE mode, also normalized to the maximum.

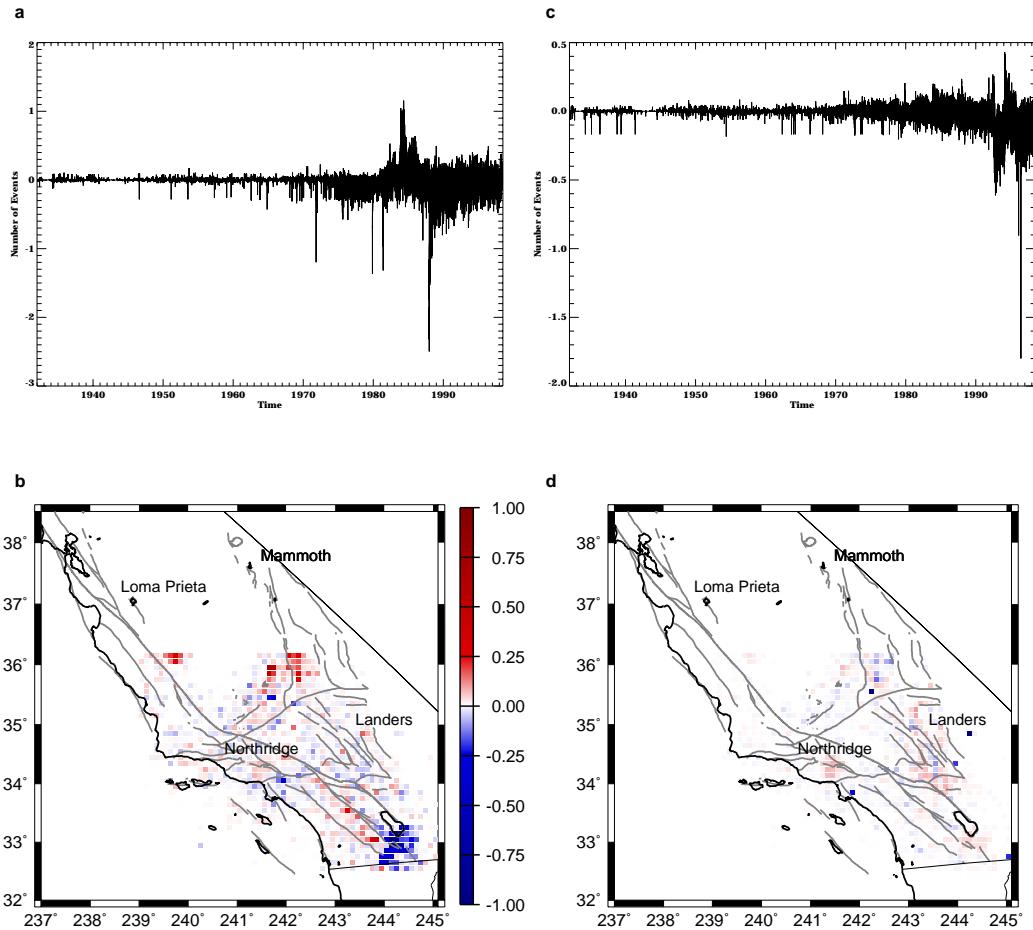


Figure 3.4: Fifth and sixth KLE modes for southern California seismicity, 1932-1998. a) PC time series for fifth KLE mode; b) fifth KLE mode, normalized to maximum; c) PC time series for sixth KLE mode; and d) sixth KLE mode, also normalized to the maximum.

and also on a smaller spatial scale than the first four modes, as would be expected in an eigenvector decomposition ordered by decreasing eigenvalue. Some signal is still apparent in these modes, however, and, while the trend of increasing noise with increasing mode number continues, the value of these lower modes cannot be discounted. To illustrate this point, Figure 3.6 shows KLE mode 19, where the 1968 Santa Barbara Channel event is clearly visible.

All the modes shown in these figures are elements of an orthonormal basis set of the dominant signals in the data. The first mode, which has the Landers event firing and everything else “off”, is orthogonal to the second mode, and all remaining modes. In addition, one would expect to see harmonics of these modes in some of the lower modes, as the expansion accounts for smaller wavelength variations. A number of the lower modes in this expansion appear to be just such higher order harmonics of earlier modes. Examples are KLE modes 7, 8, and 9. When compared to KLE3, they consist of similar patterns at smaller spatial scales. Finally, modes associated with the Landers event dominate the first ten modes of this analysis.

A similar analysis on this same data set but with a cutoff applied at events of magnitude 3.0, as described above, was accomplished in order to reduce the noise level in the data and the dominance of the Landers aftershock sequence. Figure 3.7 shows much cleaner visualizations of the first few modes - two Landers modes, Northridge and Imperial Valley. Note the absence of the first mode from the analysis above. Applying a cutoff of magnitude 3.0 has effectively removed the detectability mode. In addition, a number of earlier, more poorly instrumented events emerge in this analysis. Figure 3.8 shows the seventh through tenth KLE modes for this cutoff analysis. In Figure 3.8a, the seventh KLE mode shows the 1987 Superstition Hills and 1979 Imperial Valley events anticorrelated with each other. The eighth KLE mode shown in Figure 3.8b clearly depicts the Kern

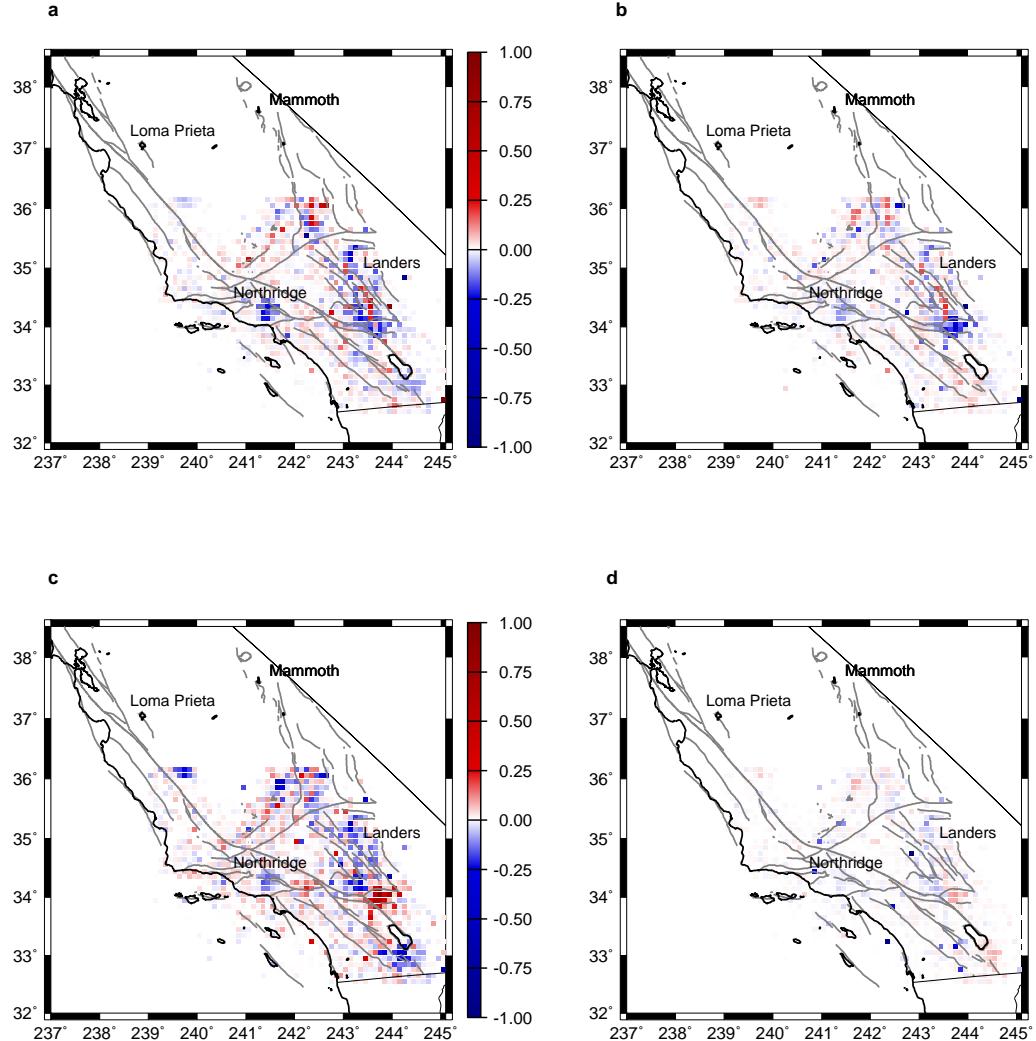


Figure 3.5: Seventh through tenth KLE modes for southern California seismicity, 1932-1998. a) Seventh KLE mode; b) eighth KLE mode; c) ninth KLE mode; and d) tenth KLE mode, all normalized to the maximum.

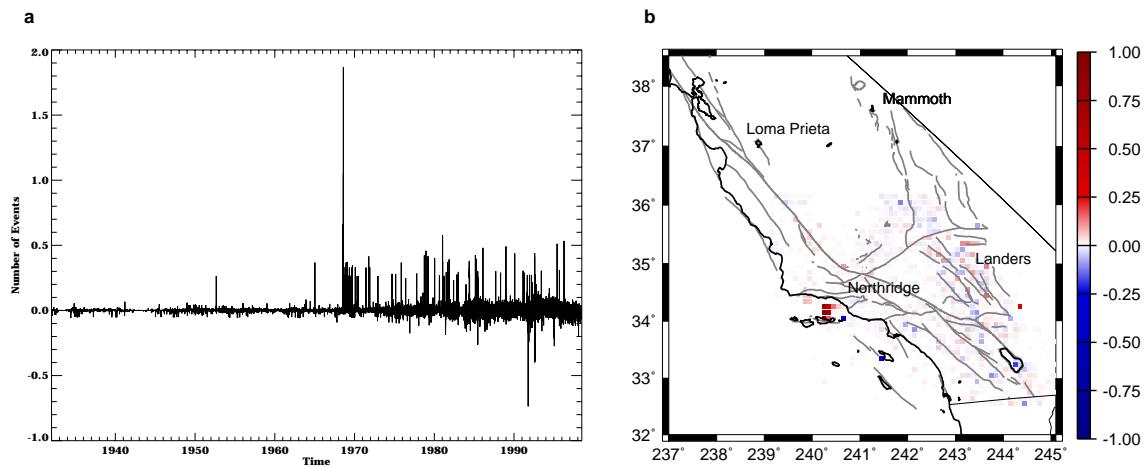


Figure 3.6: Nineteenth KLE mode for southern California seismicity, 1932-1998.  
 a) PC time series for nineteenth KLE mode; and b) nineteenth KLE mode, normalized to maximum.

County earthquake sequence of 1952, an event which never appears in the earlier analysis of the entire data set.

The 1971 San Fernando event can be differentiated from the 1994 Northridge event in this cutoff analysis, as shown in Figures 3.8c and 3.8d. The ninth KLE mode, in Figure 3.8c, includes not only those two events, but also the 1992 Landers sequence, as well as the 1986 North Palm Springs event. Figure 3.8d shows the San Fernando and Northridge earthquakes, anticorrelated with each other, and the San Fernando and the North Palm Springs correlated together.

### **3.4.2 Time Period: 1932 through August, 1999.**

A slightly longer time period, 1932 through August of 1999, was analyzed using the entire data set, including the larger areal extent and events of all quality. The time interval for this second decomposition was increased to one day, so that the total number of time steps is approximately 24333. In addition, every box location within the areal limits noted above was included in the analysis, even those where no event occurred for more than 67 years. The number of time series corresponds to the number of locations, 3162.

Figure 3.9a is a plot of the first 25 normalized eigenvalues, while Figure 3.9b is the first 1000 normalized eigenvalues, plotted on a log-normal scale. To illustrate that the modes below 100 are essentially noise, Figure 3.10 shows the 500th KLE mode, which appears to be essentially uncorrelated. The normalized eigenvalues have dropped below 0.0005, or 0.05 percent for these modes.

KLE modes one and two are shown in Figure 3.11, with their accompanying PC time series. Again, the first mode is the expression of better detection limits over time, and its PC time series displays that distinctive wave associated with the Landers sequence and its large numbers of aftershocks, punctuated by the occurrence of the Northridge event.

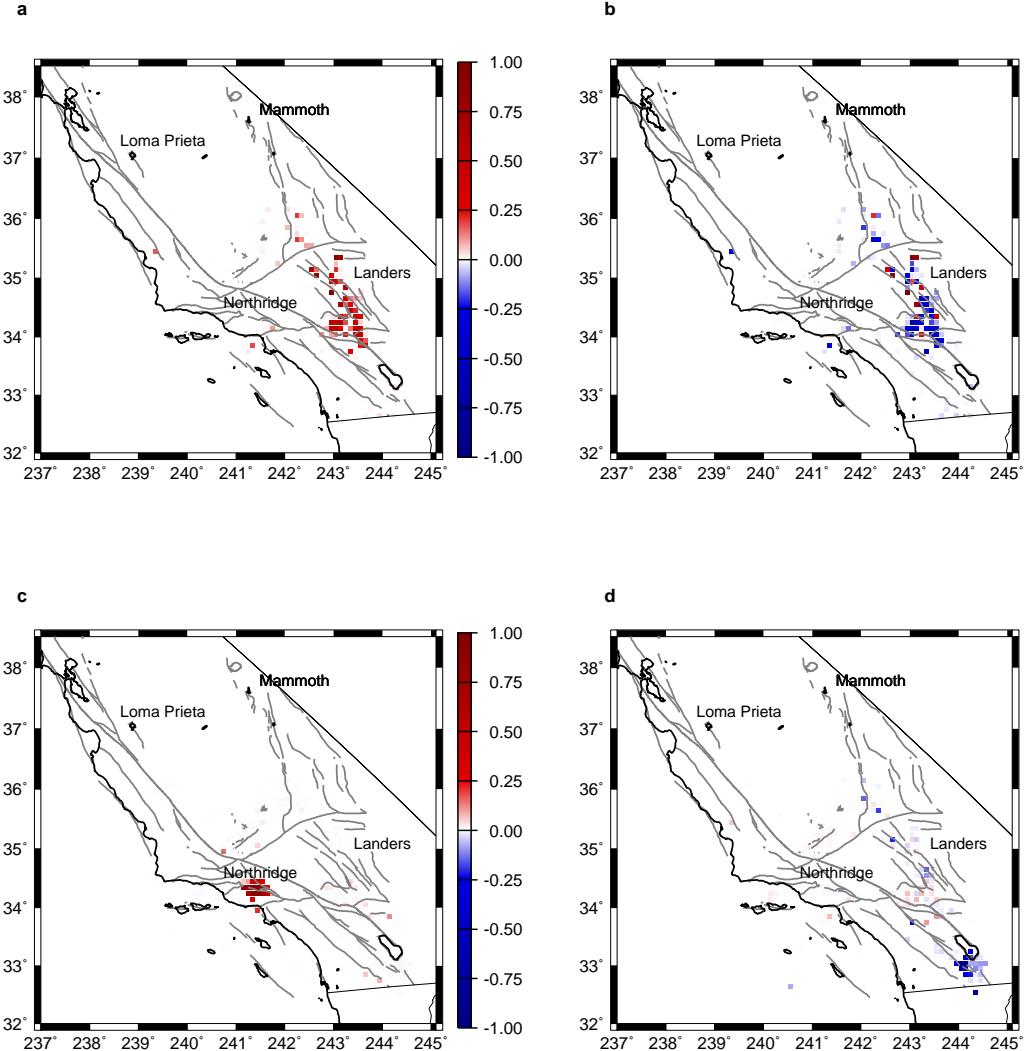


Figure 3.7: First through fourth KLE modes for southern California seismicity, 1932-1998,  $M > 3.0$ . a)First KLE mode; b) second KLE mode; c) third KLE mode; and d) fourth KLE mode, all normalized to the maximum.

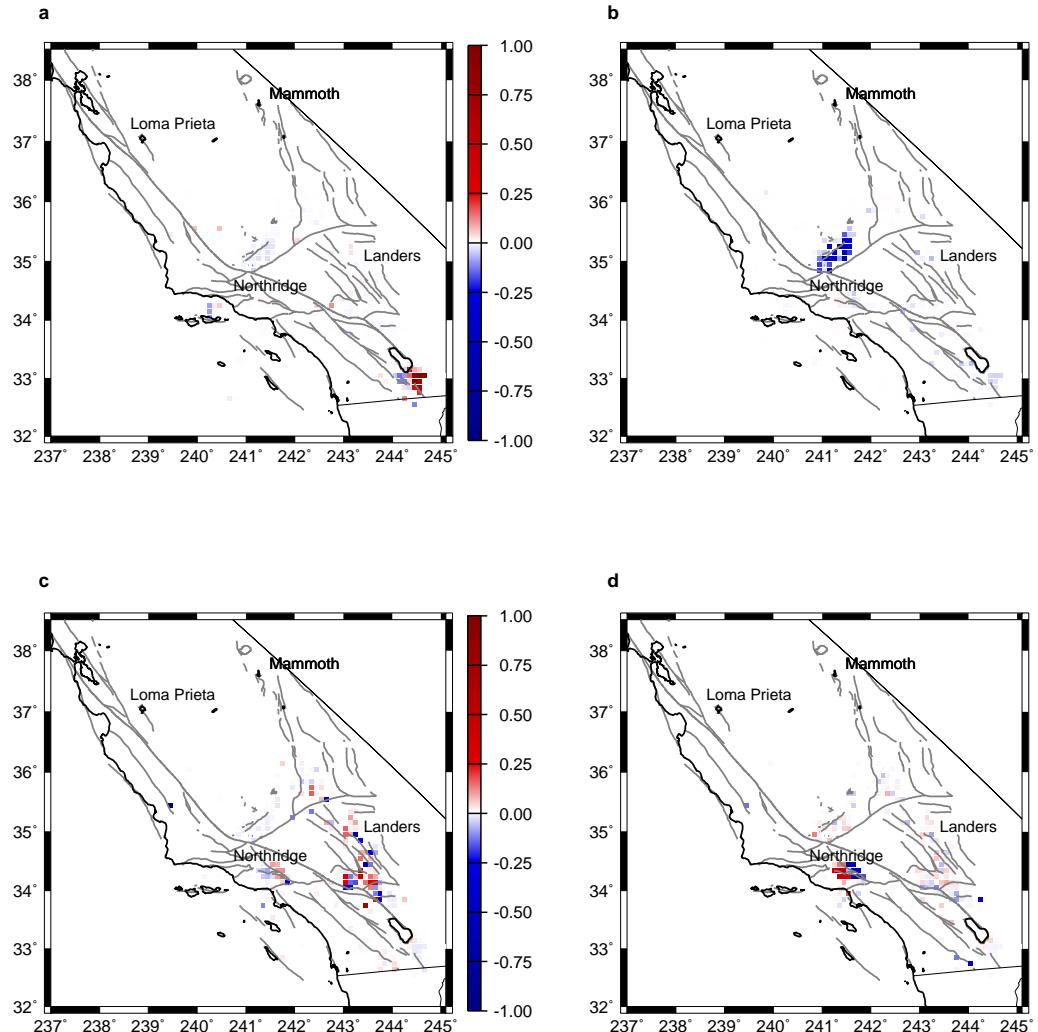


Figure 3.8: Seventh through tenth KLE modes for southern California seismicity, 1932-1998,  $M > 3.0$  a) Seventh KLE mode; b) eighth KLE mode c) ninth KLE mode; and d) tenth KLE mode, all normalized to the maximum.

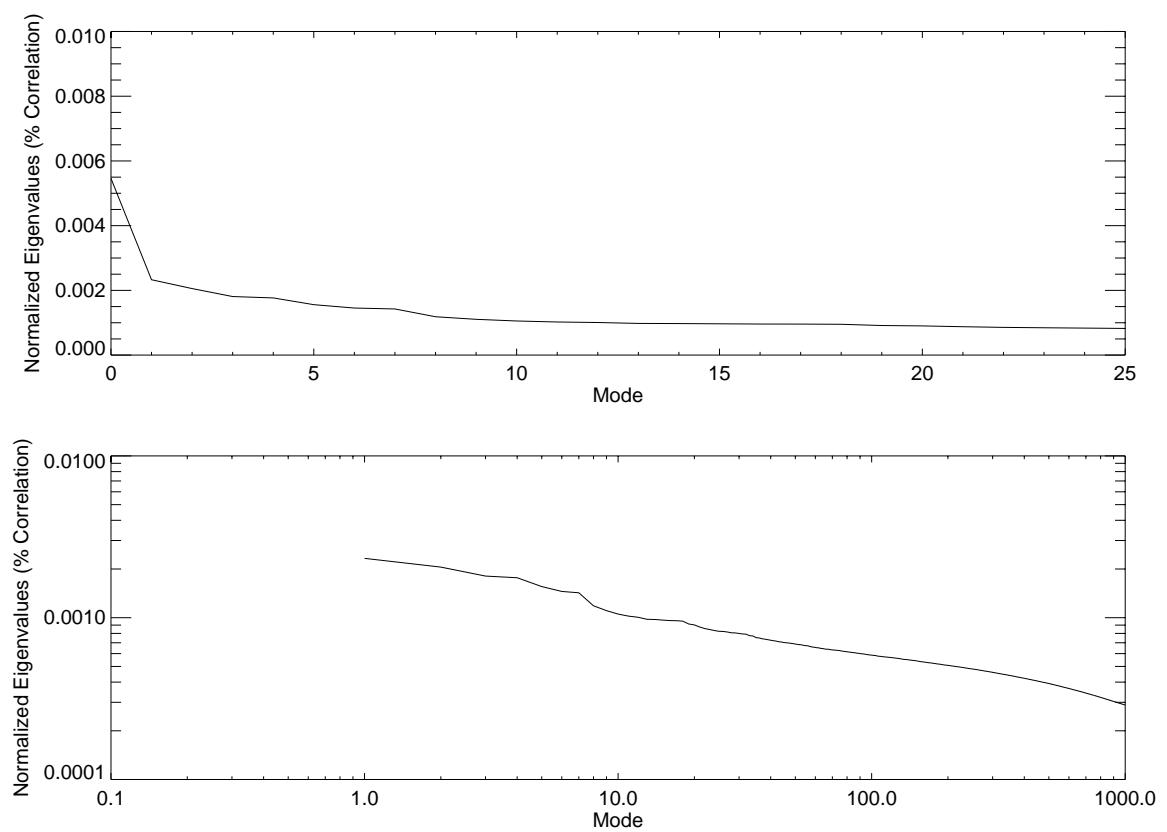


Figure 3.9: Eigenvalues,  $\lambda_i^2$ , normalized by the sum of all eigenvalues, versus mode for the KLE decomposition of the entire southern California data set, 1932-1999.

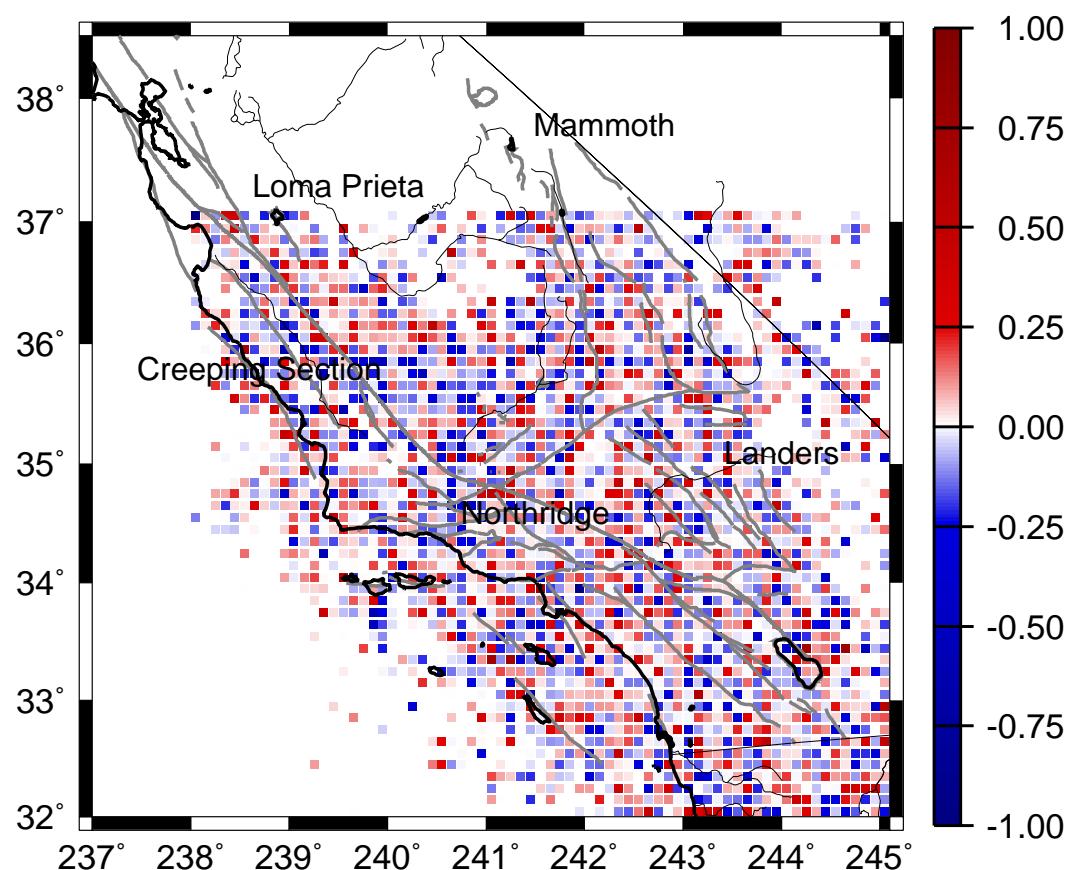


Figure 3.10: KLE mode 500, southern California seismicity 1932-1998.

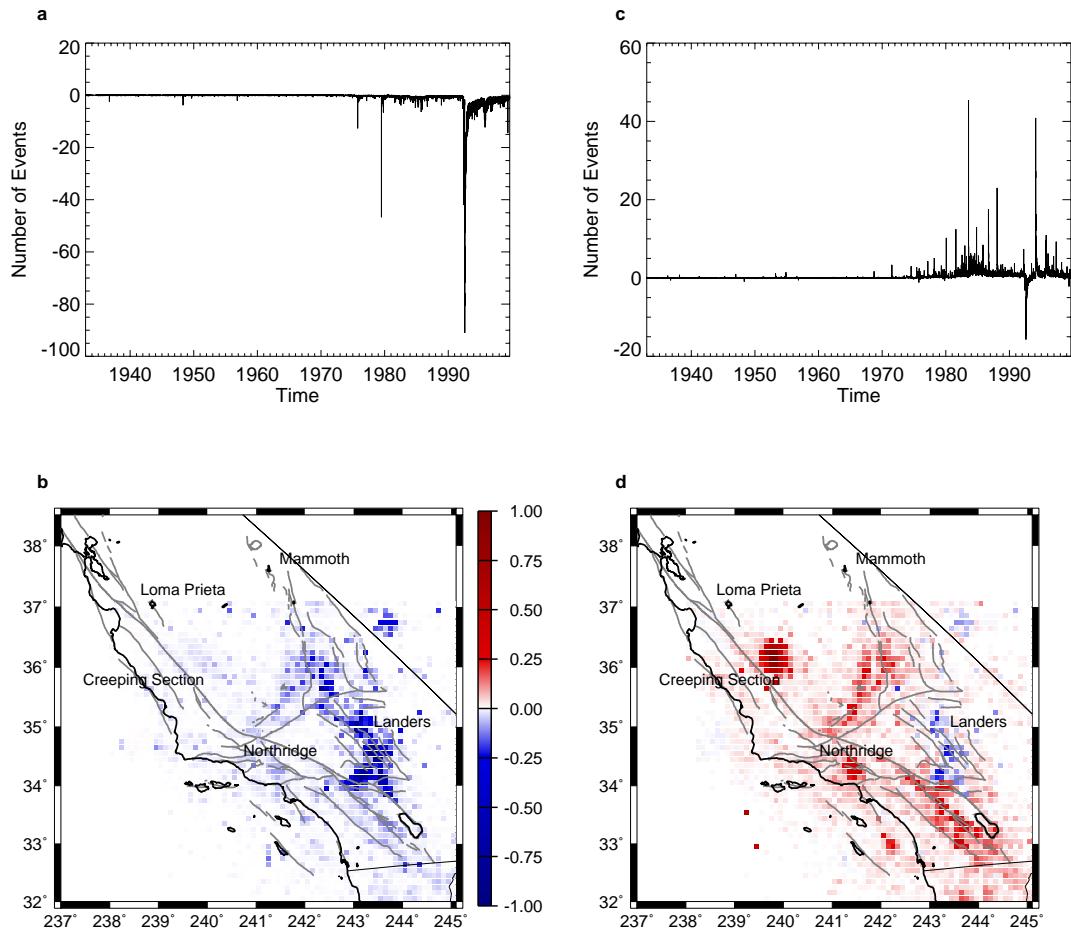


Figure 3.11: First two KLE modes for southern California seismicity, 1932-1999. a) PC time series for first KLE mode; b) first KLE mode, normalized to maximum; c) PC time series for second KLE mode; and d) second KLE mode, also normalized to the maximum.

Figure 3.11d shows the second KLE mode. Here the region surrounding the 1992 Landers event is on while the rest of the southern San Andreas fault system is off. It should be noted here that the increase in the size of the region has had little effect on the nature of the KLE modes, except to increase the relative importance of the Coalinga earthquake, and to reveal an apparent correlation between Landers and a set of events in eastern Nevada. Nor has the inclusion of events of all qualities, and those box locations without any events, affected the nature and extent of the eigenpatterns.

However, as can be seen in Figures 3.12 through 3.13, a number of the spurious modes from the original decomposition described in Section 3.4.2 above disappear in this analysis. In their place are lower order harmonics of the second mode. I attribute this to the inclusion of the entire Coalinga event, including its aftershocks, and the Avila Beach event, which were both missing in Figure 3.2d above. The increase in areal size increases both the number of eigenmodes and the correlation lengths.

Figure 3.13b shows the Avila Beach earthquake of 1969, apparently uncorrelated with other events in southern California. Figure 3.13d is another harmonic of the second mode. Interestingly, Figure 3.14b, KLE7, illustrates the correlations between the North Palm Springs event and other major southern California earthquake, but with minimal correlation with the Landers sequence. Notice the Oceanside event of 1986 in KLE8, a feature prominent in most of the KLE modes in this decomposition and notably absent from that of Section 3.4.1.

The ninth and tenth KLE modes, Figure 3.15, are on a smaller spatial scale. Figure 3.15d shows the Landers sequence, essentially isolated, with the Joshua Tree earthquake anticorrelated with the Landers event to the north and Big Bear to the northwest.

Figure 3.16 is the same eigenvalue plot, as in Figure 3.11, but for a magnitude

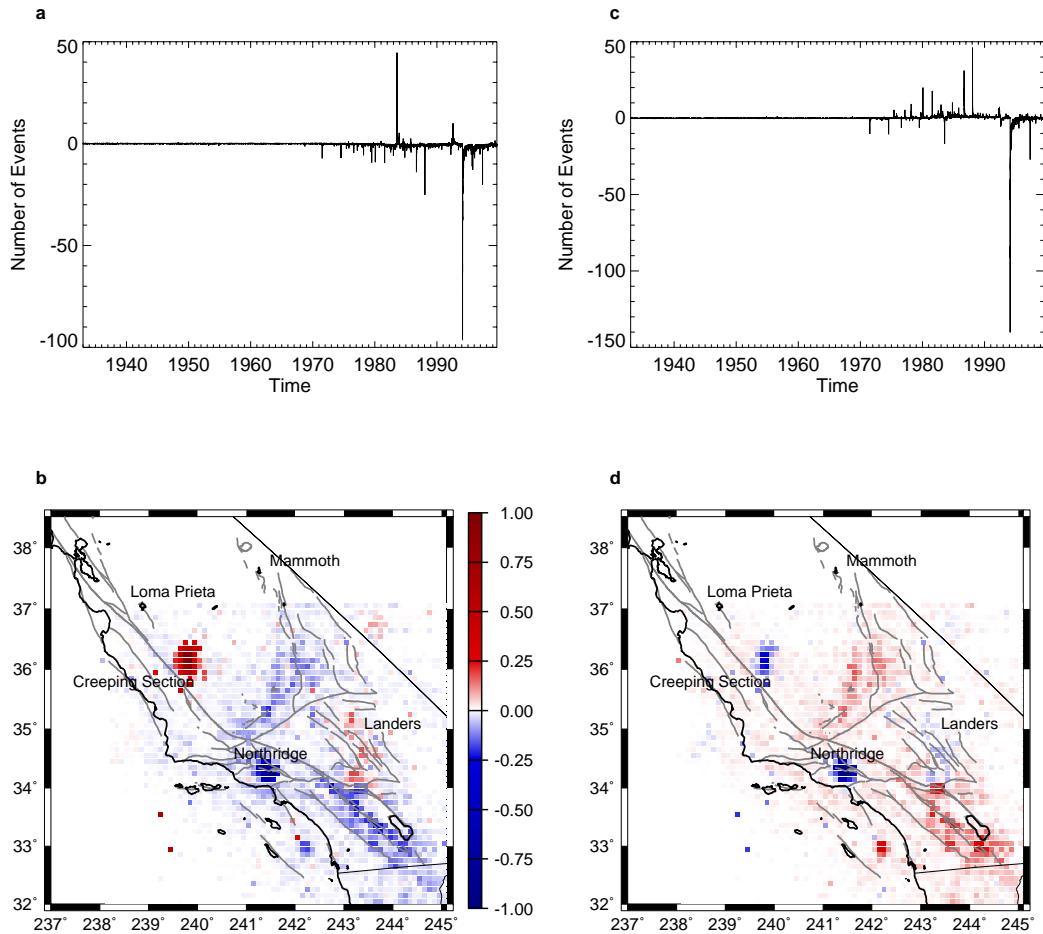


Figure 3.12: Third and fourth KLE modes for southern California seismicity, 1932-1999. a) PC time series for third KLE mode; b) third KLE mode, normalized to maximum; c) PC time series for fourth KLE mode; and d) fourth KLE mode, also normalized to the maximum.

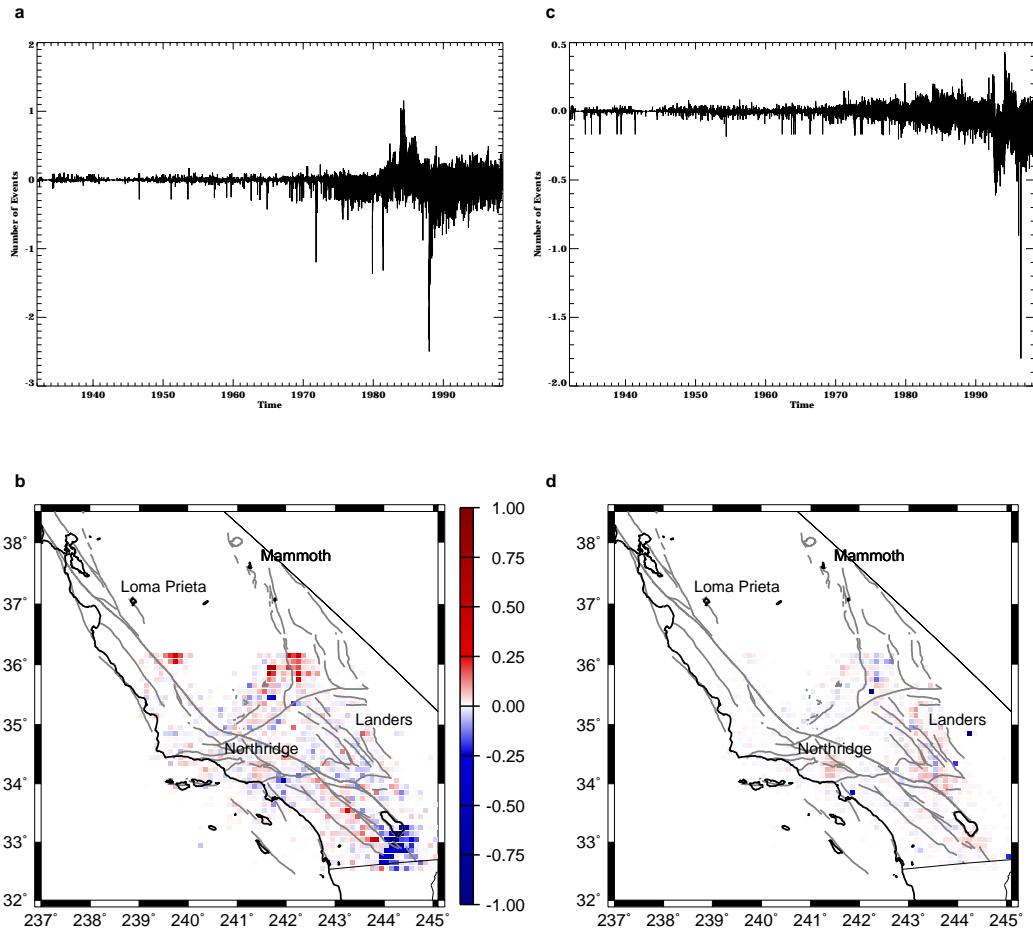


Figure 3.13: Fifth and sixth KLE modes for southern California seismicity, 1932-1999. a) PC time series for fifth KLE mode; b) fifth KLE mode, normalized to maximum; c) PC time series for sixth KLE mode; and d) sixth KLE mode, also normalized to the maximum.

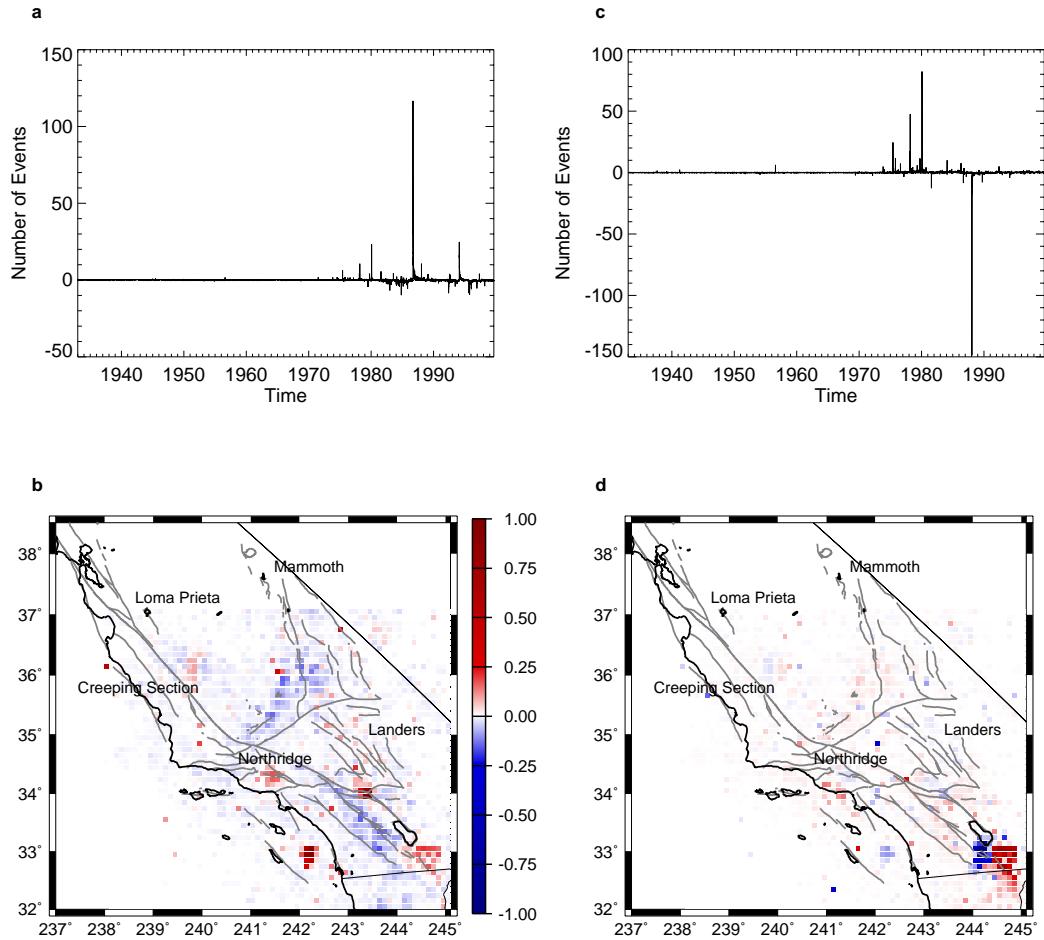


Figure 3.14: Seventh and eighth KLE modes for southern California seismicity, 1932-1998. a) PC time series for seventh KLE mode; b) seventh KLE mode, normalized to maximum; c) PC time series for eighth KLE mode; and d) eighth KLE mode, also normalized to the maximum.

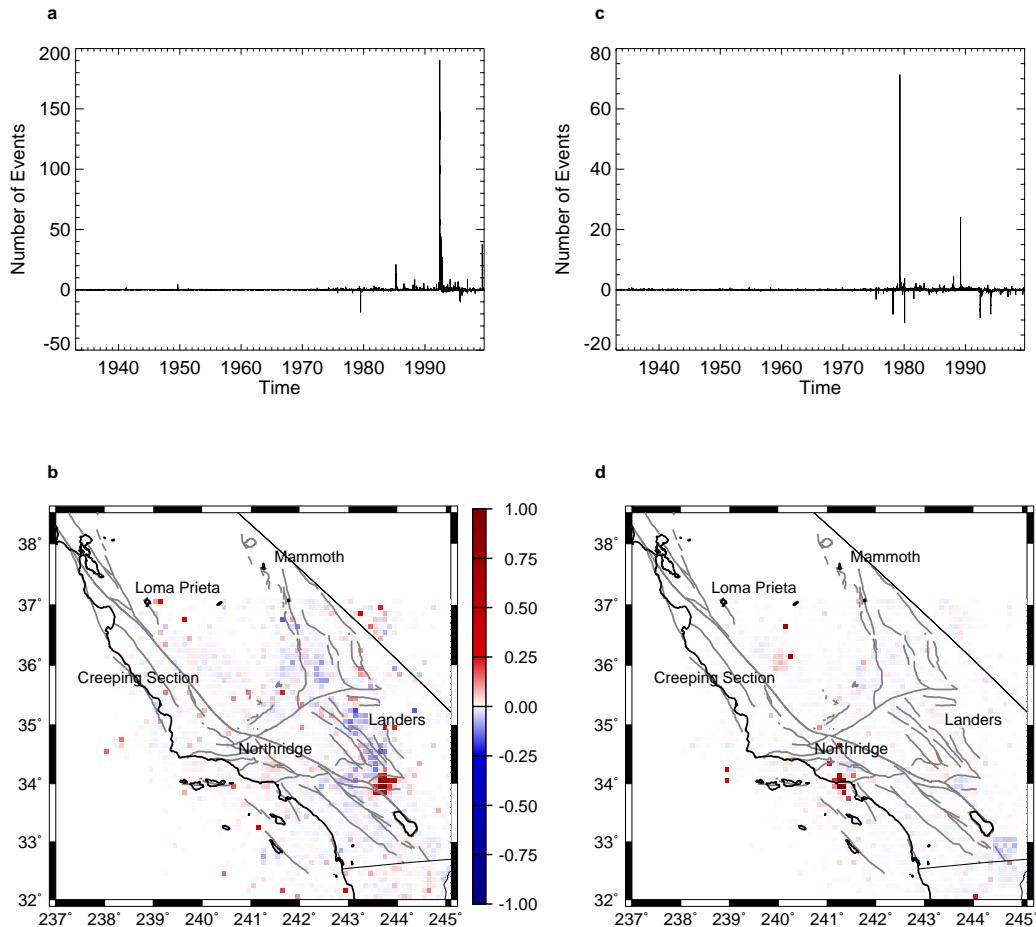


Figure 3.15: Ninth and tenth KLE modes for southern California seismicity, 1932-1999. a) PC time series for ninth KLE mode; b) ninth KLE mode, normalized to maximum; c) PC time series for tenth KLE mode; and d) tenth KLE mode, also normalized to the maximum.

cutoff of 3.0. Again, the first mode is no longer the detection limit eigenpattern, but the Landers sequence essentially uncorrelated with other events (Figure 3.16), and the second KLE mode is, somewhat surprisingly, the Avila Beach event of 1969, correlated with an unnamed event which occurred in the ocean to the south, in October of that same year.

The third KLE mode, 3.17, is a combination of the San Fernando earthquake of 1971 and the Northridge event of 1994 (visible in the associated PC time series). The fourth KLE mode is the 1983 Coalinga earthquake.

KLE modes five and six, Figure 3.18, are on a spatially smaller scale, and show the 1979 Imperial Valley event and the 1987 Superstition Hills and Elmore Ranch events, correlated and anticorrelated, respectively.

Note that, again, the application of a magnitude cut has allowed events from earlier in the data set to have a greater prominence. Figure 3.19 show the seventh KLE mode, where the 1971 San Fernando event is anticorrelated with the Northridge event, a lower order harmonic of the third eigenpattern. The 1952 Kern County event has appeared in the ninth and tenth modes 3.20.

### **3.4.3 Time Period: 1932 through December, 1991.**

One of the interesting questions which arises in studying the results in Sections 3.4.1 and 3.4.2, above, is how exactly does the dominance of the Landers sequence, and its associated instrumentation, affects the eigenpatterns. To test this effect I removed that event from the data set by cutting off the time series before its occurrence, at the end of 1991. For this time period, 1932 to 1991, the time series interval was again one day, so that the total number of time steps is approximately 21,535. Again, all locations from the entire database, and all quality events, were included in the decomposition, so that the number of location time series is 3162.

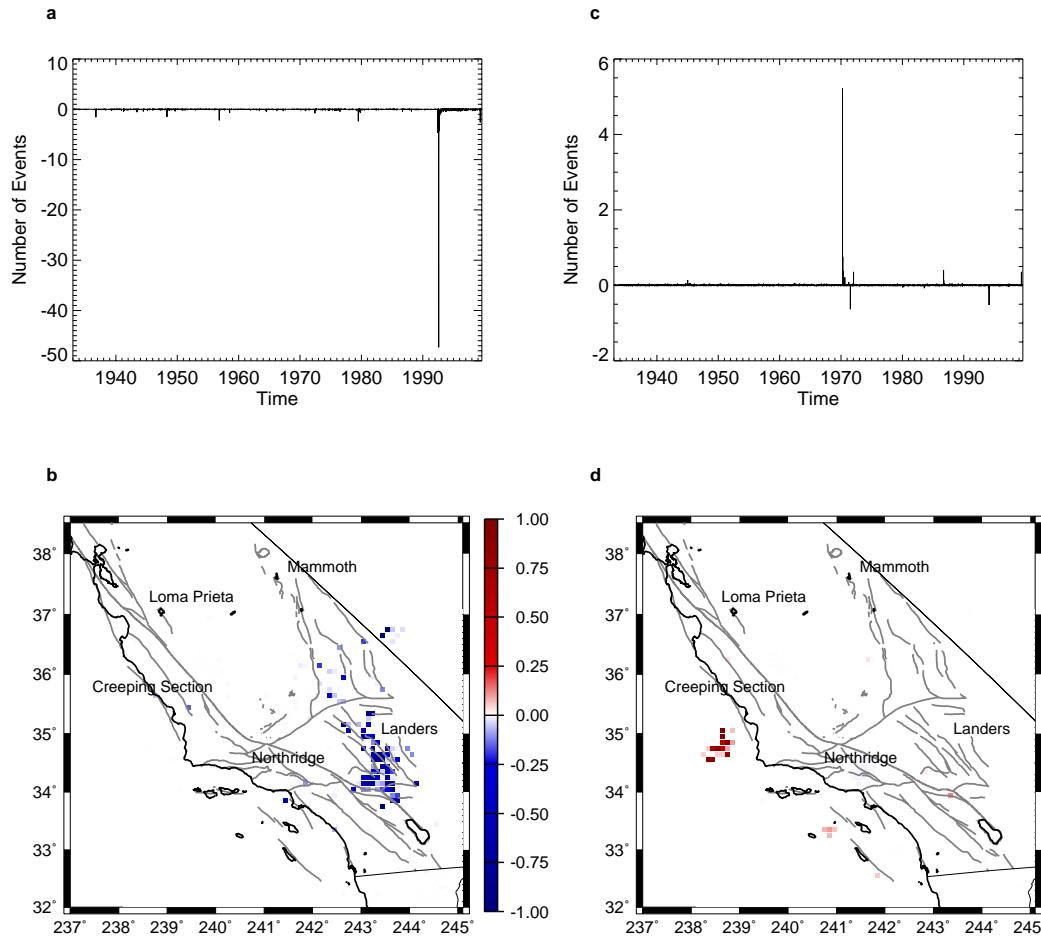


Figure 3.16: First two KLE modes for southern California seismicity, 1932-1999, with a magnitude cut of 3.0. a) PC time series for first KLE mode; b) first KLE mode, normalized to maximum; c) PC time series for second KLE mode; and d) second KLE mode, also normalized to the maximum.

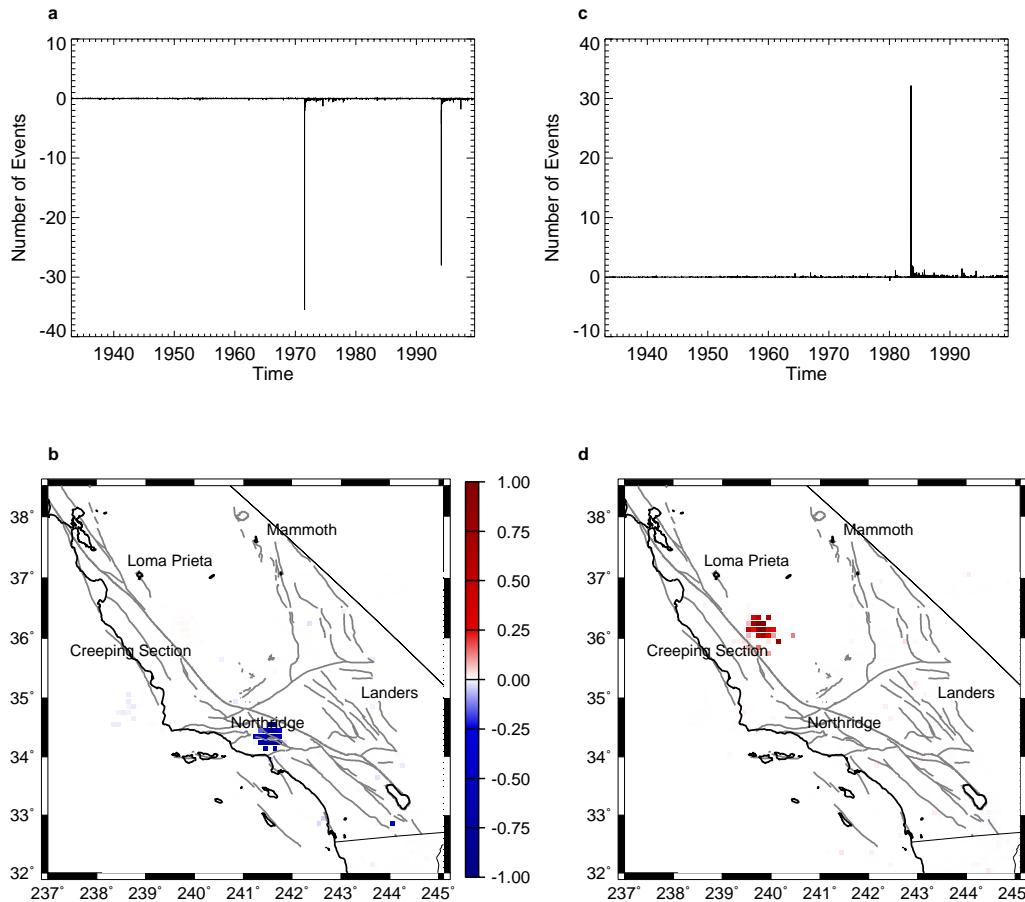


Figure 3.17: Third and fourth KLE modes for southern California seismicity, 1932-1999, with a magnitude cut of 3.0. a) PC time series for third KLE mode; b) third KLE mode, normalized to maximum; c) PC time series for fourth KLE mode; and d) fourth KLE mode, also normalized to the maximum.

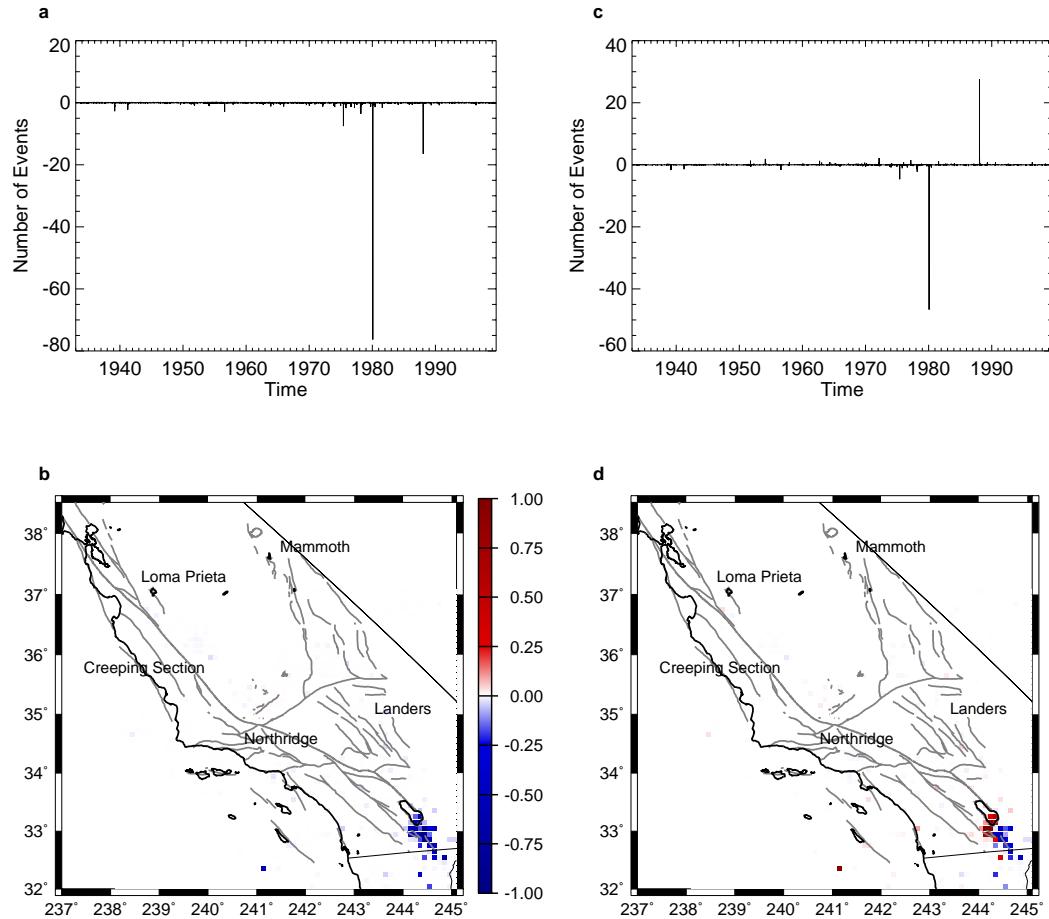


Figure 3.18: Fifth and sixth KLE modes for southern California seismicity, 1932-1999, with a magnitude cut of 3.0. a) PC time series for fifth KLE mode; b) fifth KLE mode, normalized to maximum; c) PC time series for sixth KLE mode; and d) sixth KLE mode, also normalized to the maximum.

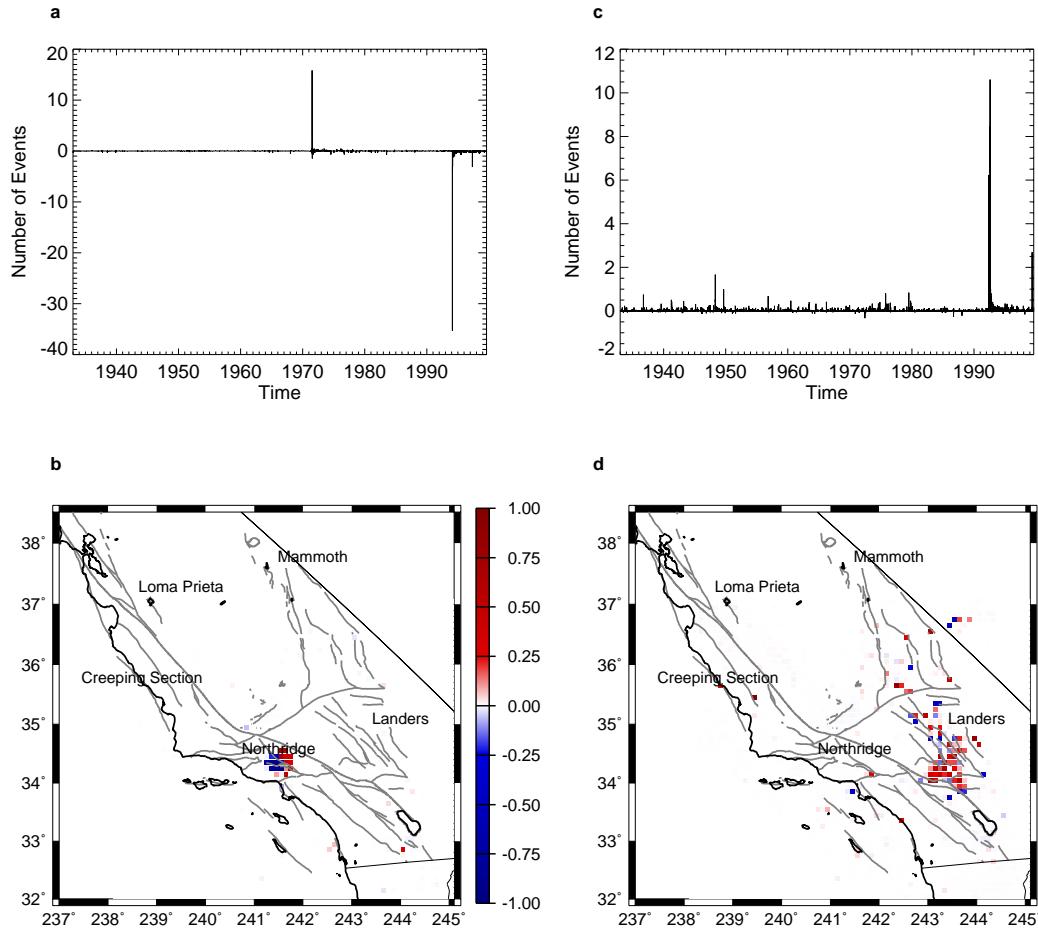


Figure 3.19: Seventh and eighth KLE modes for southern California seismicity, 1932-1999, with a magnitude cut of 3.0. a) PC time series for seventh KLE mode; b) seventh KLE mode, normalized to maximum; c) PC time series for eighth KLE mode; and d) eighth KLE mode, also normalized to the maximum.

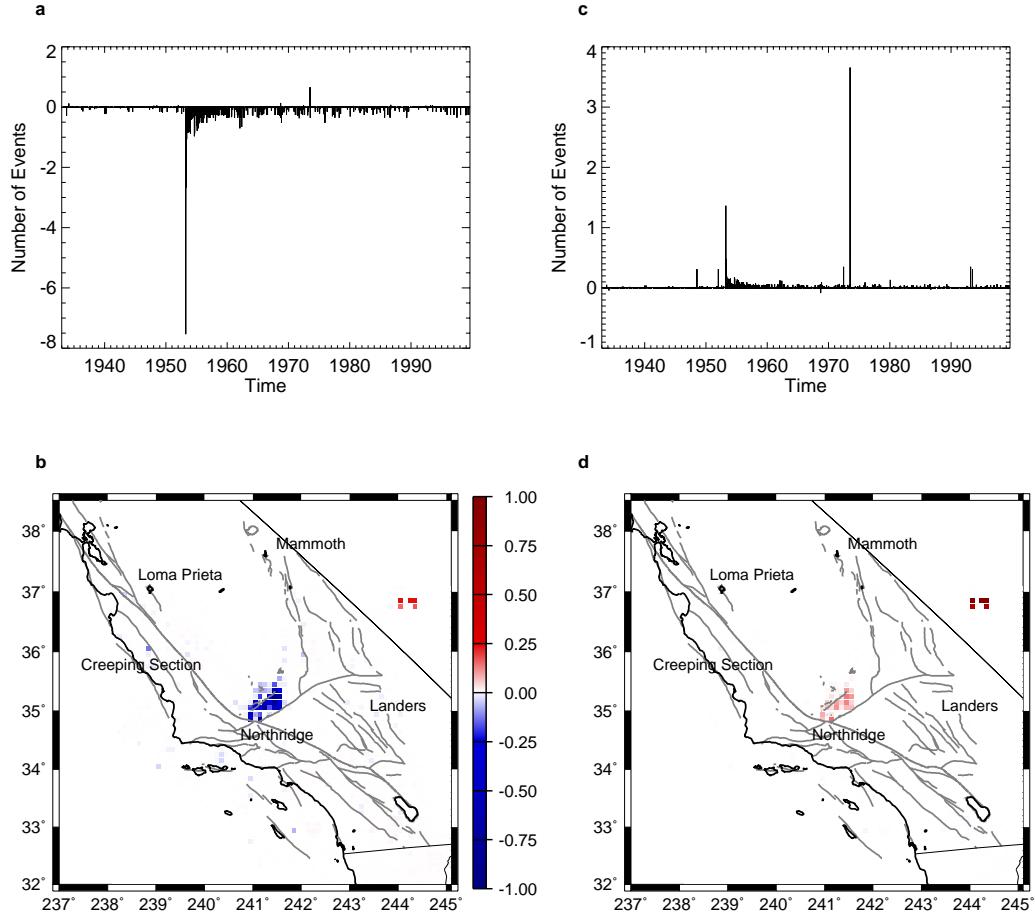


Figure 3.20: Ninth and tenth KLE modes for southern California seismicity, 1932-1999, with a magnitude cut of 3.0. a) PC time series for ninth KLE mode; b) ninth KLE mode, normalized to maximum; c) PC time series for tenth KLE mode; and d) tenth KLE mode, also normalized to the maximum.

Figure 3.21a shows the first 25 normalized eigenvalues, while Figure 3.21b is the first 1000 normalized eigenvalues, plotted on a log-normal scale. The eigenvalue plot is now smoother, without the large drop after the first mode which was a function of the large wave in the first PC time series generated by the Landers aftershocks (see Figure 3.11).

KLE modes one and two are shown in Figure 3.22, where KLE1 is again the detectability mode. Note that there is no large signal for Landers visible in Figure 3.22b. In addition, many of what were the lower modes in the previous analysis have moved up in the eigenvalue ranking, replacing the large number of Landers harmonics. KLE2 is now the 1983 Coalinga earthquake, anticorrelated with the 1986 Oceanside and North Palm Spring events. The 1971 San Fernando event is a single blue box, correlated with the Coalinga earthquake.

The third KLE mode shown in Figure 3.23 is again the 1969 Avila Beach event. KLE4, Figure 3.23, shows the 1987 Superstition Hills and Elmore ranch events correlated with several earthquakes the Whittier Narrows earthquake and anticorrelated with the North Palm Springs and Oceanside events. Figure 3.24b, KLE mode five, also shows the North Palm Springs and Oceanside events correlated with each other, while in KLE mode six Imperial Valley is anticorrelated with the Superstition Hills/Elmore Ranch sequence. Point Mugu, 1973, also has appeared south of Northridge. Note that smaller scale features again appear at higher mode numbers.

Figure 3.25 shows Point Mugu correlated with the 1971 San Fernando event. The Tejon Ranch earthquake of 1988, as well as the 1979 Homestead Valley earthquake, south of Landers, are also correlated with these events, while Imperial Valley is anticorrelated with these events. KLE8 again shows the 1979 Homestead Valley event, now anticorrelated with Point Mugu. Note the arcuate structure at Homestead Valley, which cuts across the faults in that area. This is a feature of

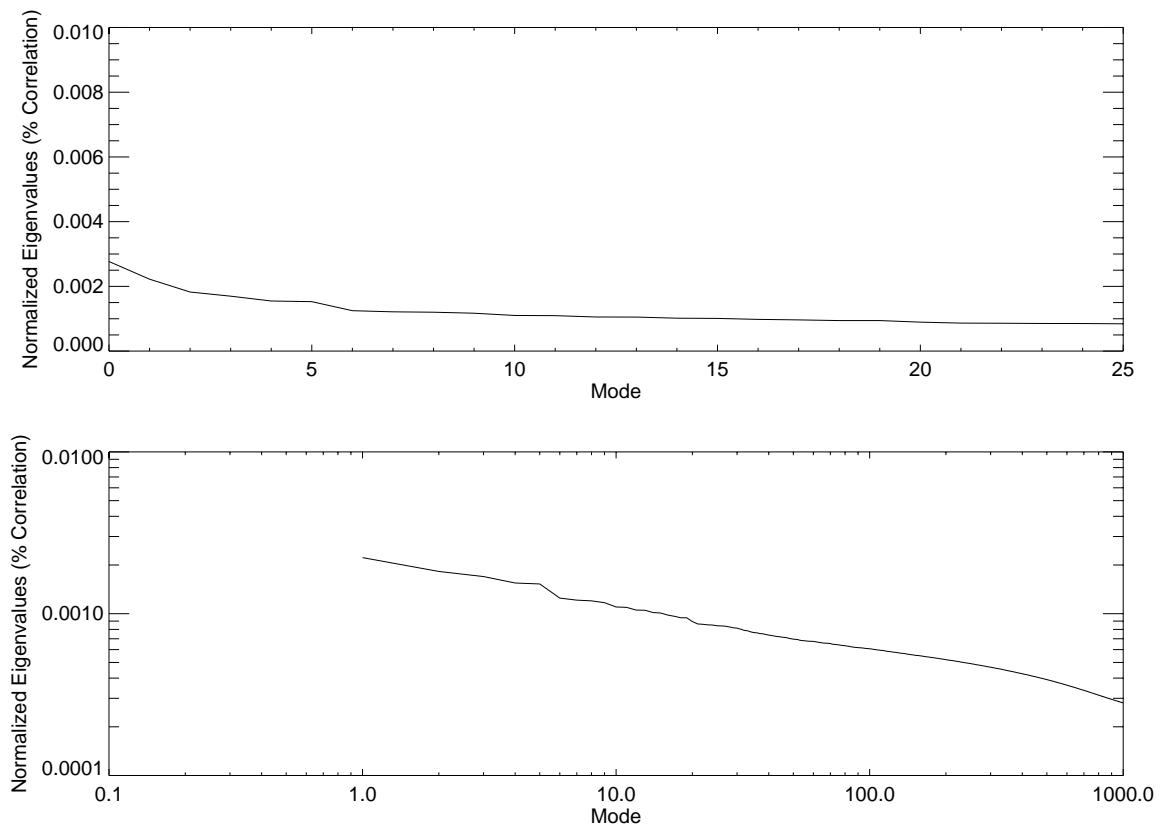


Figure 3.21: Eigenvalues,  $\lambda_i^2$ , normalized by the sum of all eigenvalues, versus mode for the KLE decomposition of the entire southern California data set, 1932-1991.

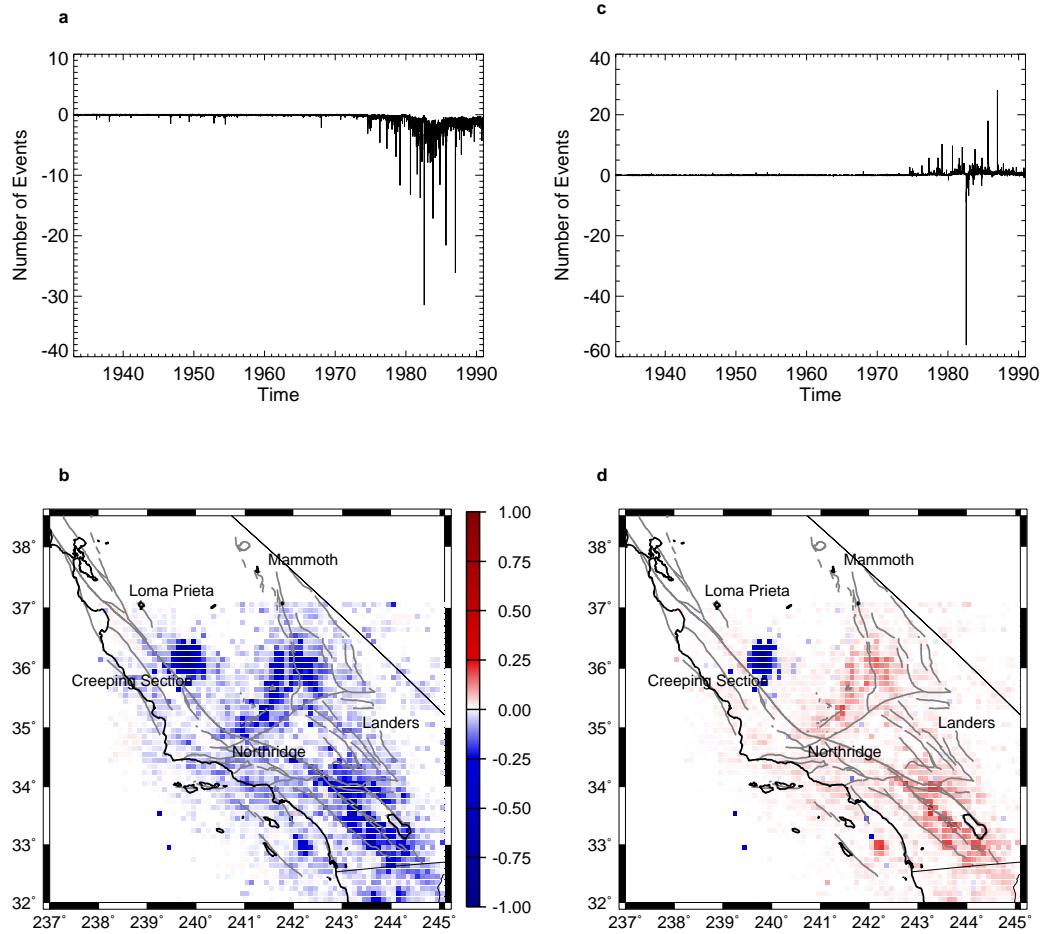


Figure 3.22: First two KLE modes for southern California seismicity, 1932-1991. a) PC time series for first KLE mode; b) first KLE mode, normalized to maximum; c) PC time series for second KLE mode; and d) second KLE mode, also normalized to the maximum.

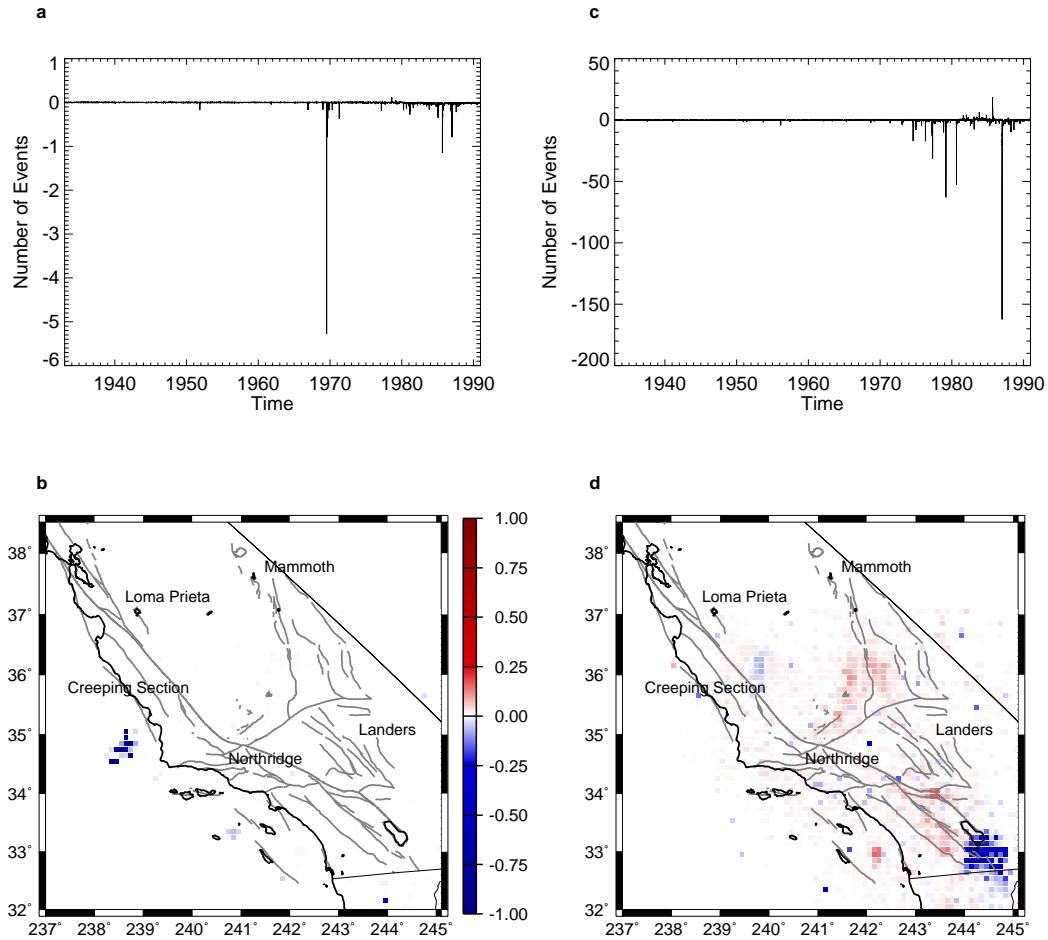


Figure 3.23: The third and fourth KLE modes for southern California seismicity, 1932-1991. a) PC time series for third KLE mode; b) third KLE mode, normalized to maximum; c) PC time series for fourth KLE mode; and d) fourth KLE mode, also normalized to the maximum.

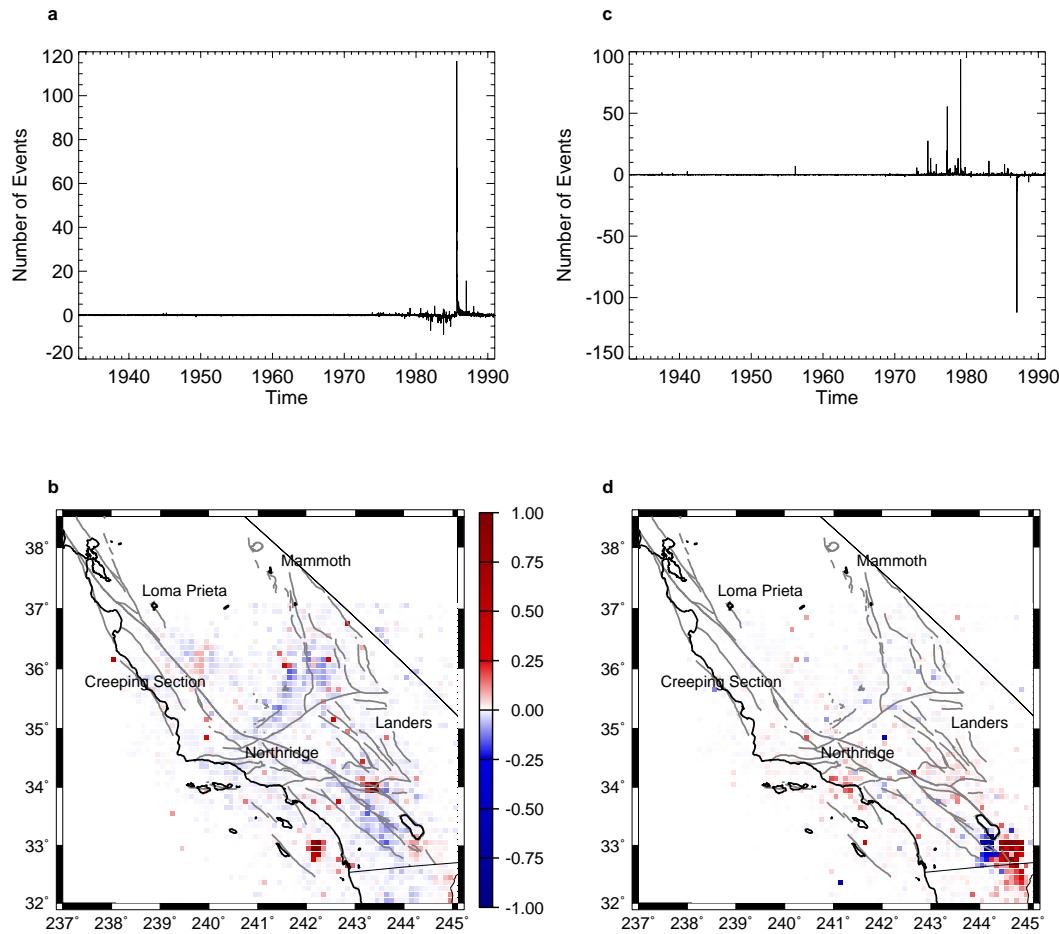


Figure 3.24: Fifth and sixth KLE modes for southern California seismicity, 1932-1991. a) PC time series for fifth KLE mode; b) fifth KLE mode, normalized to maximum; c) PC time series for sixth KLE mode; and d) sixth KLE mode, also normalized to the maximum.

the local seismicity that has only been recognized in recent years with the occurrence of the 1991 Landers sequence and the 1999 Hector Mine earthquake, but which was clearly visible in this decomposition as early as 1991 (Figure 3.26).

Analysis of this same data set, 1932 through 1991, but with a magnitude cut of 3.0 applied to the data, yields the KLE modes shown in Figures 3.27 through 3.29. Interestingly, while the detectability mode has dropped out again, the first mode is not Coalinga, the second mode in the entire data set (see Figure 3.22 above), but the 1971 San Fernando event. The 1983 Coalinga earthquake is now the third KLE mode, behind the 1969 Avila Beach earthquake (see Figure 3.28). The correlation between the 1979 Imperial Valley and 1987 Superstition Hills sequence is the fourth mode, and its anticorrelation the fifth mode (see Figures 3.28). The 1952 Kern County event has dropped to the seventh mode (Figure 3.29), while the eighth mode is new to the decomposition - the 1968 Borrego Mountain earthquake. Figure 3.29 shows KLE modes nine and ten, now harmonics of the earlier modes.

#### **3.4.4 Time Period: 1932 through December, 1978.**

The recurrent strong correlation or anticorrelation between the Imperial Valley earthquakes of the seventies and eighties, and other events of that time period, suggested that there might be some value in studying the eigenmodes for the entire data set ending just before the 1979 Imperial Valley event. Again, the time series interval used was one day, so that the total number of time steps is approximately 16790. All locations from the entire database, and all quality events, were included in the decomposition, so that the number of location time series remains at 3162.

Figure 3.30a is the first 25 normalized eigenvalues for this decomposition, while Figure 3.30b is the first 1000 normalized eigenvalues, plotted on a log-

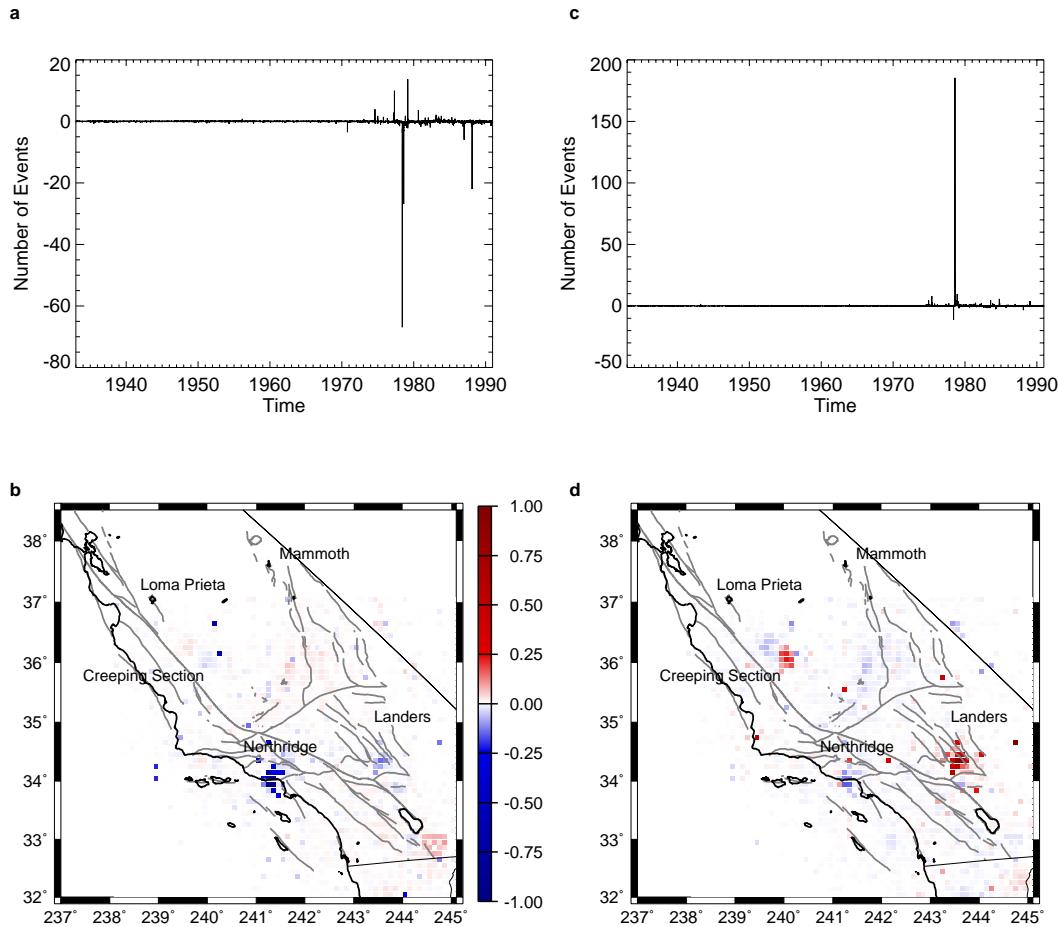


Figure 3.25: Seventh and eighth KLE modes for southern California seismicity, 1932-1991. a) PC time series for seventh KLE mode; b) seventh KLE mode, normalized to maximum; c) PC time series for eighth KLE mode; and d) eighth KLE mode, also normalized to the maximum.

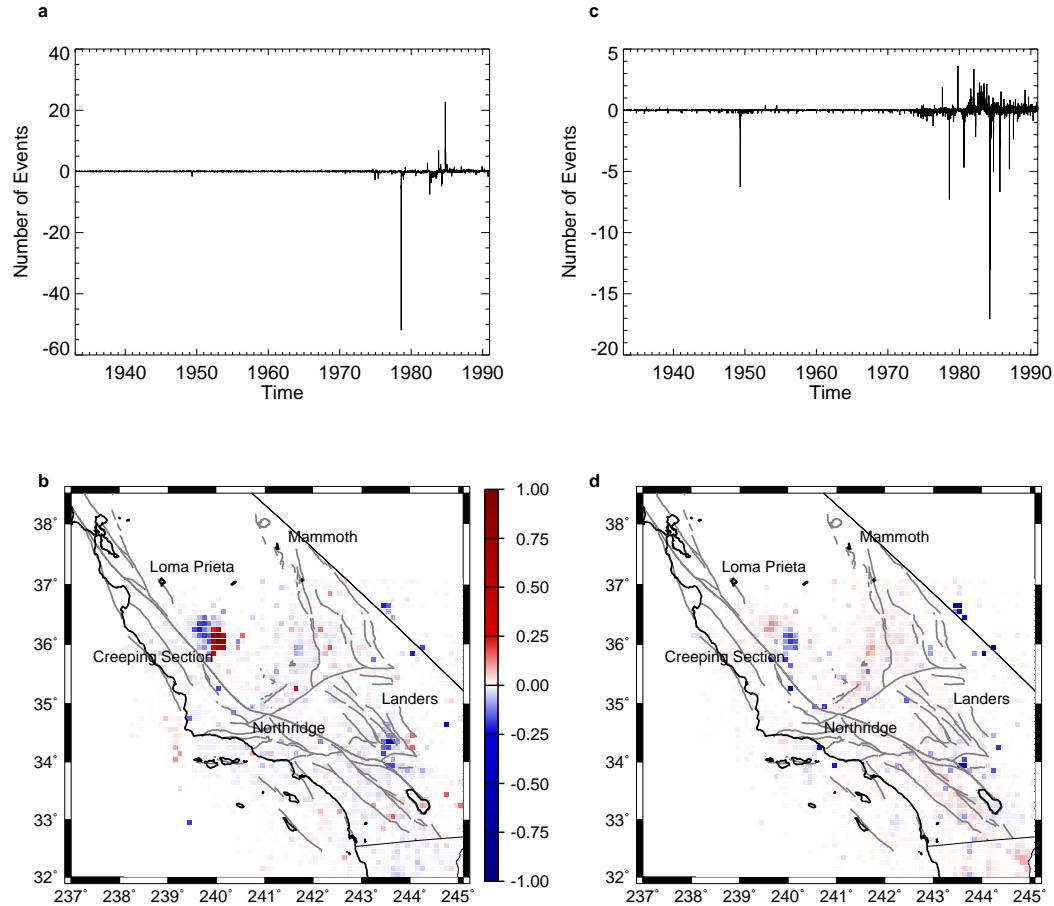


Figure 3.26: Ninth and tenth KLE modes for southern California seismicity, 1932-1991. a) PC time series for ninth KLE mode; b) ninth KLE mode, normalized to maximum; c) PC time series for tenth KLE mode; and d) tenth KLE mode, also normalized to the maximum.

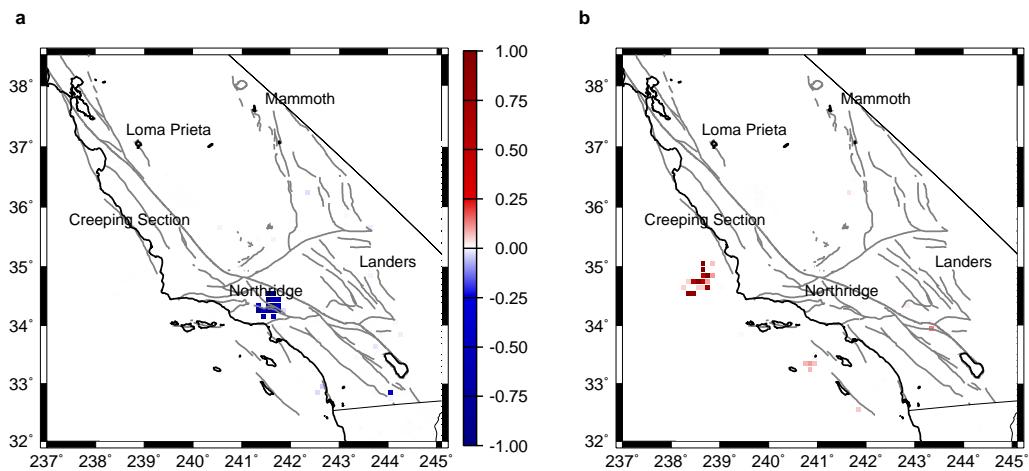


Figure 3.27: First two KLE modes for southern California seismicity, 1932-1991,  $M > 3.0$ . a) First KLE mode, normalized to maximum; b) second KLE mode, also normalized to the maximum.

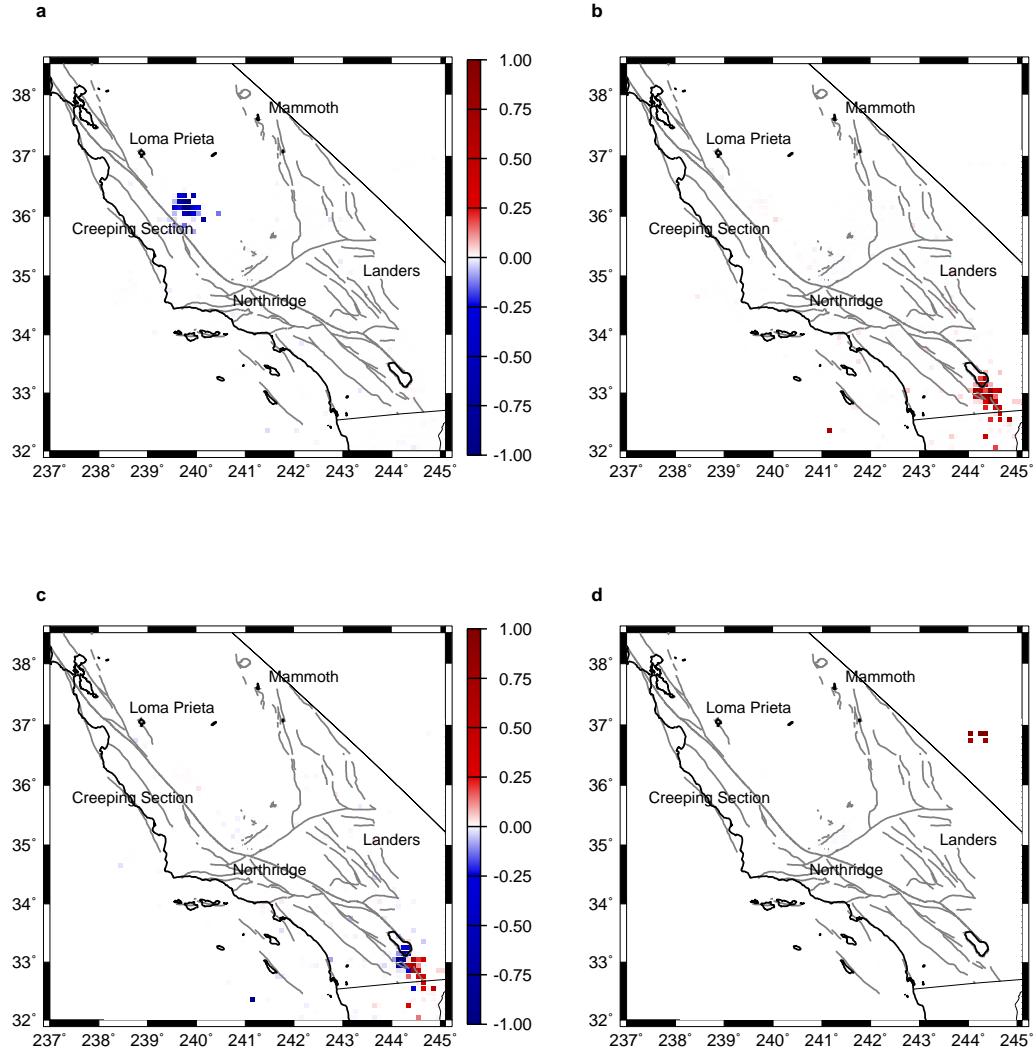


Figure 3.28: Third through sixth KLE modes for southern California seismicity, 1932-1991,  $M > 3.0$ . a) Third KLE mode; b) fourth KLE mode; c) fifth KLE mode; and d) sixth KLE mode, all normalized to the maximum.

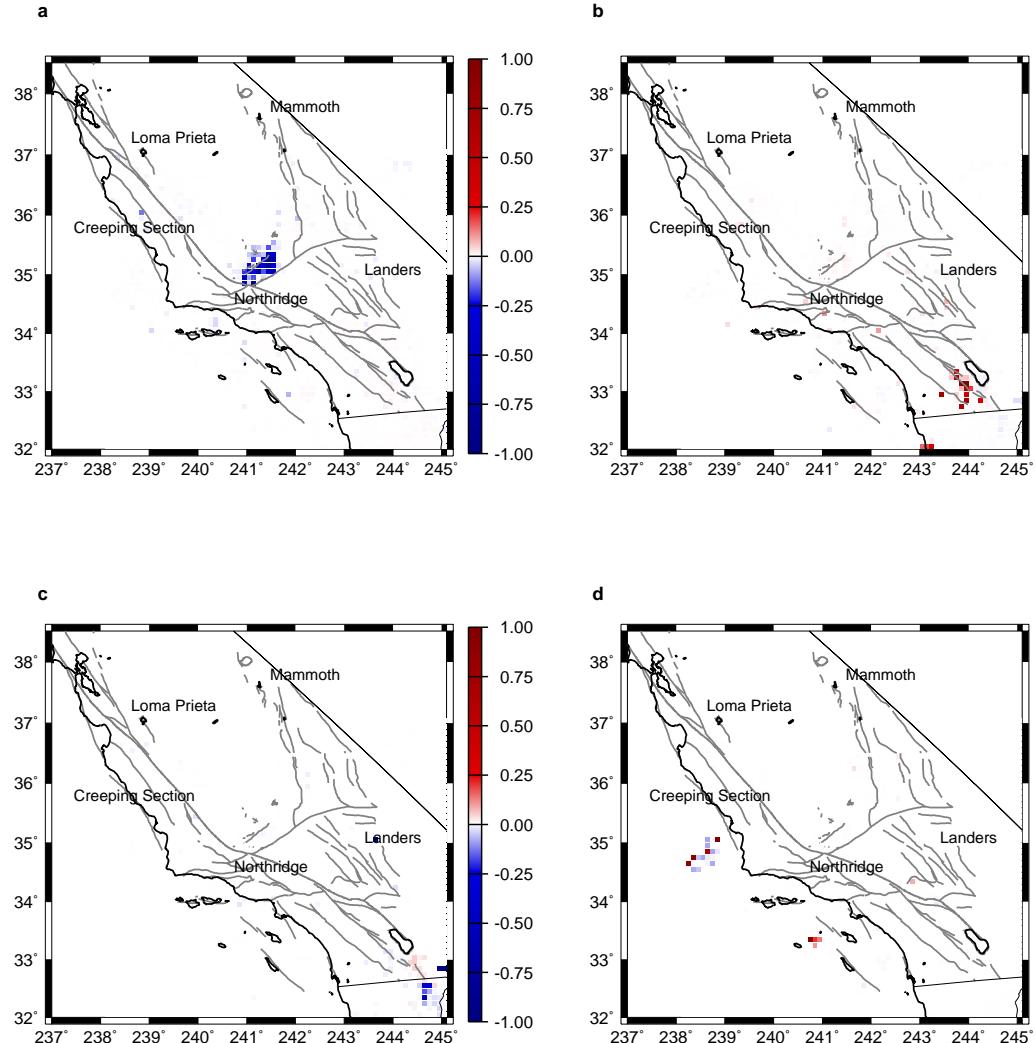


Figure 3.29: Seventh through tenth KLE modes for southern California seismicity, 1932-1991,  $M > 3.0$ . a) Seventh KLE mode; b) eighth KLE mode; c) ninth KLE mode; and d) tenth KLE mode, all normalized to the maximum.

normal scale. Figure 3.31 shows the first and second KLE modes, Avila Beach, 1969, and the detectability mode, respectively. It is interesting that the ability to detect smaller events has dropped in eigenvalue power, confirming that upgrades in the networks in recent years have significantly affected the quality and detail of seismicity data in recent years.

The remaining modes are substantially noisier, and include many smaller scale features. Figure 3.32 shows KLE modes three and four, while KLE modes seven and nine are shown in Figure 3.33. The prominence of events from the late 60s and early 70s is evident, including Point Mugu, San Fernando, and, seen for the first time in any analysis, the 1966 Parkfield event (KLE7).

A cutoff of magnitude 3.0 was applied to this data set as well, and some of those results are shown in Figure 3.34 and 3.35. The 1971 San Fernando event is now the first KLE mode, while Avila Beach is second (Figure 3.34). The 1952 Kern County event is now visible as the fifth KLE mode, while the Borrego Mountain earthquake of 1968 is now mildly anticorrelated with the 1966 Parkfield event in the sixth KLE mode (see Figure 3.35). The Parkfield earthquake is now the seventh mode (Figure 3.35), while lower order harmonics come into the decomposition in the eighth mode. Note, in Figure 3.35b, the appearance of the San Bernardino Mountain event of 1943.

### **3.4.5 Time Period: 1812 through 1994, $M \geq 6.0$ .**

The catalog for events of magnitude greater than 6.0, 1812 through 1994, compiled by Deng and Sykes in the mid-1990s is given Appendix C, C.2 [<http://www.gps.caltech.edu/~jishu>]. A KLE decomposition was applied to this data as well, where the time steps were set at one year. The time series for each location was assigned a 1.0 if an earthquake occurred at that location in a given year, or a 0.0 otherwise. The location box size was the same as above, squares of  $0.1^\circ$

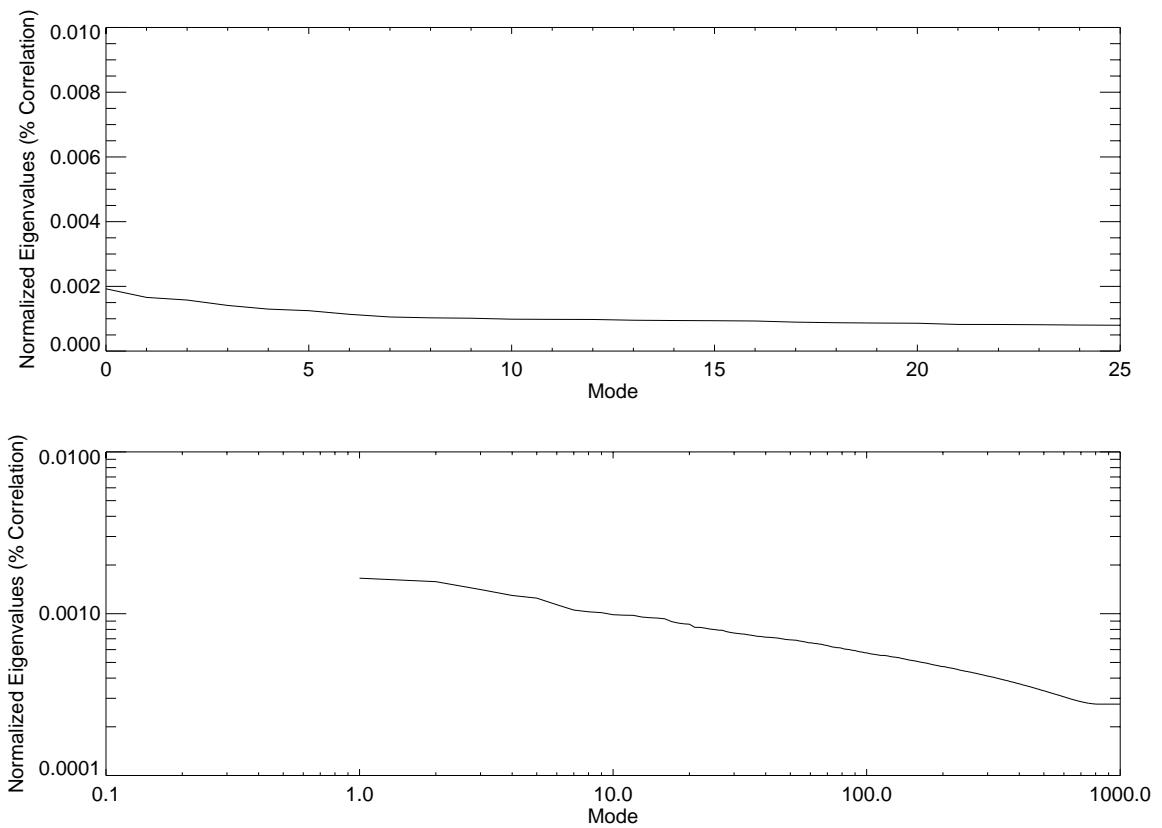


Figure 3.30: Eigenvalues,  $\lambda_i^2$ , normalized by the sum of all eigenvalues, versus mode for the KLE decomposition of the entire southern California data set, 1932-1978.

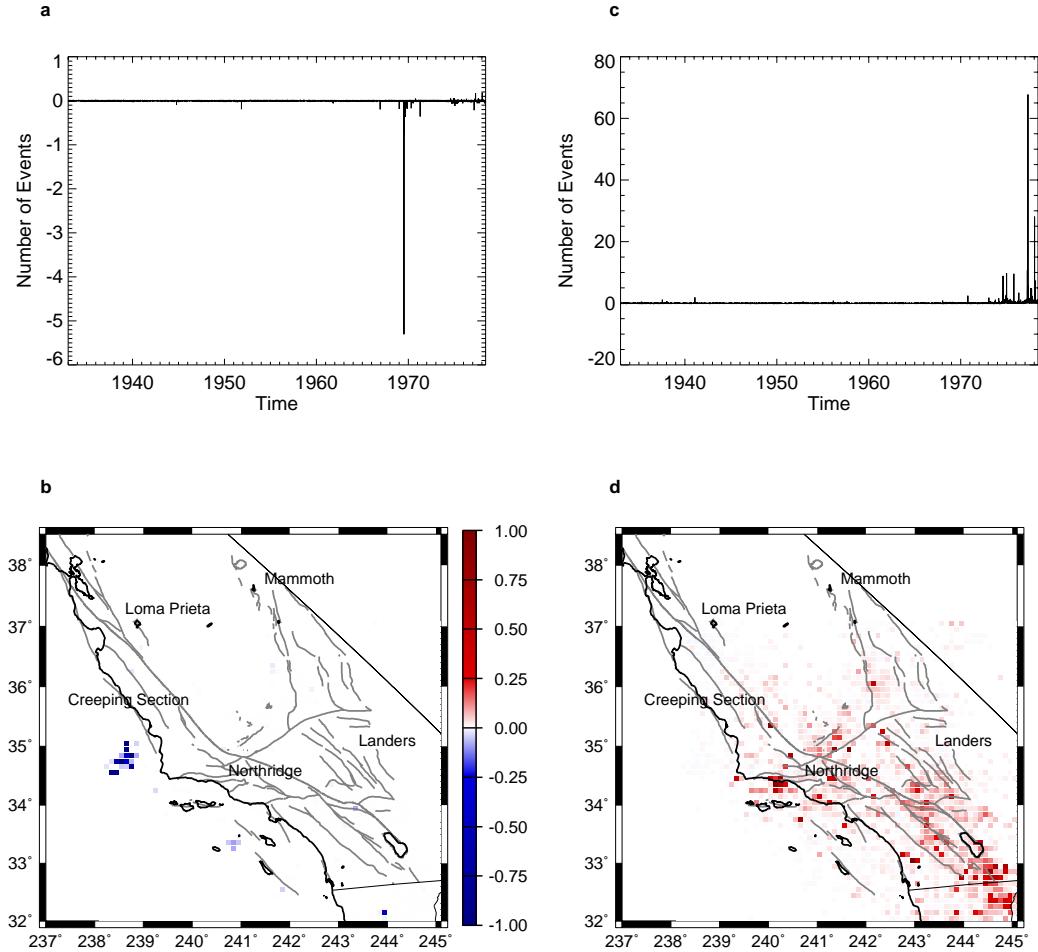


Figure 3.31: First two KLE modes for southern California seismicity, 1932-1978. a) PC time series for first KLE mode; b) first KLE mode, normalized to maximum; c) PC time series for second KLE mode; and d) second KLE mode, also normalized to the maximum.

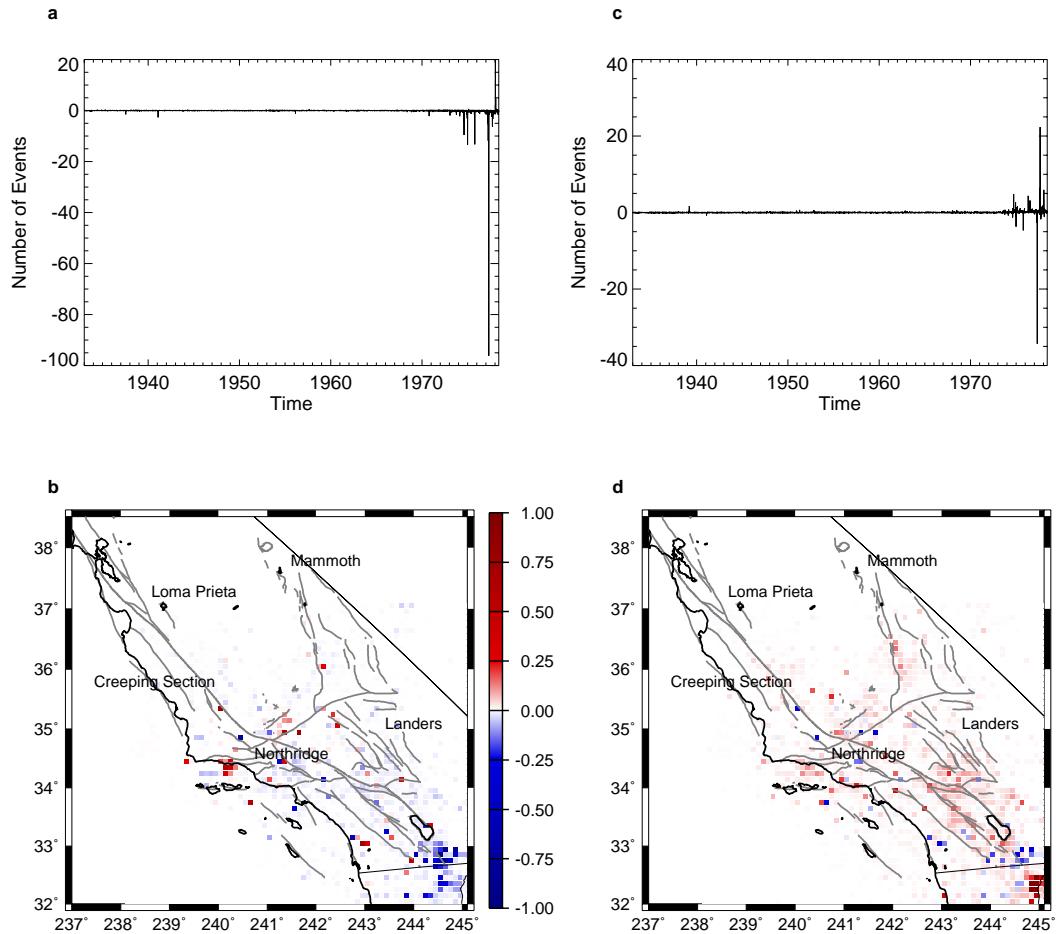


Figure 3.32: Third and fourth KLE modes for southern California seismicity, 1932-1978. a) PC time series for third KLE mode; b) third KLE mode, normalized to maximum; c) PC time series for fourth KLE mode; and d) fourth KLE mode, also normalized to the maximum.

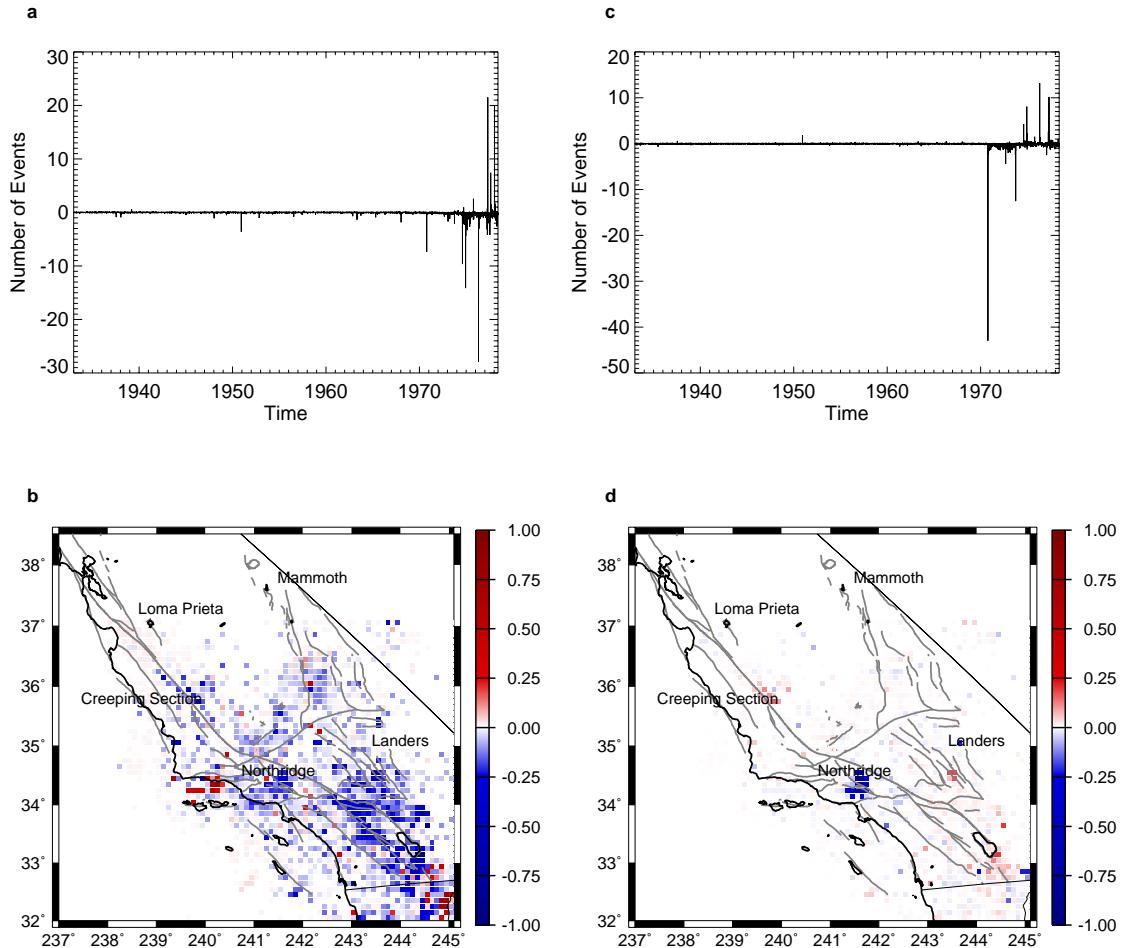


Figure 3.33: Seventh and ninth KLE modes for southern California seismicity, 1932-1978. a) PC time series for seventh KLE mode; b) seventh KLE mode, normalized to maximum; c) PC time series for ninth KLE mode; and d) ninth KLE mode, also normalized to the maximum.

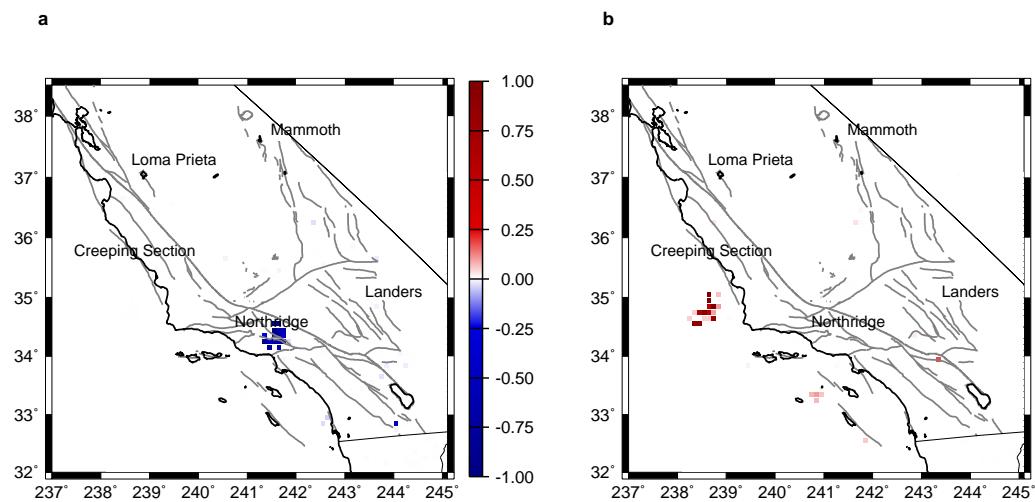


Figure 3.34: First two KLE modes for southern California seismicity, 1932-1978,  $M > 3.0$ . a) First KLE mode and b) second KLE mode, normalized to maximum.

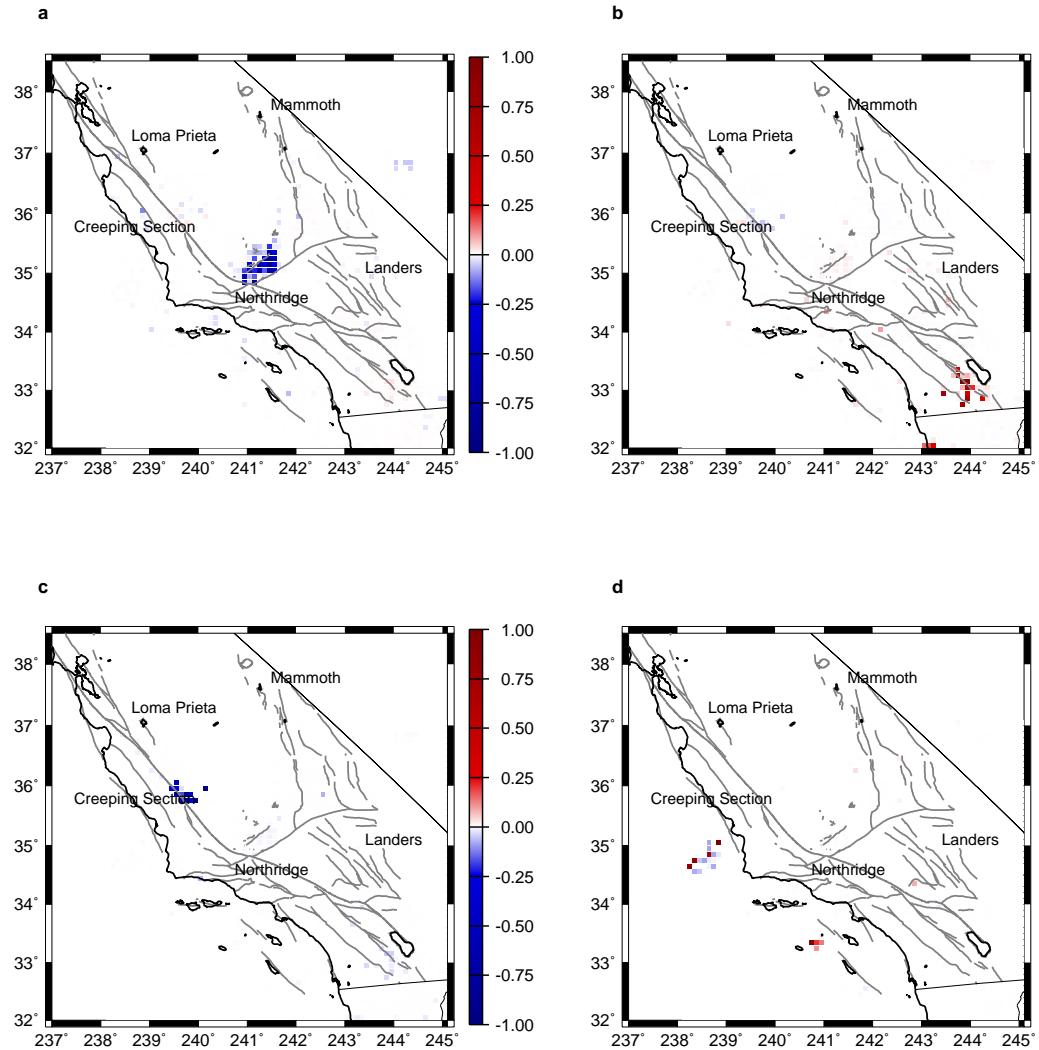


Figure 3.35: Fifth through seventh KLE modes for southern California seismicity, 1932-1978,  $M > 3.0$ . a) Fifth KLE mode; b) sixth KLE mode; c) seventh KLE mode; and d) eighth KLE mode, all normalized to the maximum.

latitude and  $0.1^\circ$  longitude to a side. The total area ranged from  $-115^\circ$  and  $-121^\circ$  longitude and  $32.5^\circ$  and  $36.5^\circ$  latitude, and only those locations in which at least one event occurred over the entire time period were included in the analysis, as in 3.4.1 above.

Figure 3.36 shows the first two KLE modes for this decomposition. The first KLE mode is the Kern County earthquake of 1952, while the second mode is the Landers sequence. The third mode and fourth mode, shown in Figure 3.37, are harmonics of the same mode, as are the fifth and sixth modes, Figure 3.38, only on a smaller scale. Note the locations of the 1812 Wrightwood earthquake, the Owens Valley sequence of 1872, the 1883 Santa Barbara Channel earthquake, the Pajaro gap event of 1890, the San Jacinto/Elsinore fault events of the early 1890s, the 1908 Death Valley earthquake, the 1952 Kern County event, and the Superstition Hills sequence of the late 1980s.

Jumping to KLE modes nine and ten, the Parkfield earthquakes from the early part of this century (1901 and 1922) are evident, correlated (or anticorrelated) with the Imperial Valley events of 1915 and 1940 and the San Jacinto earthquakes of 1899 and 1918 (Figure 3.39). The appearance of an eigenpattern for the Parkfield area is not a surprise, as it is one of the best known locations in the world repeating earthquakes are thought to occur. Figure 3.40 shows the plot of events at Parkfield versus time for this century [Turcotte, 1997]. The interesting development here is the correlation with other areas of the San Andreas system, at distances larger than those anticipated in stress transfer calculations [Harris and Simpson, 1992; Stein et al., 1992; Stein, 1999]. Note that, despite the evidence for long-range correlations in the analysis, it fails to find the same patterns after the 1930s, and may relate to the failure of the Parkfield event to occur in the late 1980s.

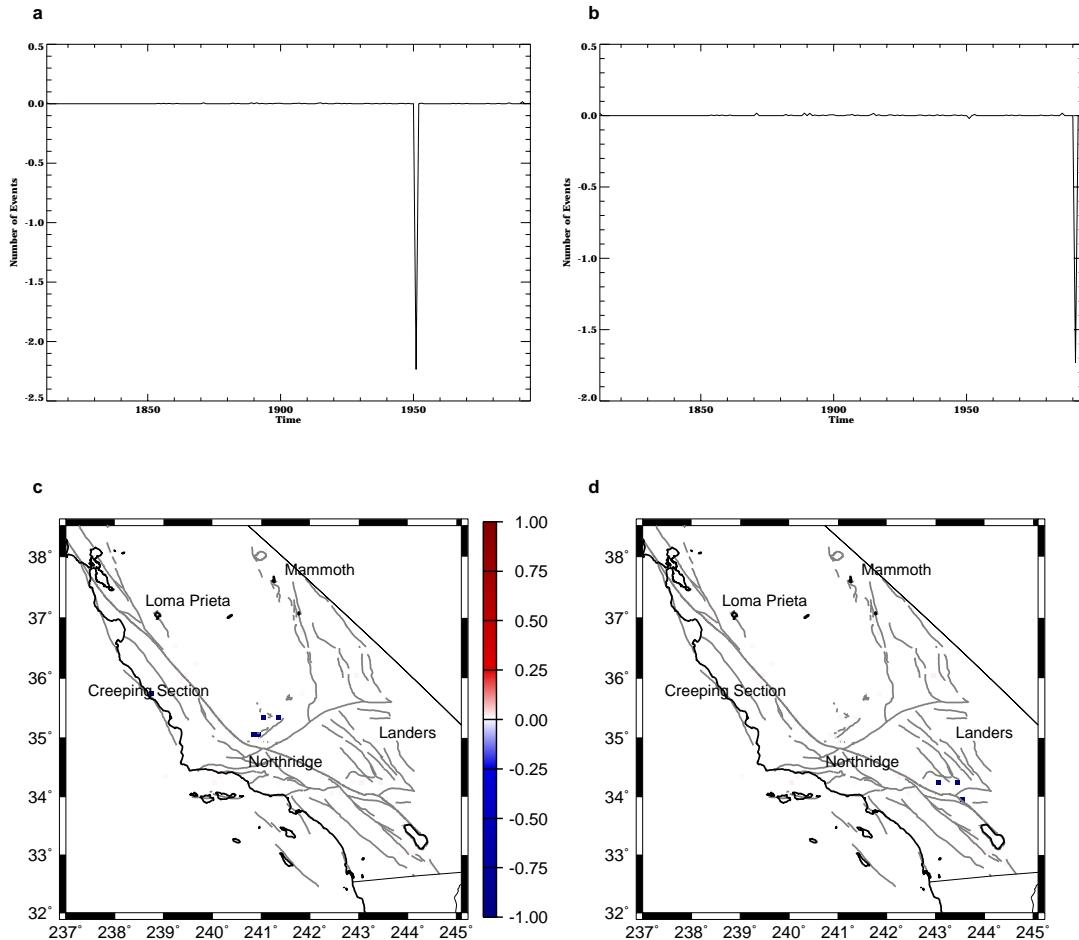


Figure 3.36: First and second KLE modes for southern California seismicity, 1812 - 1994,  $M > 6.0$ . a) PC time series to first KLE mode; b) first KLE mode; c) PC time series to second KLE mode; and d) second KLE mode, all normalized to the maximum.

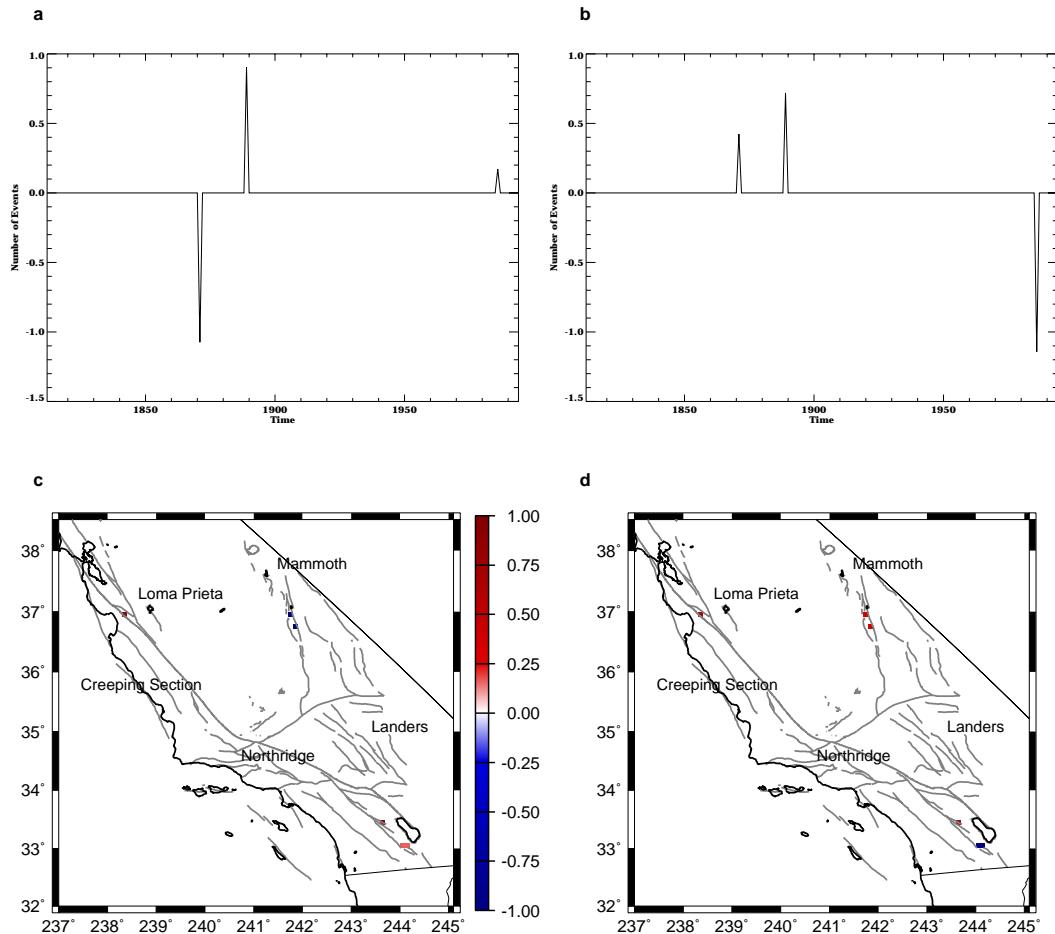


Figure 3.37: Third and fourth KLE modes for southern California seismicity, 1812 - 1994,  $M > 6.0$ . a) PC time series to third KLE mode; b) third KLE mode; c) PC time series to fourth KLE mode; and d) fourth KLE mode, all normalized to the maximum.

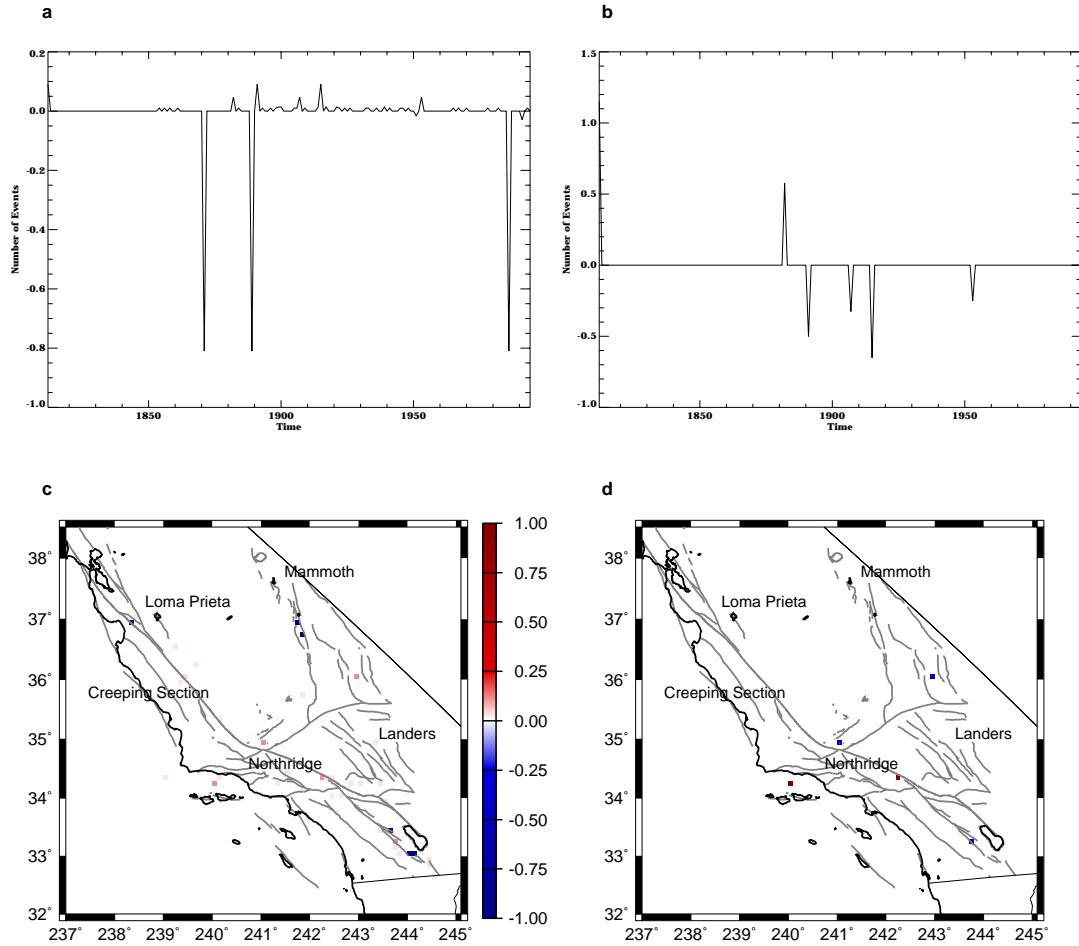


Figure 3.38: Fifth and sixth KLE modes for southern California seismicity, 1812 - 1994,  $M > 6.0$ . a) PC time series to fifth KLE mode; b) fifth KLE mode; c) PC time series to sixth KLE mode; and d) sixth KLE mode, all normalized to the maximum.

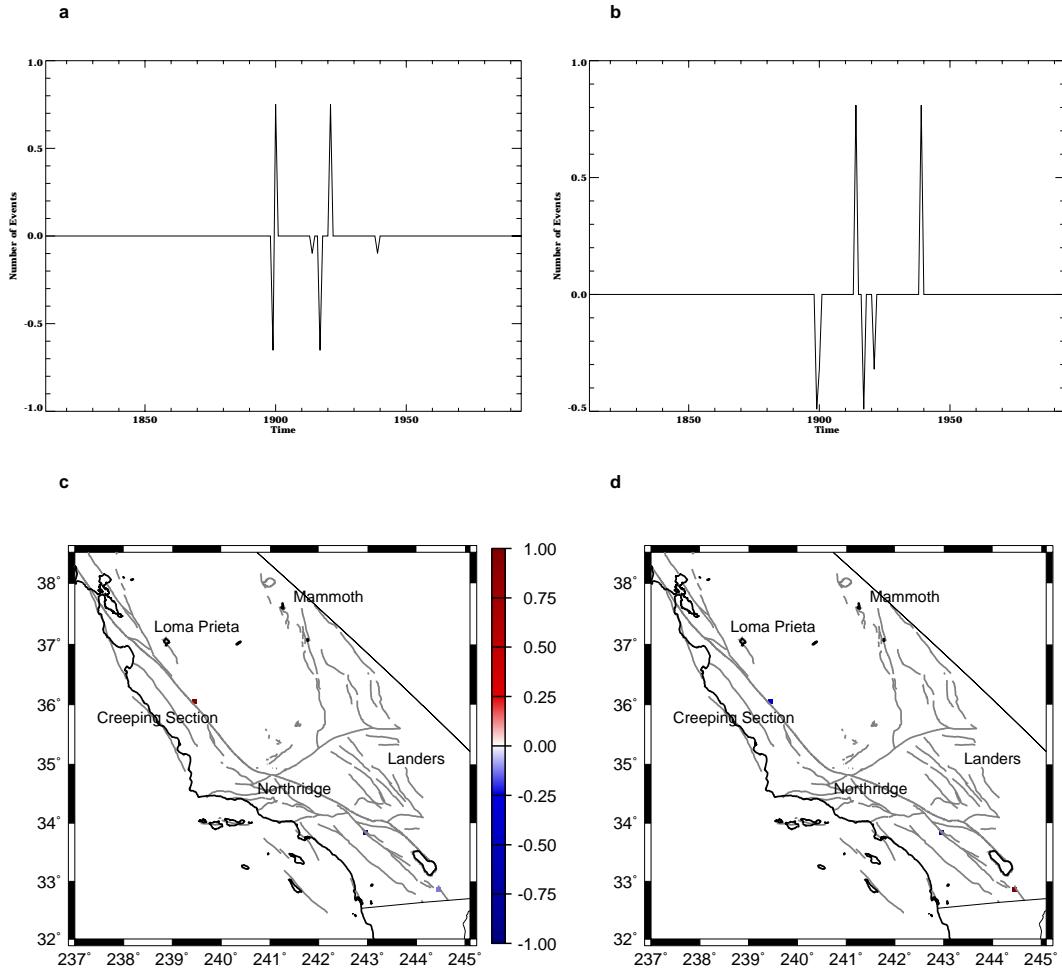


Figure 3.39: Ninth and tenth KLE modes for southern California seismicity, 1812 - 1994,  $M > 6.0$ . a) PC time series to ninth KLE mode; b) ninth KLE mode; c) PC time series to tenth KLE mode; and d) tenth KLE mode, all normalized to the maximum.

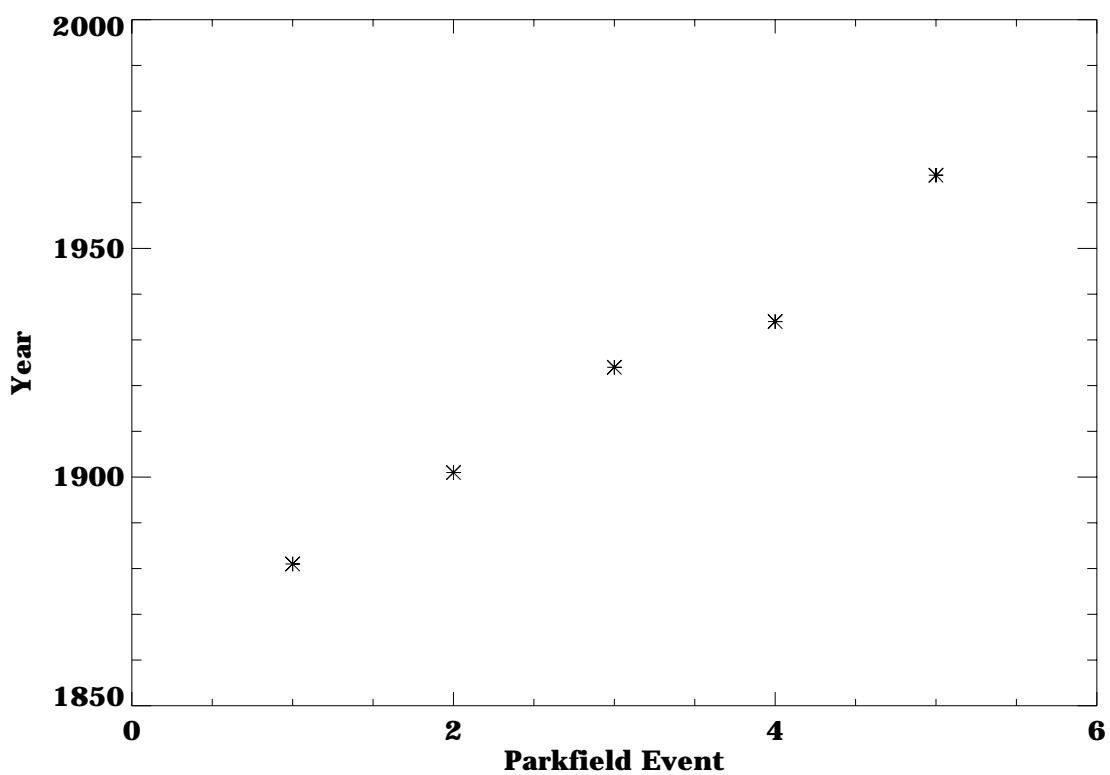


Figure 3.40: Major earthquakes at Parkfield, California, since 1850 [from Turcotte, 1997].

### 3.5 Precursory Modes

The presence of both large and small scale correlations in the data, evident in the KLE decompositions shown above, prompted a study of the change in these modes for each year, in an attempt to identify modes which consistently appear over some identifiable time period prior to an event. While the complex rate correlation operator,  $K(\mathbf{x}_i, \mathbf{x}_j)$ , can be used to compute the probability of future events on a fault patch model producing events over time periods of thousands of years (Section 2.3 above), its application to historic seismicity data is limited. Neither the long time periods nor the large number of moderate to large events produced by numerical simulations are available in the actual data, nor is the same accuracy in time and space possible. Consequently, the following method was developed.

If the seismicity in a given year,  $S$ , is known, and the eigenmodes, or eigenvectors  $e_i$ , are calculated using all seismicity data (a total of  $i = 1, \dots, N$  sites), then the eigenvectors are a complete, orthonormal set of basis functions, and any seismicity over that space can be decomposed into those eigenvectors.

$$S = \sum_{i=1}^N \alpha_i e_i, \quad (3.1)$$

where  $\alpha_i$  are the eigenvalues for that particular year. The eigenvalues,  $\alpha_i$ , are then computed from

$$\alpha_i = \sum_{i=1}^N e_i S_i. \quad (3.2)$$

Computing the  $\alpha_i$  for any given year, given the KLE decompositions above, is a relatively simple process. The data set used was that described in Section 3.4.1 above. The resulting  $\alpha_i$  for the time period prior to the 1992 Landers sequence are plotted in Figures 3.41. Events which occur in the data set, for example, the 1979 Imperial Valley event or the 1983 Coalinga earthquake, have signal in the

corresponding eigenmodes, and would be expected to produce signal prior to those events. The  $\alpha_i$  appear to change in both a systematic way in the lower eigenmodes over the several years prior to 1992. Interestingly, the modes that increase the most are those that include the Landers event - modes 2, 3, 8, and 9 (see Figures 3.31, 3.32, and 3.33 above).

The same analysis was performed with a cutoff at year 1978, just prior to the 1979 Imperial Valley event. Again,  $\alpha_i$  are plotted in Figures 3.42. Systematic changes in  $\alpha_i$ , particularly over mode five, as well as some of the lower modes, are apparent again, so that these particular modes were summed for each year, so that

$$\psi^{a-b} = \sum_{j=a}^b \alpha_j e_j, \quad (3.3)$$

where  $a$  and  $b$  are the indices to be summed over, those which appear to change prior to the event.

The important point in studying these summed modes is to remember that, in constructing the original decomposition modes, only catalog data up until December 31 of 1978 was used. In other words, no information from the upcoming Imperial Valley earthquake of 1979 was used to construct the precursory modes.

It was discovered that, while the variation over one year was not significant, a summation over several years prior to the earthquake produced an increase in signal in the area of that earthquake. A summation of  $\psi^{2-100}$  over several years, in this case 1965 to 1972, is shown in Figure 3.43. The location of the 1971 San Fernando earthquake, which has just occurred, is shown in blue. As the summation is increased for the next few years, interesting features develop (see Figure 3.44). Note the appearance of red signal at the location of the future Landers sequence, almost 20 years in the future, as well as the formation of a large red area to the south, in the location of the 1979 Imperial Valley earthquake.

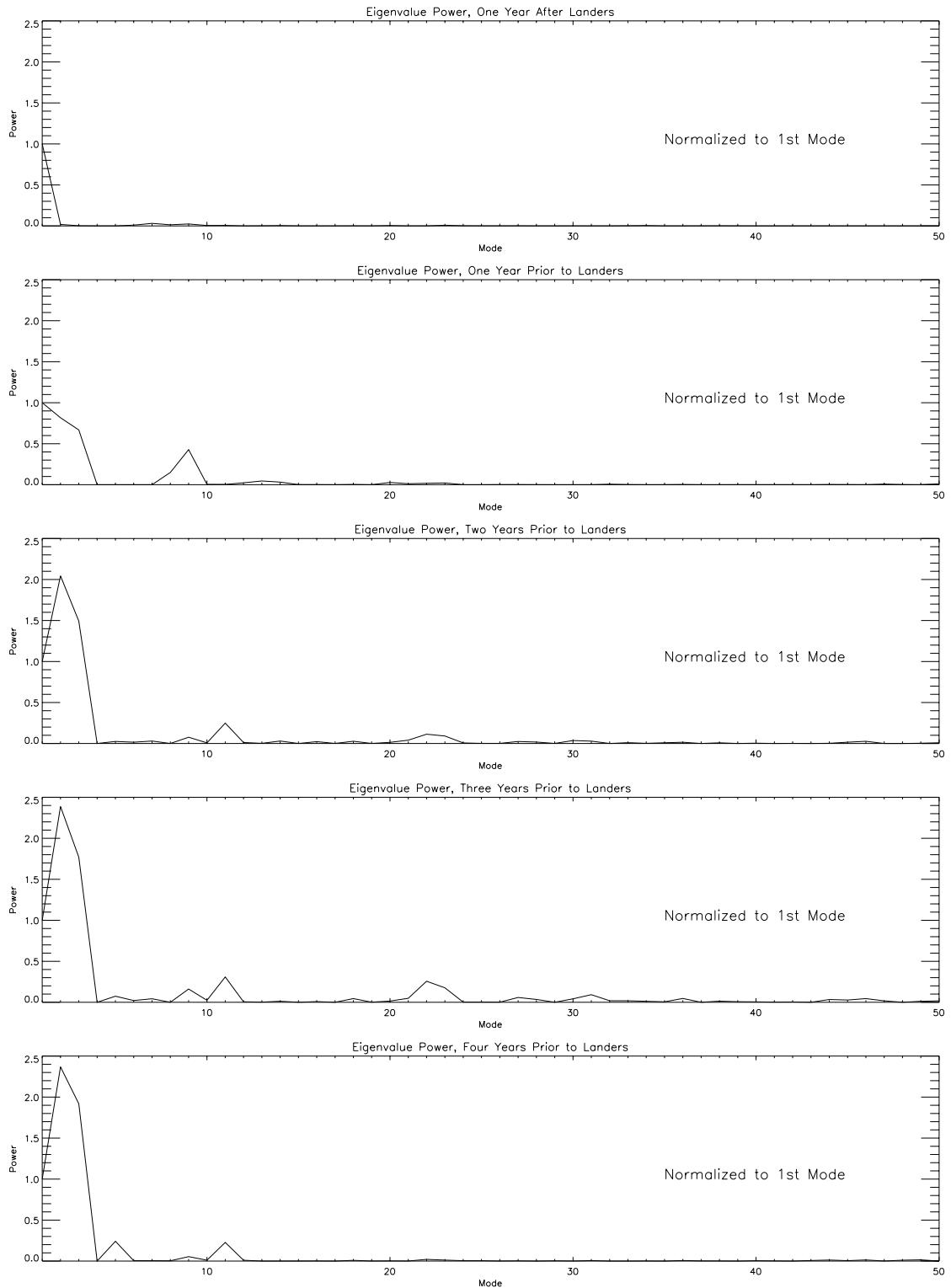


Figure 3.41:  $\alpha_i$ , plotted for the five years prior to the 1992 Landers earthquake. Values are normalized to the first mode.

Finally, the decrease in noise and the corresponding increase in size and intensity of the red signal shown in the 1979 plot, Figure 3.45, led to the formation of the pattern dynamics theory detailed in the next two chapters.

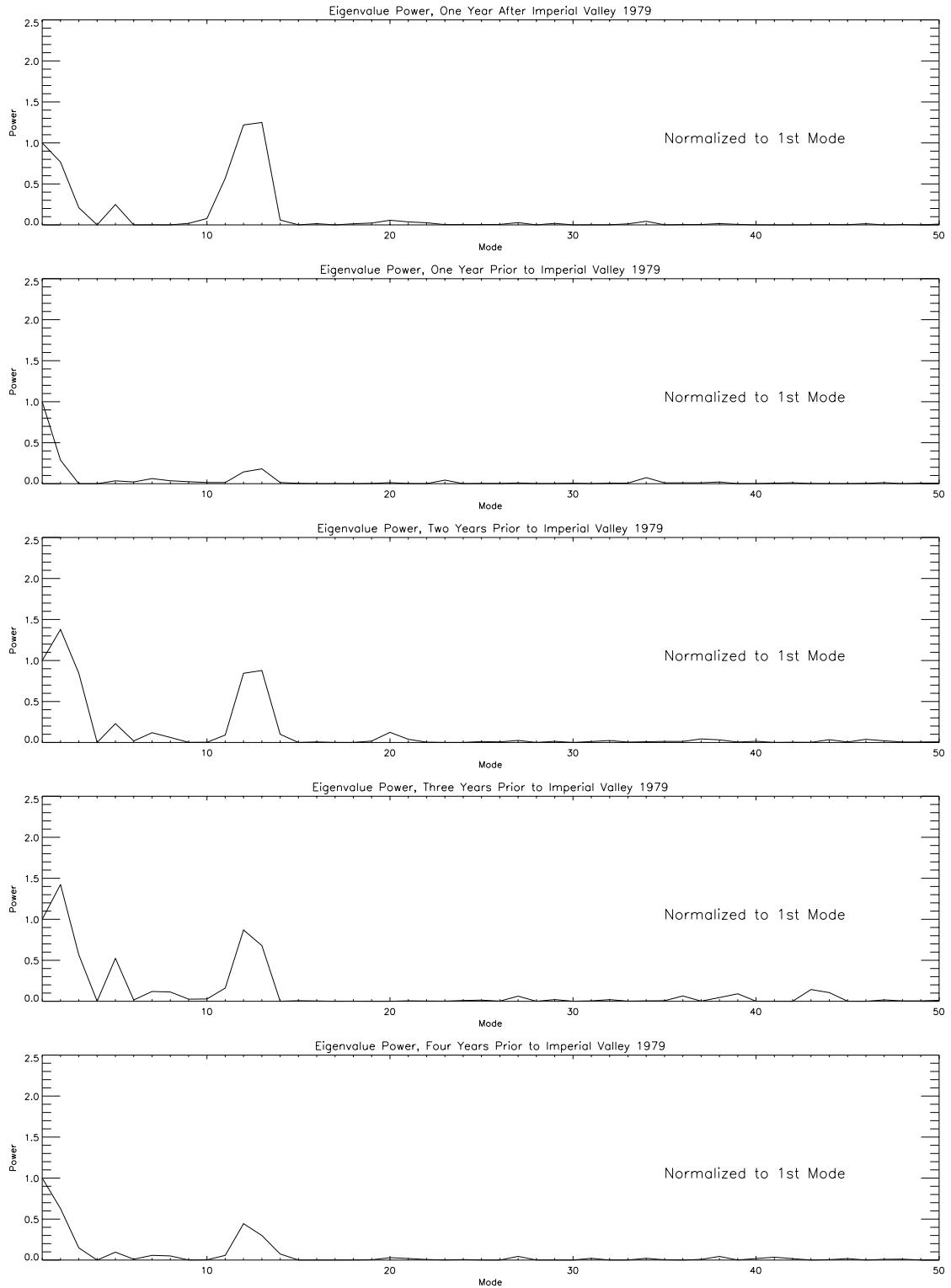


Figure 3.42:  $\alpha_i$ , plotted for the five years prior to the 1979 Imperial Valley earthquake. Values are normalized to the first mode.

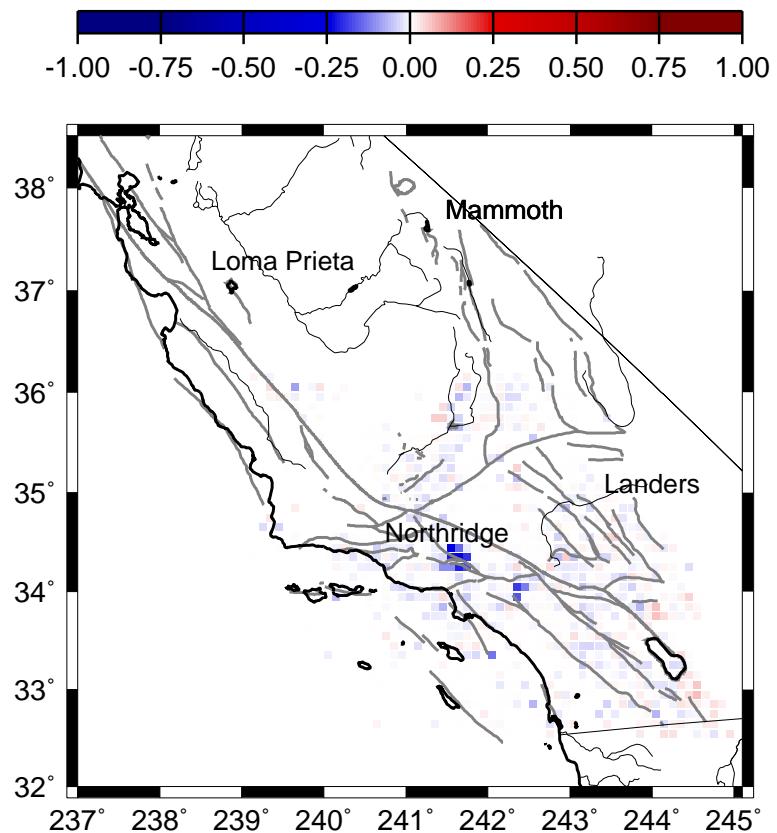


Figure 3.43:  $\psi^{2-100}$ , summed over the years 1971 to 1972. Values are normalized to the first mode.

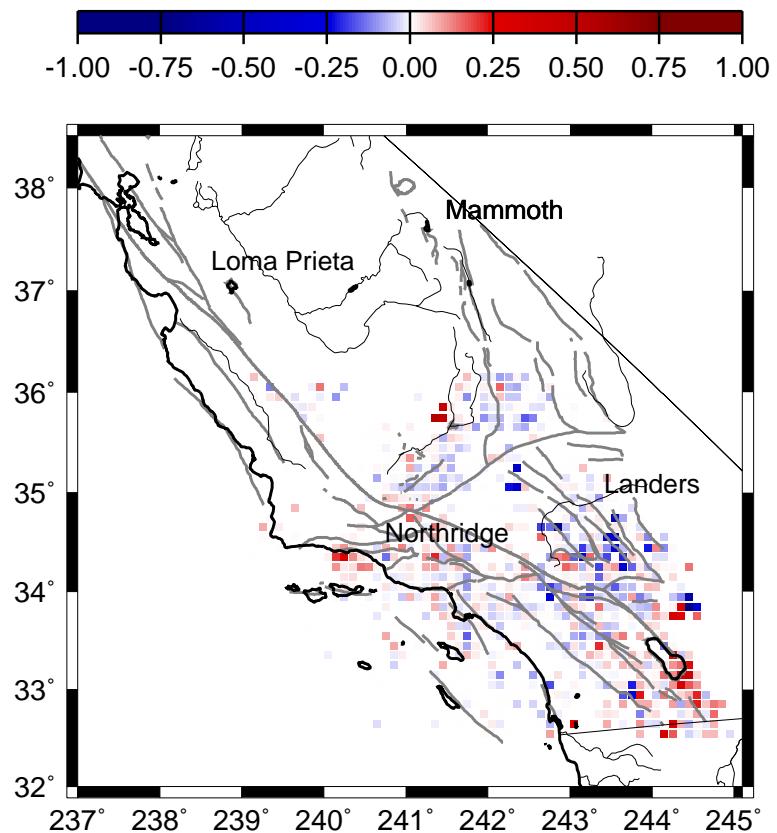


Figure 3.44:  $\psi^{2-100}$ , summed over the years 1971 to 1975. Values are normalized to the first mode.

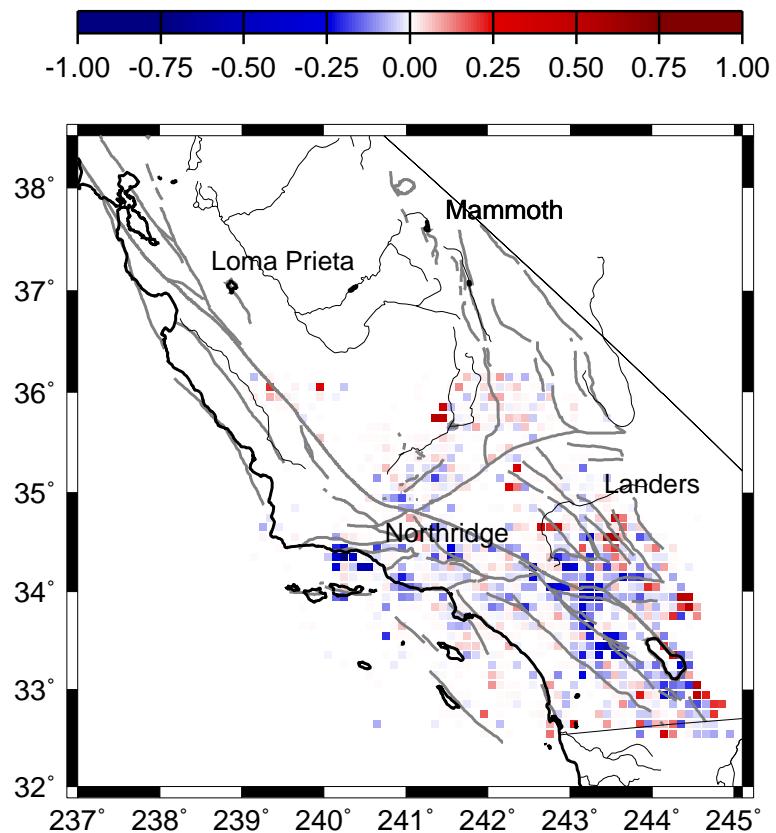


Figure 3.45:  $\psi^{2-100}$ , summed over the years 1971 to 1979. Values are normalized to the first mode.

## Chapter 4

### Dynamics - Theory

Application of the rate correlation operator,  $K(\mathbf{x}_i, \mathbf{x}_j)$ , to observed seismicity data in a region such as southern California, where the available data spans decades, not millennia, is not practical. However, several methods based upon learning patterns which are then used to forecast future events have been documented by Keilis-Borok et al. (1998) and Press and Allen (1995), for example. Here, the pattern dynamics method suggested by the data analysis above is extended in the theory first detailed by Rundle et al., 2000 (2).

Systems in which the dynamical variables of importance can be expressed as a phase angle are analyzed by Mori and Kuramoto as phase dynamical systems. Phase dynamics is a general reduction approach in which the motions exhibited by ordered patterns deformed under a variety of physical effects are interpreted in view of the phase angle variables. These phase angle variables characterize the system's development in phase space [Mori and Kuramoto, 1998]. Remember that, from equation 2.19,

$$\psi_R(\mathbf{x}_i, t) = \sum_{i=1}^N \alpha_n e^{-i\omega t} \phi_n(\mathbf{x}_i) \quad (4.1)$$

In other word, the evolution of pattern states is via rotations in an  $N$ -dimensional complex correlation space. This suggests that the time average of these rotations should be zero, as in a random walk. As the dynamics of the

system is represented by the incoherent superposition of these pattern states over sufficiently large region, the spatial mean should vanish also.

Figure 4.1 shows the variation in seismicity over time in southern California [Turcotte, 1997]. Note the increased number of events in 1987, 1992, and 1994, which raises the Gutenberg-Richter relation,  $a$ , without affecting the slope,  $b$ . This effect is attributed to the aftershocks of the Whittier, Landers, and Northridge earthquakes, respectively. If these aftershocks are removed, the background seismicity in southern California is nearly uniform in time [Turcotte, 1998]. Therefore, the mean rate of seismicity per year in southern California goes to a constant value of  $N$  over a large enough area. It is the localized, persistent deviations from this mean rate that are the surface expression of the underlying correlations.

Define the seismicity function for a region,  $S(\mathbf{x}_i, t_0, t)$ , as the time average at  $\mathbf{x}_i$  of  $\psi_{obs}(\mathbf{x}_i, t)$  over the time period  $[t_0, t]$ , so that

$$S(\mathbf{x}_i, t_0, t) = \frac{1}{t - t_0} \int_{t_0}^t \psi_{obs}(\mathbf{x}_i, t') dt'. \quad (4.2)$$

Note that the seismicity in a region such as southern California is extremely noisy, as seen in Figure 1.1. In the past, attempts to detect systematic variations in both the amplitude and phase of the seismicity, or in the rate of seismicity, were relatively unsuccessful (Sholz, 1990, Turcotte, 1997). The two most successful of these methods are the decrease in  $b$ -value in the local Gutenberg-Richter relation, considered a measure of precursory quiescence, and the time-to-failure criterion, which detects foreshocks [Bufe and Varnes, 1993; Jaume and Sykes, 1999]. Both of these methods have had intermittent results, and are mutually exclusive. They also involve the subjective identification of an appropriate area or volume for consideration over a well-defined time period.

Under phase dynamics,  $S(\mathbf{x}_i, t_0, t)$  is the real part of a complex-valued seismicity phase function [Rundle et al., 2000 (2)]. As the background rate of seis-

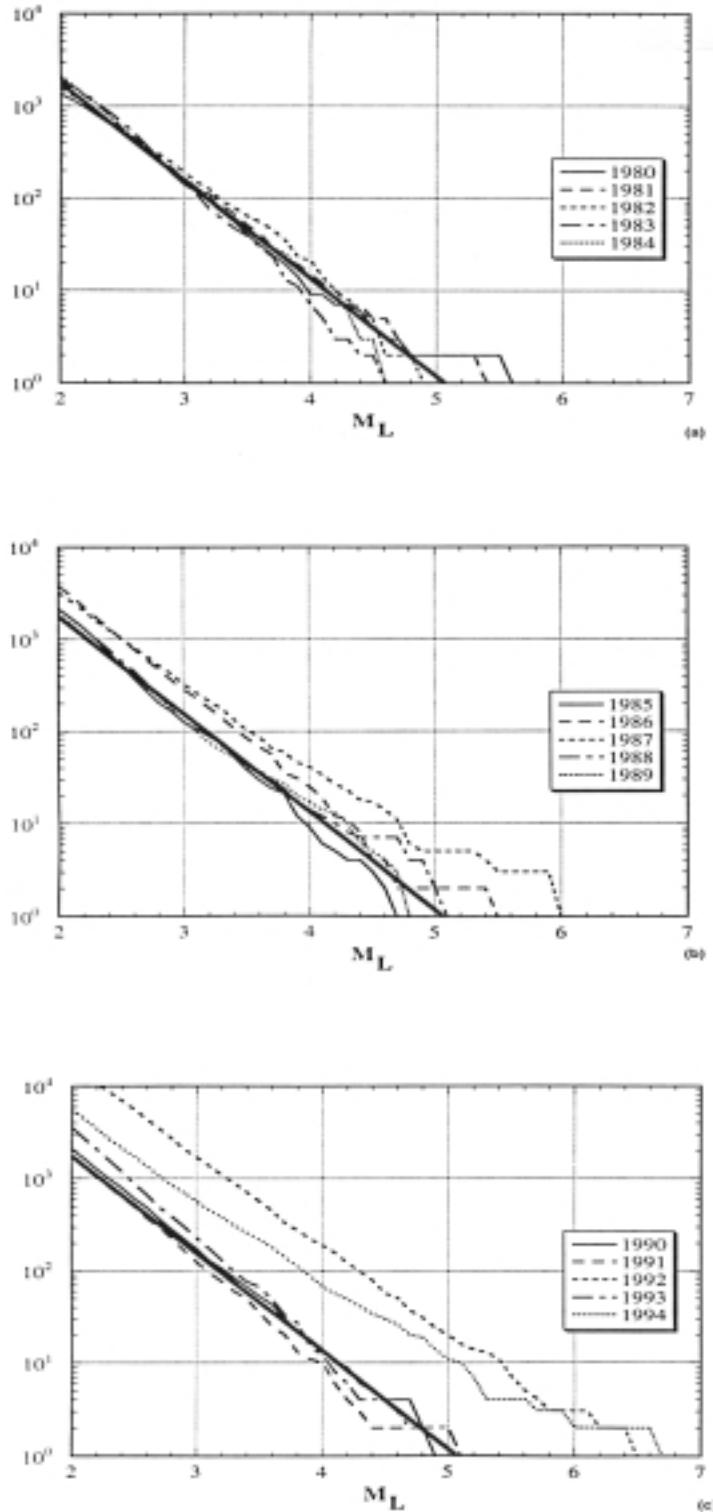


Figure 4.1: Cumulative number of earthquake,  $N$ , with magnitude greater than  $m$  for each year between 1980 and 1994 for southern California. a) 1980-1984; b) 1985-1989; c) 1990-1994. The solid straight line is the Gutenberg-Richter relation with  $b = 1.05$  and  $a = 2.06 \times 10^5 \text{ yr}^{-1}$  [Turcotte, 1997].

micity is constant over time, the amplitude of this phase function is unimportant - it is the localized spatial deviations from that background rate, whether they be increases or decreases in seismic activity, that account for the correlations in the fault system. It is the persistent rotation in space of this phase function that contains the important information about the regional seismicity.

Consider  $S(\mathbf{x}_i, t_0, t)$  to be  $N$ -dimensional vector  $\mathbf{S}(t_0, t)$  that has an inner product and a well-defined  $L^2$  norm,  $\|\mathbf{S}(t_0, t)\|$  [Holmes et al., 1996]. Then, define  $\hat{\mathbf{S}}(t_0, t)$  to be a unit vector in the  $N$ -dimensional correlation space pointing in the direction of  $S(\mathbf{x}_i, t_0, t)$ , tail fixed at the origin and head on an  $N$ -dimensional unit sphere. An incoherent superposition of functions must have a zero mean, we can remove the spatial mean, and divide by the standard deviation. Therefore, from Rundle et al., 2000 (2),

$$\hat{\mathbf{S}}(t_0, t) = \frac{S(\mathbf{x}_i, t_0, t) - \bar{S}(\mathbf{x}_i, t, t)}{\sigma}, \quad (4.3)$$

where

$$\bar{S}(\mathbf{x}_i, t_0, t) = \frac{1}{N} \int_{all \ x} S(\mathbf{x}_i, t_0, t) dx, \quad (4.4)$$

is the spatial mean of  $S(\mathbf{x}_i, t_0, t)$  and

$$\sigma^2 = \|\mathbf{S}(t_0, t)\|^2 = \frac{1}{N} \int_{all \ x} [S(\mathbf{x}_i, t_0, t) - \bar{S}(\mathbf{x}_i, t_0, t)]^2 dx. \quad (4.5)$$

Note that, while  $\mathbf{S}(t_0, t)$  is a local function, depending only on conditions at  $\mathbf{x}_i$ . However, due to the normalization, which involves information from the entire active region of  $N$  boxes,  $\hat{\mathbf{S}}(t_0, t)$  is a spatially non-local function. because the norm,  $\sigma$ , is a non-local function, integrated over all  $\mathbf{x}_i$ . The function  $\hat{\mathbf{S}}(t_0, t)$  is a function of the conditions everywhere in the system. For a system in which the correlation lengths were small, there would be little effective difference between

equation 4.2 and equation 4.3. However, in a strongly correlated system, the effect may be substantial. The recent theoretical and simulation results, as outlined in Chapter 2, that support the existence of long-range correlations in the earthquake fault are confirmed by the results in Chapter 3, above.

Remember that  $\hat{\mathbf{S}}(t_0, t)$  is an  $N$ -dimensional vector which represents a rotation of the seismicity in phase space. This persistent variation in one direction, or persistent rotation of that vector, is the distance from the time average of the random walk which is a function of the noise in the system. Quantification of that vector is a measure of the system attractor which is a function of the underlying stresses that signal a future event. Under the phase dynamics assumption, the important change in seismicity  $\Delta\hat{\mathbf{S}}(t_1, t_2)$  corresponds to rotations in correlation space over the time period  $[t_1, t_2]$ , where  $t_2 = t_1 + \Delta t$ , such that

$$\Delta\hat{\mathbf{S}}(t_1, t_2) = \hat{S}(\mathbf{x}_i, t_0, t_2) - \hat{S}(\mathbf{x}_i, t_0, t_1). \quad (4.6)$$

The physically meaningful changes in  $\hat{\mathbf{S}}(t_0, t)$  are captured by rotations about its origin [Rundle et al., 2000 (2)]. In addition, using equation 4.2 above,

$$|\Delta\hat{\mathbf{S}}(t_1, t_2)|^2 \propto |\Delta\psi_{obs}(\mathbf{x}_i, t_1, t_2)|^2. \quad (4.7)$$

From equation 2.22,

$$\Delta P(\mathbf{x}_i, t) \propto |\Delta\psi_R(\mathbf{x}_i, t)|^2 = \{\Delta\psi_{obs}(\mathbf{x}_i, t)\}^2,$$

so that

$$\Delta P(\mathbf{x}_i, t) \propto |\Delta\hat{\mathbf{S}}(t_1, t_2)|^2. \quad (4.8)$$

Because probability is a conserved quantity we can write

$$\int_{all x} \Delta P(\mathbf{x}_i, t) dx = 0, \quad (4.9)$$

then  $\Delta P(\mathbf{x}_i, t)$ , the increase or decrease in the probability of future seismicity, is

$$\Delta P(\mathbf{x}_i, t) = \left| \Delta \hat{\mathbf{S}}(t_1, t_2) \right|^2 - \mu_P, \quad (4.10)$$

where  $\mu_P$  is the spatial mean of  $\Delta P(\mathbf{x}_i, t)$  [Rundle et al., 2000 (2)],

$$\mu_P = \int_{all \, x} \left| \Delta \hat{\mathbf{S}}(t_1, t_2) \right|^2 dx. \quad (4.11)$$

In the following chapter I calculate the change in probability,  $\Delta P(\mathbf{x}_i, t)$ , for the southern California region, using observed seismic data and compare the results to randomly generated catalogs and seismic hazard analyses.

## Chapter 5

### Pattern Dynamics Applied to Southern California Seismicity

#### 5.1 Introduction

Earthquakes strike populated regions without warning, causing great destruction and loss of life. One tragic example is the recent Izmit, Turkey, earthquake of August 17, 1999, which resulted in the deaths of over 17,000 persons. Many similar examples have been documented over the course of time [Richter, 1958; Kanamori, 1981; Scholz, 1990; Turcotte, 1997]. While a long-sought goal of earthquake research is the reliable forecasting of these great events, very little progress has been made in developing a successful, consistent methodology [Kanamori, 1981; Geller et al., 1997; Sykes et al., 1999]. Most scientists agree that events of this magnitude must be preceded by some causal process. Yet despite the fact that the largest of these events span distances of more than 500 km, no reliable precursors have been detected with any repeatability. In Chapter 3, I documented the discovery of correlations in the historic seismicity data for southern California. Chapter 4 discussed the theory that fluctuations in the same seismicity measure systematic space-time variations in seismicity from southern California using a new technique. This procedure is based upon the idea that seismic activity corresponds geometrically to the rotation of a pattern state vector in the high-dimensional correlation space spanned by the eigenvectors of a correlation operator [Rundle et al., 2000 (1)]. To emphasize the connection to phase

dynamics, the function  $\Delta P$  is called the Phase Dynamical Probability Change (PDPC).

Here I show that the analysis of data obtained prior to December 31, 1991 reveals coherent regions often associated with the future occurrence of major earthquakes in the same areas, including the 1992 Landers earthquake sequence, the 1994 Northridge event, and the 1999 Hector Mine earthquake. The technique described here is not a model, rather it is a new method for processing seismicity data to reveal underlying space-time structure, based upon the physics of critical point systems and the analysis of the fault system as a phase dynamical system. In addition, I test the success of this method by applying the same method to random catalogs, and apply a likelihood test to both the actual historic and the random catalogs for comparison [Tiampo et al., 2000].

## 5.2 Data

The observed seismicity data set employed in this pattern dynamics analysis is the same as that used in Section 3.4.2 above, existing observations in southern California from the SCEC database between the years 1932 and the present [<http://www.scec.org>]. Using only the subset of this data at locations  $x$  in southern California and covering the period from January 1, 1932 through December 31, 1991, I compute the PDPC function  $\Delta P = P(\mathbf{x}_i, t)$  to detect anomalous spatial correlations at sites in southern California prior to January 1, 1992. Note that I use only events having magnitude  $M > 3.0$ , to ensure completeness of the catalog. The hypothesis to be tested is that these anomalous correlated regions are associated with large main shocks that occurred after January 1, 1992. Figure 1.1 shows the relative seismic activity for this period,  $S(\mathbf{x}_i, 1932, 1991)$ , superimposed on a map of southern California. It is clear that  $S(\mathbf{x}_i, 1932, 1991)$ , is an unremarkable function. For example, there is little evidence of any phenomena

precursory to the  $M \sim 7.3$  Landers, California event that occurred on June 28, 1992.

### 5.3 Results

The application of the PDPC method is as discussed in Chapter 4 above.  $S(\mathbf{x}_i, t_0, t)$  was computed by summing the seismicity from  $t_o$  to  $t$ , usually some year between 1970 and 1991. The function  $\hat{S}(t, t_0)$  is computed for two separate times. In the following cases,  $t_0$ , the base year, was 1965. A variety of initial and final years were chosen,  $t_{in}$  and  $t_{fi}$  respectively, such that

$$\Delta\hat{S}(\mathbf{x}_i, t_{in}, t_{fi}) = S(\mathbf{x}_i, t_0, t_{fi}) - S((\mathbf{x}_i, t_0, t_{in})) \text{ and} \quad (5.1)$$

$$\Delta P(\mathbf{x}_i, t_{fi}) = [\Delta\hat{S}(\mathbf{x}_i, t_{in}, t_{fi})]^2 - \mu_P. \quad (5.2)$$

$\Delta P(\mathbf{x}_i, t_{fi})$ , or PDPC, is normalized to the maximum value, and plotted for  $\mathbf{x}_i$ .

#### 5.3.1 The Phase Dynamical Probability Change, $\Delta P$

Figure 5.1 shows plots of  $\Delta P$ , using only existing seismicity data acquired prior to January 1, 1992, six months before the occurrence of the June 28, 1992 Landers, CA,  $M \sim 7.3$ , earthquake. Recall that an increase in  $\Delta P$  above the background level measured by  $\mu_P$  should be interpreted as a tendency to form a spatially correlated region of seismic activity, and that such regions evidently must be present for larger earthquakes to occur. The color coding on Figure 5.1 is scaled to the largest value of  $\Delta P$  on any of Figures 5.1a, b, c or d. The largest 30 percent of points is represented by red, approximately 20 percent by yellow, and approximately 50 percent are between white and green. The inverted blue

triangles represent events that occurred during the time period covered by the plot, to indicate colored boxes that need not be analyzed further.

Blue circles represent more recent events of magnitude  $M > 5.0$  that occurred after January 1, 1992. It should be emphasized again that no data for these more recent events was used in constructing the colored boxes in Figure 5.1. In particular, I include circles representing the 1992 Landers sequence and the recent  $M \sim 7.1$  Hector Mine events. These earthquakes are evidently associated with a long-lived arcuate structure of colored boxes cutting across the local fault geometry that began forming prior to 1980. This is confirmed by the arcuate structure seen in the correlations of Figures 3.25 and 3.26, and in the precursory modes for 1975 and 1979, from eigenvectors constructed only from data through 1978 (Figures 3.44 and 3.45). This structure continues down to the southeast of the Landers mainshock, east of the 1992 Joshua Tree earthquake, and the lack of subsequent activity to date may indicate this site as a potential rupture zone in the future.

Visual examination of Figure 5.1 indicates that recent large events (blue circles) that occurred after January 1, 1992 are clearly associated with detectable locations of positive  $\Delta P$  that formed prior to January 1, 1992. However there is clearly some variability, particularly for smaller events, depending on the choice of time interval  $(t_1, t_2)$ . Larger events tend to be associated with larger colored regions that form earlier and persist longer after the event. Since earthquake fault dynamics are now believed to be associated with critical phenomena (33, 40, 41 42), there may be a scaling relation between the area  $A$  of the correlated region and the time interval  $t$  prior to the main shock at which the correlation begins to form (43). Since the linear size of our boxes is approximately 11 km, one should not expect events significantly smaller than  $M \sim 6$ , whose characteristic linear source dimension is 10 km, to be well resolved by our procedure. Yet even the

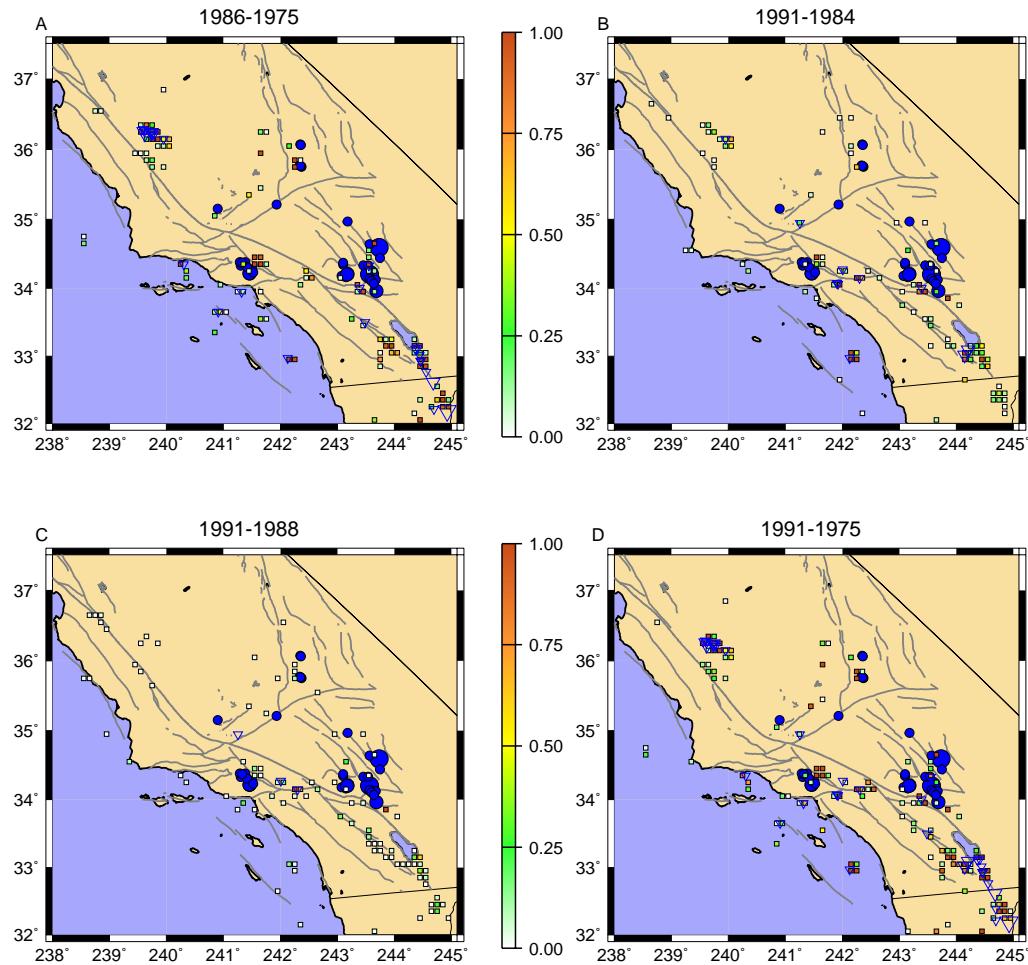


Figure 5.1: Maps of  $\Delta P$  for the time intervals a) 1986-1975, b) 1991-1984, c) 1991-1988 and d) 1991-1975. In constructing these maps, no data is used from time periods after December 31, 1991. Inverted triangles represent events that occurred during the indicated time periods, with three sizes corresponding to magnitudes of:  $5 < M < 6$ ,  $6 < M < 7$ , and  $M > 7$ . Filled circles represent events that occur after January 1, 1992. Three increasing circle sizes again correspond to the same magnitude ranges as for the inverted triangles.

smaller circles associated with  $M \sim 5 - 6$  events seem to occur in proximity to colored boxes, albeit those that appear only a short time interval before the main shock.

Also shown in Figure 5.2 is the progression of  $\Delta P(\mathbf{x}_i, t_{fi})$  from  $t_{fi} = 1978$  to  $t_{fi} = 1991$ . Note the increasing signal strength at both the site of the future Landers event and the Northridge earthquakes.

### 5.3.2 Phase Angle Rotation, $\Delta\hat{S}$

As discussed below, it can be seen that regions with  $\Delta P > 0$  may correspond either to anomalous seismic activity or anomalous quiescence. In some locations, a region that was the site of a recent major earthquake may evidently indicate values of  $\Delta P > 0$  that are linked with a future large event that will occur at a neighboring, but disjoint, location. The positive value of  $P$  may appear only at the location of the past event, rather than at the neighboring location of the future event. In these cases, the neighboring aftershock zone of the past event is participating in the anomalous activity or quiescence that defines the future event. In addition, anomalous activation or quiescence may appear in events having magnitudes less than the uniform cutoff,  $\Delta M > 3$ .

An example of this phenomenon is the colored area at the location of the 1971 San Fernando earthquake, shown in Figure 5.1a, that are evidently associated with the coming 1994 Northridge event. However, this area is present not because it represents aftershock activity from the 1971 San Fernando event (anomalous seismic activity), rather it represent an area of anomalous quiescence.

In order to illustrate that these areas of anomalous correlated regions are not simply an artifact of aftershock sequences to large events and that the San Fernando region is anomalously quiescent prior to Northridge, I plot the change in the normalized seismicity  $\Delta\hat{S}(\mathbf{x}_i, 1978, 1991)$ , as defined below, over the years

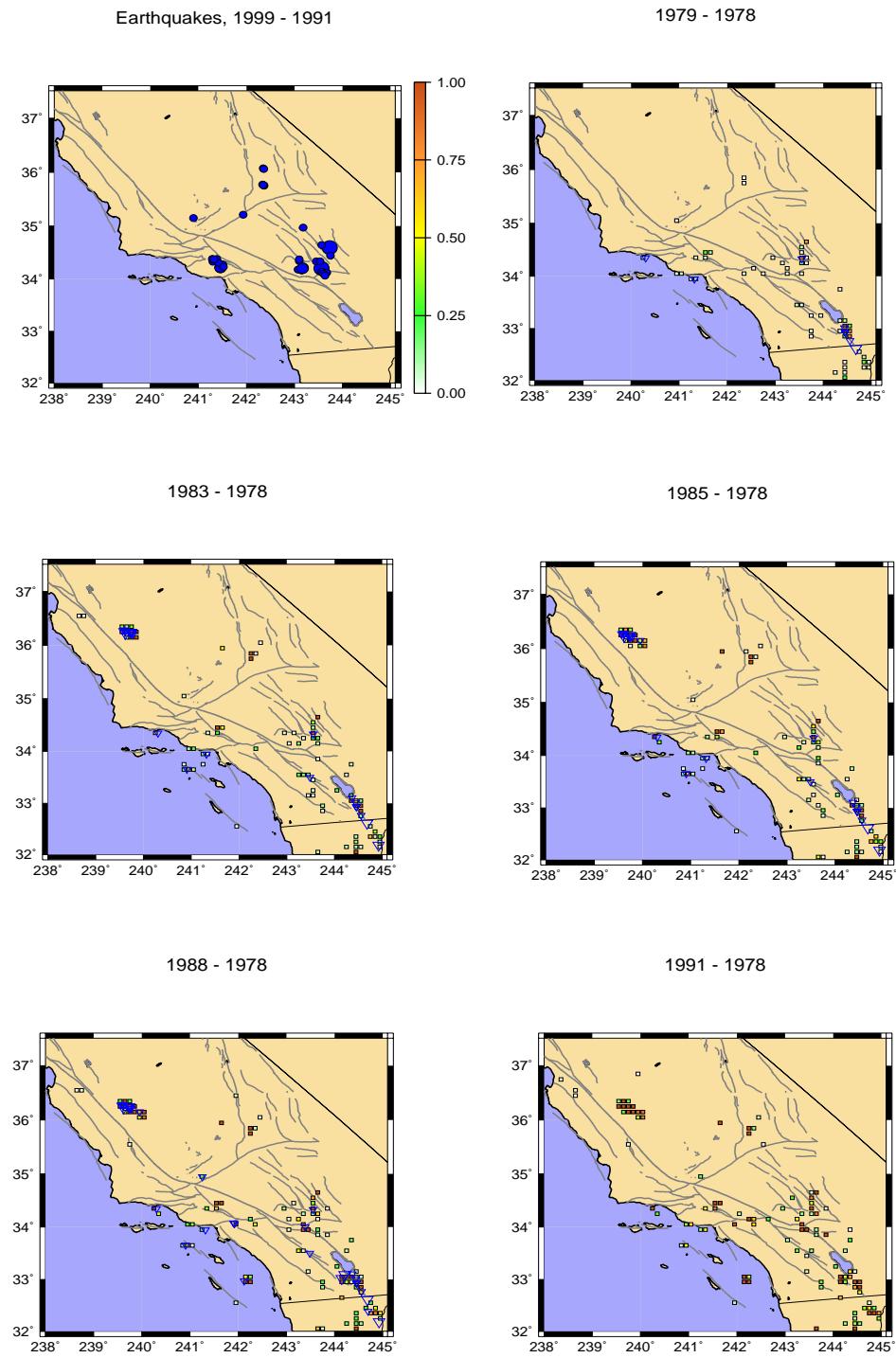


Figure 5.2: Maps of  $\Delta P$  for the time intervals shown. The plot in the top right-hand corner shows the events which occur between 1991 and 1999. Again, no data is used from time periods after December 31, 1991. Inverted triangles represent events that occurred during the indicated time periods. The size of the symbols corresponds to magnitudes as in Figure 5.1 above.

1978-1991. In Figure 5.3, a positive (red) value of  $\Delta\hat{S}$  represents anomalous seismic activity, and a negative (blue) value of  $\Delta\hat{S}$  represents anomalous quiescence. The area that appear red in the plot of  $\Delta P$  in figure a is seen to be blue in Figure 5.3, indicating that the positive value of  $\Delta P$  arises from anomalous quiescence in the San Fernando aftershock zone. In addition, it will be seen below that the construction of  $P(\mathbf{x}_i, 1978, 1991)$  actually subtracts away the effect of any changes in seismicity prior to 1978, so San Fernando aftershock activity between 1971-1978 does not contribute. Figure 5.3 also shows that the 1983 Coalinga earthquake displays seismic activation, while the Landers sequence is a complicated mix of positive and negative  $\Delta\hat{S}(\mathbf{x}_i, 1978, 1991)$ . These results support the conclusion that the PDPC function does not simply identify areas associated with past events and their aftershock sequences, rather it quantifies the underlying stress coherence and correlations associated with the regional seismicity.

### 5.3.3 Cumulative probability

If the quantity  $|\Delta\hat{S}(\mathbf{x}_i, t_1, t_2)|^2$  is proportional to the change in probability accumulated over time, as seen above, then it is, by definition, proportional to the cumulative probability at each particular location. Here I look at the plot of  $|\Delta\hat{S}(\mathbf{x}_i, t_1, t_2)|^2$  at the site of the historic Parkfield events.

Figure 5.4 is a plot of  $|\Delta\hat{S}(\mathbf{x}_i, t_1, t_2)|^2$  vs. year for the time period 1970 to 1990. Note that this time period begins a few years after the 1966 Parkfield event and continues past the time when the next Parkfield event was considered due, 1988, based on the recurring earthquakes theory. The initial time  $t_1$  is fixed at 1970, the final time,  $t_2$ , varies through 1990. The cumulative probability is decreasing in the early 1970s, as would be expected from the fact that the 1966 event has just occurred. The probability begins to rise in 1971, after the San Fernando event, but drops steeply again in 1978, when the Santa Barbara Channel event

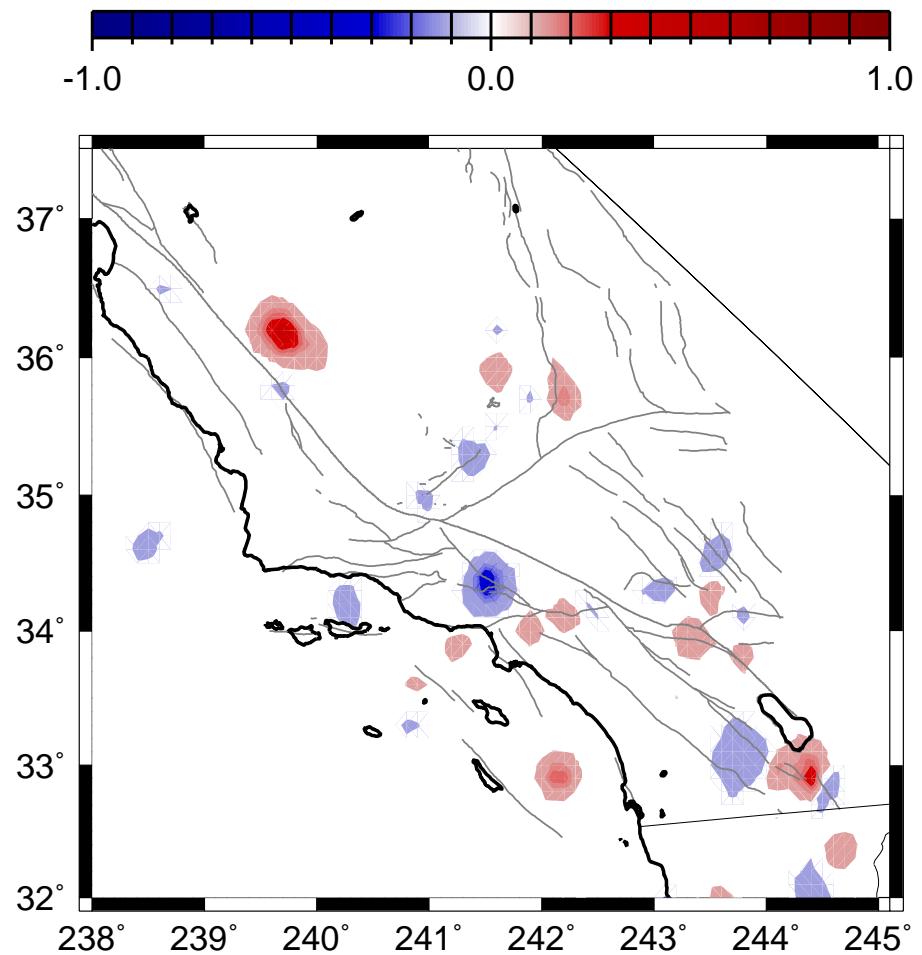


Figure 5.3: Map of  $\Delta\hat{S}(x_i, 1978, 1991)$ , normalized to the maximum absolute value. Red indicates an increase in  $\Delta\hat{S}$ , blue a decrease in  $\Delta\hat{S}$ .

occurs. It begins to rise again in 1979. Interestingly, there is no corresponding drop in 1983, with Coalinga earthquake, an event that occurred in close proximity to the east of Parkfield. The cumulative probability continues to rise past 1988, and should be studied for later time periods to assess the current status.

While these results are not definitive, as there is no guarantee that the Parkfield event will again nucleate in the same location, there is certainly evidence to support the theory that  $|\Delta \hat{S}(\mathbf{x}_i, t_1, t_2)|^2$  is related to cumulative probability.

#### **5.4 Random Catalogs**

Each random catalog was constructed from the instrumental catalog by using the same total number of events, but assigning occurrence times from a uniform probability distribution over the years 1932-1991, and distributing them uniformly over the original locations. This procedure produces a Poisson distribution of events in space with an exponential distribution of interevent times. Randomizing the catalog in this way destroys whatever coherent space-time structure may exist in the data, thus effectively declustering the catalog. I applied the method to this random catalog and obtained the colored boxes shown in Figure 5.5, which corresponds to the same time period of Figure 5.1d, 1975-1991. There are many more colored boxes in Figure 5.5 than in 5.1d, and that the boxes are more broadly distributed in space.

#### **5.5 Likelihood Tests**

To test the hypothesis that the formation of correlated regions identified by this method are coherent space-time structures related to future large events, I carried out a likelihood ratio test on the PDPC solution to the actual historic catalog used in the above analysis.

In the likelihood ratio test method, I wish to compare the quality of forecasts

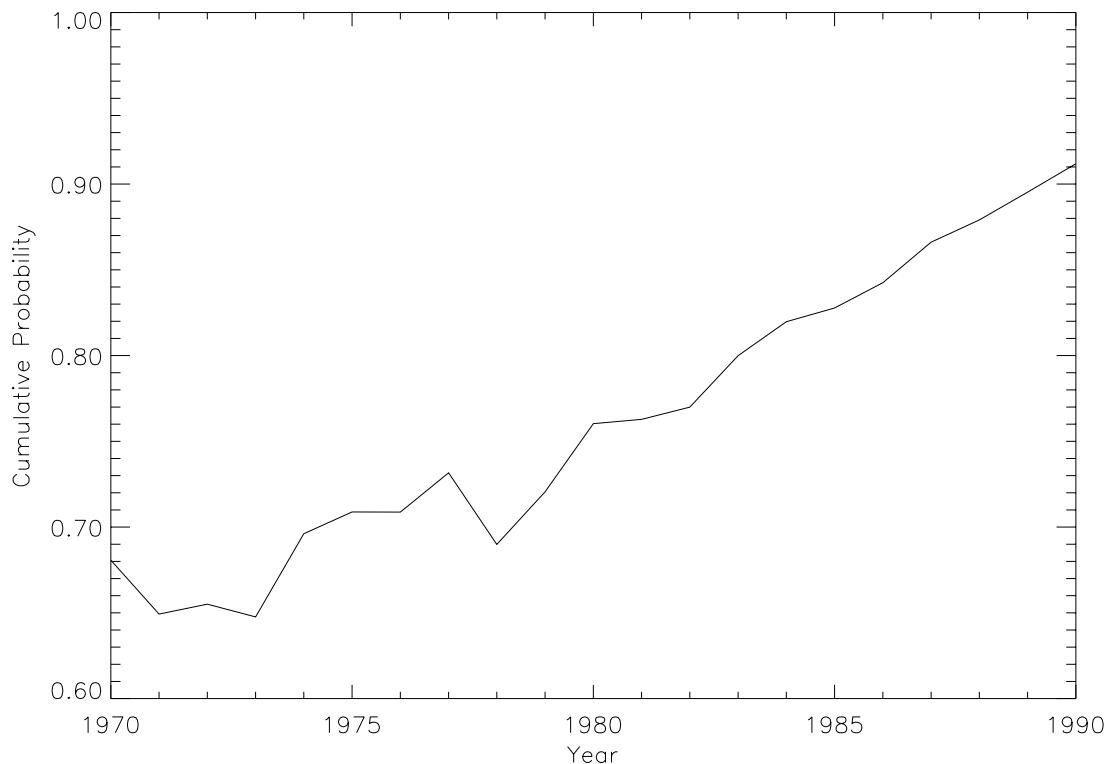


Figure 5.4:  $\left| \Delta \hat{S}(\mathbf{x}_i, t_1, t_2) \right|^2$  plotted for the time period 1970 to 1990, at the location of the historic Parkfield event.

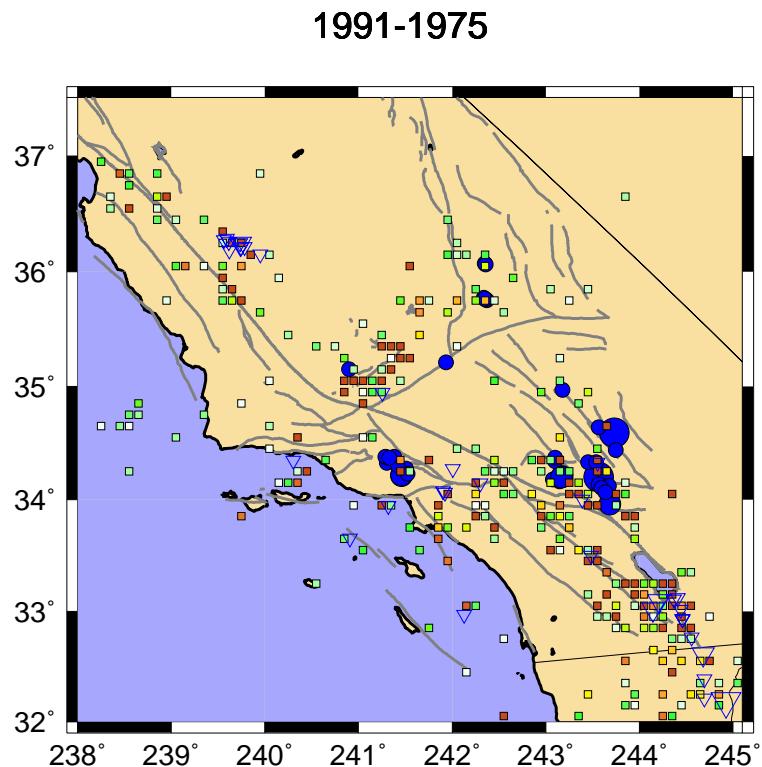


Figure 5.5: Map of  $\Delta P$  for one random catalog for the time period 1991-1978. Inverted triangles represent events that occurred during the indicated time periods. The sizes of both circles and triangles corresponds to magnitudes as in Figure 5.1 above.

given by two probability distributions, in this case the PDPC computed from actual seismicity data, and the PDPC computed on random catalogs. In the likelihood ratio test, a probability density function (PDF) is required. This test used a Gaussian distribution whose width is that of the original location grid, approximately 11 km, and whose peak value is given by  $\Delta P + \mu_P$ , superimposed over each location. Generally, for the normalized probability function, designated  $P(\mathbf{x}_i, t)$ , with  $N$  locations, then the likelihood function,  $L$ , is

$$L = \prod_{i=1}^N P(\mathbf{x}_i, t). \quad (5.3)$$

The likelihood test must compare the forecast with the actual events, in this case the location of the circles, which identify the locations of future events, as shown in Figure 5.1a. In this case, the likelihood provides a measure of how well the colored regions predict the actual events as quantified by the locations of the circles. The corresponding probability distributions are designated  $P_{act}(\mathbf{x}_i, t)$  and  $P_{circles}(\mathbf{x}_i, t)$ , respectively. To compare the forecasting ability, define a composite likelihood function,  $L(r)$ , which involves a mixing ratio  $r$ ,

$$L(r) = \prod_{i=1}^N P(\mathbf{x}_i, t, r), \quad (5.4)$$

where

$$P(\mathbf{x}_i, t, r) = rP_{actual}(\mathbf{x}_i, t) - (r - 1)P_{circles}(\mathbf{x}_i, t). \quad (5.5)$$

To compare two probability distributions, calculate the log-likelihoods  $\log L_{act}$  and  $\log L_{random}$ ,

$$\log L = \log_{10}\left(\frac{L(r)}{L_{circles}}\right). \quad (5.6)$$

$\log L_{random}$  is the log-likelihood computed for the random catalogs, i.e. instead of  $P_{act}(\mathbf{x}_i, t)$  used in the calculation of  $P(\mathbf{x}_i, t, r)$  in 5.5,  $P_{random}(\mathbf{x}_i, t)$  is used instead. The value of  $r$  used is that of the box size, 11km.

The log-likelihood, here designated  $\log L$ , is computed for each catalog, random and actual. A log-likelihood is considered a better fit to the forecast test if its value is larger, i.e., if

$$\log L_{act} \gg \log L_{random} \quad (5.7)$$

means that the PDPC method applied to the actual catalog produces better predictions than the PDPC method applied to a random catalog.

Note that this comparison assumes that the PDPC anomalies are statistically independent, an assumption that is probably not correct. However, statistical independence is routinely assumed in earthquake forecasting, and the Poisson model is currently the most widely used method for calculating earthquake hazard, so that this comparison is equivalent to the current industry standard [Sholz, 1990; Gross and Rundle, 1998].

I compared the result to one hundred different likelihood ratio tests [Gross and Rundle, 1988; Bevington et al., 1992] on the method using values of  $\Delta P$  obtained from random seismicity catalogs that were used as null hypotheses. The log likelihood is then calculated for the circles,

In addition, I applied the likelihood ratio test to both Figures 5.1d and 5.5. I calculated values of log likelihood of -11.25 for Figure 5.1 and -53.5762 for Figure 5.5. Note that the log likelihood value for this random catalog is approximately the mean of the histogram shown in Figure 5.6. These values correspond to a likelihood ratio of  $e^{42.33} \sim 10^{18}$ , indicating that the colored boxes obtained from the actual instrumental catalog are far more likely to be associated with the locations of the blue circles than the colored boxes obtained from the random

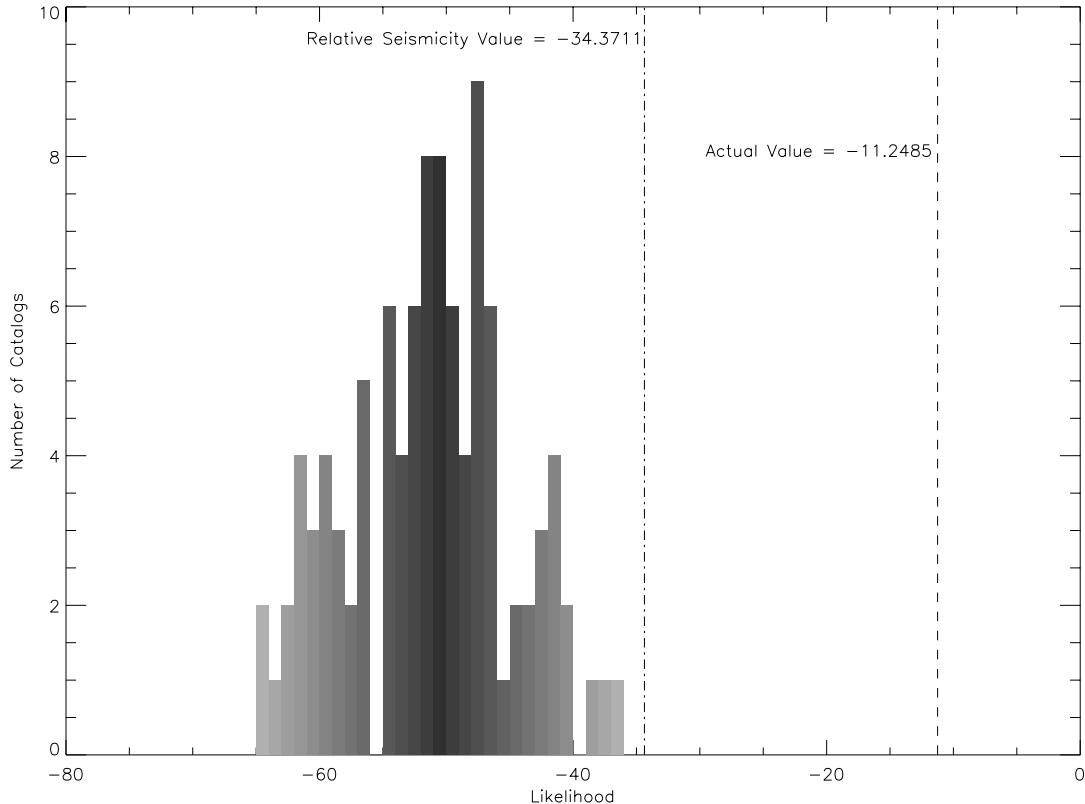


Figure 5.6: Histogram of likelihood values for one hundred random catalogs of southern California seismicity, time period 1991-1978. In addition, the likelihood value for the actual catalog, and the same time period, as shown in Figure 5.2 above, is superimposed on this plot as a dashed line. The likelihood value for the normalized southern California seismicity of 1.1 is plotted as a dash-dotted line.

catalog. The physical reason for this large ratio is that the likelihood test invokes a penalty for colored boxes that are not sufficiently near to blue circles (“false positives”), and there are many more such boxes in Figure 5.5 than in Figure 5.1d.

Finally, I apply the likelihood ratio test to the normalized seismicity for southern California, 1931 through 1991 (Figure 1.1, p. 4). This normalized seismicity plot effectively maps to a plot of seismic hazard for southern California. I calculate a log likelihood value of -34.3711 for Figure 1.1. This log-likelihood value for this hazard map corresponds to a likelihood ratio to the PDPC map of Figure 5.1 of  $e^{23.12} \sim 10^{10}$ . The colored boxes obtained from the PDPC analysis are still far more likely to be associated with the locations of the blue circles than the hazard map probabilities.

From these tests, I conclude that there are coherent space-time correlation structures in the instrumental catalog that the PDPC method identifies.

## 5.6 Discussion

While this PDPC method may identify higher risk areas, there is no certainty at this time that every location of increased  $\Delta P$  will be located near the site of a future large earthquake. There are a number of examples in Figure 5.1 where a box appears without the occurrence of a major earthquake between 1992 and 1999 (“false negative”). There also are a number of examples where a box appears during one time period, then disappears over a longer time period without the occurrence of a major earthquake. One example of this is the colored box which appears near  $34.1^\circ$  latitude,  $241^\circ$  longitude, during the period 1986 to 1991, but disappears during the period 1988 to 1991.

However, the success of this method in forecasting future events is evident visually, and in the statistical tests in Section 5.4. Anomalous regions of increased

$\Delta P$  are associated with future events (see Figure 5.2), and are not associated simply with areas of past seismic activation and aftershocks (Section 5.3.2). This analysis identifies systematic variations in seismicity prior to recent southern California earthquakes and employs data from existing seismic monitoring networks to quantify coherent space-time structures in this seismicity. These space-time patterns in the seismic activity directly reflect the existence of correlated structure in the underlying stress and strain fields, a necessary precondition for the occurrence of large earthquakes. Finally, if the space-time correlations or patterns that lead to a uniformly high stress field on the fault represent emergent space-time structures, these evidently form and evolve over time intervals of years preceding the main shock. Longer time intervals and larger correlated areas should be associated with larger main shocks. Depending on the nature of future seismic activity in the region, as well as ongoing modifications and extensions of the theory and technique, this procedure may prove useful in analysis of future trends in seismic activity.

## Chapter 6

### Summary and Future Work

#### 6.1 Summary

In any attempt to quantify the space-time patterns in southern California seismicity, a necessary first step is the identification of those patterns which are precursory to large events. This leads to the development of a new pattern dynamics approach to the study and forecasting of earthquakes.

Here I apply this pattern dynamics method [Tiampo et al., 1998; Rundle et al., 1999; Tiampo et al., 1999; Rundle et al., 2000 (1)] to the analysis of observed seismicity data from southern California in order to:

- Identify the basis patterns for all possible space-time seismicity configurations using a KL expansion technique. These basis states represent a complete, orthonormal set of eigenvectors and associated eigenvalues that are obtained from the diagonalization of the correlation operators computed for the regional historic seismicity data.
- Provide a new methodology for classifying all of the possible seismicity patterns that can exist in terms of mutually orthogonal eigenstates. In fact, a number of the descriptive patterns cited in Chapter 1 can be readily identified among the eigenstates. For example, Figure 3.11, page 57, can be interpreted either as seismic activation within the Landers epicentral

region, or else as quiescence near Landers coincident with seismic activity surrounding the region, i.e. a Mogi donut.

- Expand on this approach to seismicity as a phase angle and apply the PDPC method to the historic data in order to identify the precursory modes and the cumulative change in probability of an event at a particular location in southern California.
- Support recent work that suggests that correlated structure exists in the underlying stress and strain fields, a necessary precondition for the occurrence of large earthquakes. The success of both the KL decomposition on this data set, and the PDPC method for using anomalous seismicity or precursory quiescence to identify an increased probability of future events, directly reflects the existence of these stress correlations. This supports the interpretation of the earthquake fault system as a critical point system with large correlation lengths.

Although the historic data set is clearly not long enough to accurately identify all the primary modes that characterize seismicity over the last several thousand years within the region, this analysis nonetheless illustrates the effectiveness of these two phase dynamical applications. The application of phase dynamics to other aspects of the seismic system and the phase angle interpretation of its activity is a breakthrough in the analysis of earthquake fault networks.

## 6.2 Future Work

### 6.2.1 KL Decomposition

The application of this technique to a wider variety of solid earth geophysical data holds the potential for a greater understanding of their source mechanisms.

In the area of seismicity, this technique should be applied to larger regions, both in California and in other areas of the world. Specifically, quantification of the appropriate regions sizes should be investigated, as well as the effect of magnitude cutoffs. For example, recent evidence suggests that small events are as important as large events in the redistribution of stress, and that aseismic events also play a large role in that redistribution [Hanks, 1992]. It should be possible to determine, from the seismicity patterns themselves, the importance of each magnitude level in that stress redistribution.

This technique should also be applied to deformation and strain data. For example, the continuous GPS network in southern California, SCIGN, produces time series for the deformation with a reasonable density of sites, which is improving every year. The different KLE modes in that data could be inverted to determine which seismic sources cause them, in much the manner of Savage, 1988.

In addition, signal reconstruction using the KLE modes, in both seismicity and deformation, applied in the same way as the eigenface reconstruction discussed in Chapter 1, could assist in the removal of noise from the data. Systematic quantification of the error in the modes themselves should also be performed, and a measure of the accuracy of their reconstructions developed.

Finally, these results argue strongly for the development of realistic numerical simulations of fault systems such as those in southern California [Rundle, 1988]. Because the historic data set is incomplete worldwide, construction of such numerical simulations is necessary to more accurately define the most significant eigenpatterns, which can be applied to understanding the nature of the observed seismicity. Moreover, such simulations will be of critical importance for relating the observable pattern basis set for the seismicity data to the pattern basis set for stress, strains, and displacements through time. While seismicity is readily observable by standard methods, stress and strain within the earth are not.

However, stress and strain are the primary dynamical variables, and are also the Markov variables in which the underlying nonlinear dynamics are almost certainly formulated. It will be most important to relate a readily observable, seismicity pattern basis set to the actual, unobservable dynamical pattern basis set, so that mode-shaping techniques can be applied to the underlying dynamics [Holmes et al., 1996]. In this manner it may be possible to characterize the spatially coarse-grained features of local and regional stress levels, coefficients of friction, failure and residual stress levels, and fault interactions.

### **6.2.2 PDPC Method**

Initial efforts should be directed at determining both the magnitude and time-to-failure for the events forecasted by the PDPC method. In addition, both an absolute scale for  $\Delta P$  and the error should be quantified in order to develop a better understanding of the forecasting details, and the limits of its application.

In addition, the PDPC method is being applied to other databases from around the world in order to determine to other historical data sets and seismic regimes.

Finally, a standardized computer package is being developed concurrently with the analysis above, in order that independent researchers can apply this method to other data and test its applicability and robustness.

## Bibliography

- [1] Aubrey, D.G. and Emery, K.O., 1983. Eigenanalysis of recent United States sea levels, Continental Shelf Research, **2**, 21-33.
- [2] Aubrey, D.G. and Emery, K.O., 1986. Relative sea levels of Japan from tide gauge records, Geol. Soc. Am. Bull., **97**, 194-205.
- [3] Bak, P. and Tang, C., 1989. Earthquakes as a self-organized critical phenomena. J. Geophys. Res., **94**, 15,635-15637.
- [4] Bak, P., Tang, C., and Weisenfield, K., 1987. Self-organized criticality: An explanation of the  $1/f$  noise. Phys. Rev. Lett., **59**, 381-384.
- [5] Bakun, W.H., King, G.C.P., and Cockerham, R.S., 1986. Seismic slip, aseismic slip, and the mechanics of repeating earthquakes on the Calaveras fault, California. Earthquake Source Mechanics, AGU Monograph, AGU, Washington, D.C., 195-208.
- [6] Bakun, W.H., and McEvilly, T.V., 1984. Recurrence models and Parkfield, California, earthquakes. J. Geophys. Res., **89**, 3051-3058.
- [7] Bailey, R., Dalrymple, G.B. and Lanphere, M.A., 1976. Volcanism, structure, and geochronology of Long Valley Caldera, Mono County, California. J. Geophys. Res., **81**, 725-744.
- [8] Benjamin, J.R. and Cornell, C.A., 1970. Probability, Statistics and Decision for Civil Engineers, McGraw-Hill, N.Y.
- [9] Ben-Zion, Y. and Rice, J.R., 1993. Earthquake failure sequences along a cellular fault zone in a 3D elastic solid containing asperity and non-asperity regions. J. Geophys. Res., **98**, 14,109-14,131.
- [10] Bevington, P.R. and Robinson, D.K., 1992. Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill, Inc., N.Y., 102, 196.

- [11] Bhattacharyya, J., Sheehan, A.F., Tiampo, K.F., and Rundle, J.B., 1998. Using genetic algorithms to model regional waveforms for crustal structure in the western United States. BSSA, **89**, 202-214.
- [12] Billings, S., Kennett, B., and Sambridge, M., 1994. Hypocenter location: genetic algorithms incorporating problem specific information. Geophys. J. Int., **118**, 693-706.
- [13] Boschetti, F., Dentith, M.C., and List, R.D., 1996. Inversion of seismic refraction data using genetic algorithms. Geophysics, **61**, 1715-1727.
- [14] Bowman, D.D., Ouillon, G., Sammis, C.G., Sornette, A., and Sornette, D., 1998. An observational test of the critical earthquake concept. J. Geophys. Res., **103**, 24,359-24,372.
- [15] Brehm, D.J. and Braile, L.W., 1999. Intermediate-term earthquake prediction using the modified time-to-failure method in southern California. BSSA, **89**, 275-293.
- [16] Broomhead, D.S. and King, G.C.P., 1986. Extracting qualitative dynamics from experimental data. Physical D, **20**, 217-236.
- [17] Bufe, C.G. and Varnes, D.J., 1993. Predictive modeling of the seismic cycle of the greater San Francisco bay region. J. Geophys. Res., **98**, 9871-9883.
- [18] Castle, R.O., Estrem, J.E. and Savage, J.C., 1984. Uplift across Long Valley Caldera, California. J. Geophys. Res., **89**, 11,507-11,516.
- [19] Chang, J. and Mak, M., 1993. A dynamical empirical orthogonal function analysis of the intraseasonal disturbances. Jour. Atm. Sc., **50**, no. 4, 613-630.
- [20] Curtis, A., Dost, B., Trampert, J. and Snieder, R., 1995. Shear wave velocity structure beneath Eurasia from surface wave group and phase velocities in an inverse problem reconditioned using the genetic algorithm. Eos Trans. AGU, **76**, 386.
- [21] Dalrymple, G.B., Cox, A., and Doell, R.R., 1965. Potassium-argon age and paleomagnetism of the Bishop Tuff, California. Geol. Soc. Am. Bull., **76**, 665-673.
- [22] Davis, P.M. Surface deformation due to inflation of an arbitrarily oriented triaxial ellipsoidal cavity in an elastic half-space, with reference to Kilauea Volcano, Hawaii. J. Geophys. Res., **91**, 7429-7438.
- [23] Deng, J.S. and Sykes, L.R., 1996. Triggering of 1812 Santa Barbara earthquake by a great San Andreas shock: Implications for future hazards in southern California. Geophys. Res. Lett., **23**, 1155-1158.

- [24] Denlinger, R.P. and Riley, F., 1984. Deformation of Long Valley Caldera, Mono County, California, from 1975-1982. *J. Geophys. Res.*, **89**, 8303-8314.
- [25] Dieterich, J.H., 1994. A constitutive law for rate of earthquake production and an application to earthquake clustering. *J. Geophys. Res.*, **1994**, 2601-2618.
- [26] Dodge, D.A., Beroza, G.C., and Ellsworth, W.L., 1996. Detailed observations of California foreshock sequences: Implications for the earthquake initiation process. *J. Geophys. Res.*, **101**, 22,371-22,392.
- [27] Douma, H., Sneider, R., and Lomax, A., 1996. Ensemble inference in terms of empirical orthogonal functions. *Geophys. J. Int.*, **127**, 363-378.
- [28] Drijkoningen, G.G. and White, R.S., 1995. Seismic velocity structure of oceanic crust by inversion using genetic algorithms. *Geophys. J. Int.*, **123**, 653-664.
- [29] Egolf, D.A., 2000. Equilibrium regained: From nonequilibrium chaos to statistical mechanics. *Science*, **287**, 101-104.
- [30] Elbring, G.J. and Rundle., J.B., 1986. Analysis of borehole seismograms from Long Valley, California: Implications for caldera structure. *J. Geophys. Res.*, **91**, 12,651-12,660.
- [31] Ellsworth, W.I., and A. T. Cole, A.T., 1997. *Seis. Res. Lett.*, **68**, 298.
- [32] Eneva, M. and Ben-Zion, Y., 1997. Techniques and parameters to analyze seismicity patterns associated with large earthquakes. *J. Geophys. Res.*, **102**, 17,785-17,795.
- [33] Eshelby, J.D., 1957. The determination of the elastic field of an ellipsoidal inclusion and related problems. *Proc. R. Soc. London*, Ser. A, **241**, 376-396.
- [34] Ferguson, C., Klein, W., and Rundle, J.B., 1999. Spinodals, scaling and ergodicity in a model of an earthquake fault with long-range stress transfer. *Phys. Rev. E.*, **60**, 1359-1373.
- [35] Fisher, D.S., Dahmen, K., Ramanathan, S., and Ben-Zion, Y., 1997. Statistics of earthquakes in simple models of heterogeneous faults. *Phys. Rev. Lett.*, **78**, 4885-4888.
- [36] Frohlich, C., 1987. Aftershocks and temporal clustering of deep earthquakes. *J. Geophys. Res.*, **92**, 13,944.
- [37] Fukunaga, K., 1970. Introduction to Statistical Pattern Recognition, Academic Press, N.Y.

- [38] Fung, Y.C., 1965. Foundations of solid mechanics. Prentice-Hall, Englewood Cliffs, NJ.
- [39] Garcia, A. and Penland, C., 1991. Fluctuating hydrodynamics and principal oscillation pattern analysis. J. Stat. Phys., **64**, 1121-1132.
- [40] Geller, R.J., Jackson, D.D., Kagan, Y.Y., and Mulargia, F., 1997. Enhanced: Earthquakes cannot be predicted. Science, **275**, 1616-1620.
- [41] Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison Wesley, Reading, MA.
- [42] Gomberg, J., 1996. Stress/strain changes and triggered seismicity following the  $M_w$  7.3 Landers, California, earthquake. J. Geophys. Res., **101**, 751-764.
- [43] Goswami, B.N., Rajendran, K., and Sengupta, D., 1997. Source of seasonality and scale dependence of predictability in a coupled ocean-atmosphere model. Monthly Weather Review, **125**, 846-858.
- [44] Grant, L. and Sieh, K., 1994. Paleoseismic evidence of clustered earthquakes on the San Andreas fault in the Carrizo Plain. J. Geophys. Res., **99**, 6819-6842.
- [45] Gray, C.M., 1997. Synchronous oscillations in neuronal systems: Mechanisms and functions. Pattern Formation in the Physical and Biological Sciences, Lecture Notes V, SFI, Addison-Wesley, Reading, MA, 93-134.
- [46] Gross, S.J. and Kisslinger, C., 1994. Tests of models of aftershock rate decay. BSSA, **84**, 1571-1579.
- [47] Gross, S. and Rundle, J.B., 1998. A systematic test of time-to-failure analysis. Geophys. J. Int., **133**, 57-64.
- [48] Haberman, R.E., 1981. Precursory seismicity patterns: stalking the mature seismic gap. Earthquake Prediction: An International Review, AGU Monograph, AGU, Washington, D. C. 29-42.
- [49] Hanks, T.C., 1992. Small earthquakes, tectonic forces. Science, **256**, 1430-1432.
- [50] Harris, R.A. and Simpson, R.W., 1992. Changes in static stress on Southern California faults after the 1992 Landers earthquake. Nature, **360**, 251-254.
- [51] Hasselman, K., 1988. PIPs and POPS: The reduction of complex dynamical systems using principal interaction and oscillation patterns. J. Geophys. Res., **93**, 11,015-11,021.

- [52] Hendricks, J.R., Leben, R.R., Born, G.H., and Koblinsky, C.J., 1996. Empirical orthogonal function analysis of global TOPEX/POSEIDON altimeter data and implications for global sea level rise. *J. Geophys. Res.*, **101**, 14,131-14,145.
- [53] Hertz, J., Korgh, A., and Palmer, R.G., 1990. Introduction to the Theory of Neural Computation, Lecture Notes I, SFI, Addison-Wesley, Reading, MA.
- [54] Herz, A.V., and Hopfield, J.J., 1995. Earthquake cycles and neural reverberations: Collective oscillations in systems with pulse-coupled threshold elements. *Phys. Rev. Lett.*, **75**, 1222-1225.
- [55] Hill, D.P., Bailey, R.A. and Ryall, S.A., 1985. Active tectonic and magmatic processes beneath Long Valley Caldera, Eastern California: An overview. *J. Geophys. Res.*, **90**, 11,111-11,120.
- [56] Hill, D., Eaton, J.P., and Jones, L.M., 1990. Seismicity, 1980-86. The San Andreas Fault System, California, USGS Prof. Paper 1515, U.S. GPO, Washington, D.C., 115-152.
- [57] Hill, D.P., Johnston, M.J.S., Langbein, J.O., and Bilham, R., 1995. Response of Long Valley caldera to the  $M_w = 7.3$  Landers, California, earthquake. *J. Geophys. Res.*, **100**, 12,985-13,005.
- [58] Hill, D.P., Reasenberg, P.A., Michael, A., Arabaz, W.J., Beroza, G., Brumbaugh, D., Brune, J.N., Castro, R., Davis, S., Depolo, D., Ellsworth, W.L., Gomberg, J., Harmsen, S., House, L., Jackson, S.M., Johnston, M.J.S., Jones, L., Keller, R., Malone, S., Munguia, L., Nava, S., Pechmann, J.C., Sanford, A., Simpson, R.W., Smith, R.B., Stark, M., Stickney, M., Vidal, A., Walter, S., Wong, V., Zollweg, J., 1993. Seismicity remotely triggered by the Magnitude 7.3 Landers, California, earthquake. *Science*, **260**, 1617-1623.
- [59] Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. MIT Press, Cambridge, MA.
- [60] Holmes, P., Lumley, J.L., and Berkooz, G., 1996. Turbulence, Coherent Structures, Dynamical Systems and Symmetry, Cambridge University Press, Cambridge, U.K.
- [61] Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *Jour. Educ. Psych.*, **24**, 417-520.
- [62] House, L.S., Sykes, L.R., Davies, J.N., and Jacob, K.H., 1981. Identification of a possible seismic gap near Unalaska Island, eastern Aleutians, Alaska. Earthquake Prediction: An International Review, AGU Monograph, AGU, Washington, D. C. 81-92.

- [63] Huang, Y., Saleur, H., Sammis, C., and Sornette, D., 1998. Precursors, aftershocks, criticality and self-organized criticality, *Europhys. Lett.*, **41**, 43-49.
- [64] Jaume, S.C. and Sykes, L.R., 1999. Evolving towards a critical point: a review of accelerating seismic moment/energy release prior to large and great earthquakes. *Pure Appl. Geoph.*, **155**, 279-306.
- [65] Jin, S. and Madariaga, R., 1993. Background velocity inversion with a genetic algorithm. *Geophys. Res. Lett.*, **20**, 93-96.
- [66] Jones, L.M. and Hauksson, E., 1997. The seismic cycle in southern California: Precursor or response? *Geophys. Res. Lett.*, **24**, no. 4, 469-472.
- [67] Kagan, Y.Y. and Jackson, D.D., 1992. Seismic gap hypothesis, ten years after. *J. Geophys. Res.*, **96**, 21,419-21,431.
- [68] Kanamori, H., 1981. The nature of seismicity patterns before large earthquakes. *Earthquake Prediction: An International Review*, AGU Monograph, AGU, Washington, D. C., 1-19.
- [69] Kato, N., Ohtake, M., and Hirasawa, T., 1997. *Pure Appl. Geoph.*, **150**, 249.
- [70] Kerr, R.A., 1998. A slow start for earthquakes. *Science*, **279**, 985.
- [71] King, G.C.P., Stein, R.S., and Lin, J., 1994. Static stress changes and the triggering of earthquakes. *BSSA*, **84**, 935-953.
- [72] Klein, W., Ferguson, C., and Rundle, J.B., 1996. Spinodals and scaling in slider-block models. *Reduction and Predictability of Natural Disasters*, SFI series in the science of complexity, XXV, Addison-Wesley, Reading, MA, 223-242.
- [73] Klein, W., Rundle, J.B., and Ferguson, C., 1997. Scaling and nucleation in models of earthquake faults. *Phys. Rev. Lett.*, **78**, 3793-3796.
- [74] Kossobokov, V. and Carlson, J.M., 1995. Active zone size versus activity - study of different seismicity patterns in the context of the prediction algorithm M8. *J. Geophys. Res.*, **100**, 6431-6441.
- [75] Langbein, J., 1989. Deformation of the Long Valley Caldera, Eastern California from mid-1983 to mid-1988: Measurements using a two-color geodimeter. *J. Geophys. Res.*, **94**, 3833-3849.
- [76] Langbein, J., Linker, M. and Tupper, D., 1987. Analysis of two-color geodimeter measurements of deformation within the Long Valley Caldera: June 1983 to October 1985. *J. Geophys. Res.*, **91**, 9423-9442.

- [77] Langbein, J., Hill, D.P., Parker, T.N. and Wilkinson, S.K., 1993. An episode of reinflation of the Long Valley Caldera, Eastern California: 1989-1991. *J. Geophys. Res.*, **98**, 15,851-15,870.
- [78] Langbein, J., Dzurisin, D., Marshall, G., Stein, R. and Rundle, J., 1995. Shallow and peripheral volcanic sources of inflation revealed by modeling two-color geodimeter and leveling data from Long Valley Caldera, California, 1988-1992. *J. Geophys. Res.*, **100**, 12,487-12,495.
- [79] Lee, T. and Cornillon, P., 1995. Temporal variation of meandering intensity and domain-wide lateral oscillations of the Gulf Stream. *J. Geophys. Res.*, **100**, 13,603-13,613.
- [80] Linde A.T., Gladwin, M.T., Johnston, M.J.S., Gwyther, R.L., Bilham, R.G., 1996. A slow earthquake sequence on the San Andreas fault. *Nature*, **383**, 65-68.
- [81] Lomax, A. and Snieder, R., 1995. The contrast in upper mantle shear-wave velocity between the East European Platform and tectonic Europe obtained with genetic algorithm inversion of Rayleigh-wave group dispersion. *Geophys. J. Int.*, **123**, 169-182.
- [82] Lyzenga, G.A., Raefsky, A., and Mulligan, S.G., 1991. Models of recurrent strike-slip earthquake cycles and the state of crustal stress. *J. Geophys. Res.*, **96**, 21,623-21,640.
- [83] Marone, C., Vidale, J.E., and Ellsworth, W.L., 1995. Fault healing inferred from time-dependent variations in source properties of repeating earthquakes. *Geophys. Res. Lett.*, **22**, 3095-3098.
- [84] McGuire, J.J., Imhle, P.F., and Jordan, T.H., 1996. Time-domain observations of a slow precursor to the 1994 Romanche transform earthquake. *Science*, **274**, 82-85.
- [85] McTigue, D.F., 1987. Elastic stress and deformation near a finite spherical magma body: Resolution of the point source paradox. *J. Geophys. Res.*, **92**, 12,931-12,940.
- [86] Michalewicz, Z., 1992. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, New York, NY.
- [87] Mindlin, R.D., 1936. Force at a point in the interior of a semi-infinite solid. *Physics*, **7**, 195-202.
- [88] Mitchell, M., 1992. The royal road for genetic algorithms: Fitness landscapes and GA performance. *Proc. First Int. Conf. Artificial Life*. Paris, France, 245-254.

- [89] 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.
- [90] Mogi, K., 1958. Relations between the eruptions of various volcanos and the deformations of the ground surfaces around them. Bull. Earth. Res. Inst. Tokyo, **36**, 99-134.
- [91] Mogi, K., 1969. Some features of recent seismic activity in and near Japan, (2) Activity before and after great earthquakes. Bull. Earthquake Res. Inst., Tokyo Univ., **47**, 395-417.
- [92] Mogi, K., 1977. Seismic activity and earthquake predictions. Proc. Symp. on Earthquake Prediction, Seis. Soc. Japan, 203-214.
- [93] Mori, H. and Kuramoto, Y., 1998. Dissipative Structures and Chaos, Springer-Verlag, Berlin.
- [94] Nanjo, K., Nagahama, H., and Satomura, M., 1998. Rates of aftershock decay and the fractal structure of active fault systems. Tectonophysics, **287**, 173-186.
- [95] Navarra, A., 1993. A new set of orthonormal modes for linearized meteorological problems. Jour. Atm. Sc., **50**, no. 16, 2569-2582.
- [96] Neves, F.A., Singh, S.C. and Priestley, K.F., 1996. Velocity structure of upper-mantle transition zones beneath central Eurasia from seismic inversion using genetic algorithms. Geophysics, **22**, 523-552.
- [97] Nijhout, H.F., 1997. Pattern formation in biological systems. Pattern Formation in the Physical and Biological Sciences, Lecture Notes V, SFI, Addison Wesley, Reading, MA, 269-298.
- [98] North, G.R., 1984. Empirical orthogonal functions and normal modes. Jour. Atm. Sc., **41**, no. 5, 879-887.
- [99] Ouchi, T., 1993. Population dynamics of earthquakes and mathematical modeling. Pure Appl. Geophys., **140**, 15-28.
- [100] Pacheco, J.F., Scholz, C.H., and Sykes, L.R., 1992. Changes in frequency-size relationship from small to large earthquakes. Nature, **355**, 71-73.
- [101] Penland, C., 1989. Random forcing and forecasting using principal oscillation pattern analysis. Monthly Weather Review, **117**, 2165-2185.
- [102] Penland, C. and Magorian, T., 1993. Prediction of Nino 3 sea surface temperatures using linear inverse modeling. J. Climate, **6**, 1067-1076.

- [103] Penland, C. and Sardeshmukh, P.D., 1995. The optimal growth of tropical sea surface temperature anomalies. *J. Climate*, **8**, 1999-2024.
- [104] Pollitz, F.F., and Sacks, I.S., 1997. The 1995 Kobe, Japan, earthquake: A long-delayed aftershock of the offshore 1944 Tonankai and 1946 Nankaido earthquakes. *BSSA*, **87**, 1-10.
- [105] Preisendorfer, R.W., 1988. Principle Component Analysis in Meteorology and Oceanography, Elsevier, Amsterdam.
- [106] Press, F. and Allen, C., 1995. Patterns of seismic release in the southern California region. *J. Geophys. Res.*, **100**, 6421-6430.
- [107] Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T., 1992. Numerical Recipes in C, 2nd Edition, Cambridge University Press, Cambridge, U.K.
- [108] Richter, C.F., 1958. Elementary Seismology, Freeman, San Francisco, California.
- [109] Romanowicz, B. and Rundle, J.B., 1993. On scaling relations for large earthquakes. *BSSA*, **83**, 1294-1297.
- [110] Rundle, J.B., 1982. Deformation, gravity and potential changes due to volcanic loading of the crust. *J. Geophys. Res.*, **87**, 10,729-10,744.
- [111] Rundle, J.B., 1988. A physical model for earthquakes: 2. Applications to southern California. *J. Geophys. Res.*, **93**, 6255-6274.
- [112] Rundle, J.B., 1989. Derivation of the complete Gutenberg-Richter magnitude-frequency relation using the principle of scale invariance. *J. Geophys. Res.*, **94**, 12,337-12,342.
- [113] Rundle, J.B., 1993. Magnitude-frequency relation for earthquakes using a statistical mechanical approach. *J. Geophys. Res.*, **98**, 21,943-21,950.
- [114] Rundle, J.B., Gross, S., Klein, W., Ferguson, C., and Turcotte, D.L., 1997. The statistical mechanics of earthquakes. *Tectonophysics*, **277**, 147-164.
- [115] Rundle, J.B. and Hill, D.P., 1988. The geophysics of a restless caldera - Long Valley, California. *Ann. Rev. Earth Plan. Sci.*, **16**, 251-271.
- [116] Rundle, J.B., and Klein, W., 1995. New ideas about the physics of earthquakes. *Rev. Geophys. Space Phys. Suppl. (July)* **283**, 283-286.
- [117] Rundle, J.B., Klein, W., and Gross, S., 1996. Rupture characteristics, recurrence, and predictability in a slider-block model for earthquakes. Reduction and Predictability of Natural Disasters, SFI series in the science of complexity, XXV, Addison-Wesley, Reading, MA, 167-203.

- [118] Rundle, J.B., Klein, W., and Gross, S., 1999. Physical basis for statistical patterns in complex earthquake populations: Models, predictions and tests, Pure Appl. Geoph., **155**, 575-607.
- [119] Rundle, JB, Klein, W., Tiampo, K., and Gross, S., 2000 (1). Linear pattern dynamics in nonlinear threshold systems, Phys. Rev. E., **61**, 2418-2432.
- [120] Rundle, J.B., Klein, W., Tiampo, K.,and Gross, S, to be published June 2000 (2). Dynamics of seismicity patterns in systems of earthquake faults, The Physics of Earthquakes, AGU Monograph, AGU, Washington, D.C.
- [121] Rundle, J.B., Tiampo, K., and Klein, W., 1998. Dynamics of seismicity patterns in complex earthquake fault systems. Eos Trans. AGU, **79**, 635.
- [122] Rundle, J.B., Tiampo, K., and Klein, W., 1999. Non-local space-time patterns in seismicity, with application to Bufe-Varnes scaling. Eos Trans. AGU, **80**, 924.
- [123] Ryall, F. and Ryall, A., 1981. Attenuation of P and S waves in a magma chamber in Long Valley Caldera, California. Geophys. Res. Lett., **8**, no. 6, 557-560.
- [124] Saleur, H., Sammis, C.G., and Sornette, D., 1995. Discrete scale invariance, complex fractal dimensions, and log-periodic fluctuations in seismicity. J. Geophys. Res., **17**,661-17,677.
- [125] Sammis, C.G., Sornette, D., and Saleur, H., 1996. Complexity and earthquake forecasting. Reduction and Predictability of Natural Disasters, SFI series in the science of complexity, XXV, Addison-Wesley, Reading, MA, 143-156.
- [126] Sanders, C.O., 1984. Location and configuration of magma bodies beneath Long Valley, California, determined from anomalous earthquake signals. J. Geophys. Res., **89**, 8287-8302.
- [127] Sanders, C.O. and Ryall, F., 1983. Geometry of magma bodies beneath Long Valley, California determined from anomalous earthquake signals. Geophys. Res. Lett., **10**, no. 8, 690-692.
- [128] Sanders, C.O., Ponko, S.C., Nixon, L.D. and Schwartz, E.A., 1995. Seismological evidence for magmatic and hydrothermal structure in Long Valley Caldera from local earthquake attenuation and velocity tomography. J. Geophys. Res., **100**, 8311-8326.
- [129] Savage, J.C., 1988. Principal component analysis of geodetically measured deformation in Long Valley caldera, eastern California, 1983-1987, J. Geophys. Res., **93**, 13,297-13,305.

- [130] Savage, J.C., 1993. The Parkfield prediction fallacy. BSSA, **83**, 1-6.
- [131] Scholz, C.H., 1990. The Mechanics of Earthquakes and Faulting, Cambridge University Press, Cambridge, U.K.
- [132] Sen, M.K. and Stoffa, P.L., 1992. Rapid sampling of model space using genetic algorithms: Examples from seismic waveform inversion. Geophys. J. Int., **108**, 281-292.
- [133] Shatz, C.J., 1997. Impulse activity and the patterning of connections during CNS development. Pattern Formation in the Physical and Biological Sciences, Lecture Notes V, SFI, Addison-Wesley, Reading, MA, 299-322.
- [134] Sieh, K., Stuiver, M. and Brillinger, D., 1989. A more precise chronology of earthquakes produced by the San Andreas fault in southern California. J. Geophys. Res., **94**, 603-623.
- [135] Stark, M.A. and Davis, S.D., 1996. Remotely triggered microearthquakes at The Geysers geothermal field, California. Geophys. Res. Lett., **23**, n. 9, 945-948.
- [136] Stein, R.S., 1999. The role of stress transfer in earthquake occurrence. Nature, **402**, 605-609.
- [137] Stein, R.S., King, G.C.P., and Lin, J., 1992. Change in failure stress on the southern San Andreas fault system caused by the 1992 Magnitude = 7.4 Landers earthquake. Science, **258**, 1328-1331.
- [138] Stoffa, P.L. and Sen, M.K., 1991. Nonlinear multiparameter optimization using genetic algorithms: Inversion of plane-wave seismograms. Geophysics, **56**, 1794-1810.
- [139] Swan, F.H., Schwartz, D.P., Cluff, L.S., 1980. Recurrence of moderate to large magnitude earthquakes produced by surface faulting on the Wasatch fault zone, Utah. BSSA, **70**, 1431-1462.
- [140] Sykes, L.R., Shaw, B.E., Scholz, C.H., 1999. Rethinking earthquake prediction. Pure Appl. Geoph., **155**, 207-232.
- [141] Taylor, J.R., 1982. An Introduction to Error Analysis. University Science Books, USA, 225-230.
- [142] Tiampo, K.F., Rundle, J.B., Fernandez, J.M., and Langbein, J.O., 2000. Spherical and ellipsoidal volcanic sources at Long Valley caldera, California, using a genetic algorithm inversion technique. J. Vol. Geophys. Res., accepted for publication, October 2000.

- [143] Tiampo, K.F., Rundle, J.B., Gross, S., and Klein, W. Dynamics of seismicity patterns in complex fault systems. Eos, Transactions, AGU, **79**, n. 45, p. 635, 1998.
- [144] Tiampo, K.F., Rundle, J.B., Gross, S.J., and Klein, W., 1998. Dynamics of seismicity patterns in southern California. Eos Trans. AGU, **79**, 644.
- [145] Tiampo, K.F., Rundle, J.B., Klein, W., and Gross, S.J., 1999. Systematic evolution of nonlocal space-time earthquake patterns in southern California. Eos Trans. AGU, **80**, 1013.
- [146] Tiampo, K.F., Rundle, J.B., Klein, W., McGinnis, S., and Gross, S.J., to be published June 2000. Observation of systematic variations in non-local seismicity patterns from southern California. The Physics of Earthquakes, AGU Monograph, Washington, D.C.
- [147] Turcotte, D.L., 1991. Earthquake prediction. Ann. Rev. Earth Planet. Sci., **19**, 263-281.
- [148] Turcotte, D.L., 1997. Fractals and chaos in geology and geophysics, 2nd editiion, Cambridge University Press, Cambridge, U.K.
- [149] Urbach, J.S., Madison, R.C., and Markert, J.T., 1995. Interface depinning, self-organized criticality, and the Barkhausen effect. Phys. Rev. Lett., **75**, 276-279.
- [150] Vautard, R. and Ghil, M., 1989. Singular spectrum analysis in nonlinear dynamics, with applications to paleodynamic time series. Physical D, **35**, 395-424.
- [151] Vicsek, T. Fractal Growth Phenomena, World Scientific, Singapore, 19-25.
- [152] Winchester, J.P., Creager, K.C. and McSweeney, T.J., 1993. Better alignment through better breeding: Phase alignment using genetic algorithms and cross-correlation techniques. Eos Trans. AGU, **74**, 394.
- [153] Wright, A.H., 1991. Genetic algorithms for real parameter optimization. In: Rawlins (Editor), Foundations of Genetic Algorithms. Morgan Kaufmann Publishers, San Mateo, CA.
- [154] Wyss, M., Aceves, R.L., Park, S.K., Geller, R.J., Jackson, D.D., Kagan, Y.Y., and Mulargia, F., 1997. Science, **278**, 487-490.
- [155] Wyss, M., Schorlemmer D., Wiemer, S., 2000. Mapping asperities by minima of local recurrence time: San Jacinto-Elsinore fault zones. J. Geophys. Res., **105**, 7829-7844.

- [156] Wyss, M., Shimaziki, K., and Urabe, T., 1996. Quantitative mapping of a precursory seismic quiescence to the Izu-Oshima 1990 (M6.5) earthquake, Japan. Geophys. Jour. Int., **127**, 735-743.
- [157] Wyss, M. and Wiemer, S., 1999. How can one test the seismic gap hypothesis? The case of repeated ruptures in the Aleutians. Pure Appl. Geoph., **155**, 259-278.
- [158] Yamashita, T. and Knopoff, L., 1989. A model of foreshock occurrence. Geophys. Jour., **96**, 389-399.
- [159] Yang, X.M., Davis, P.M. and Dieterich, J.H., 1988. Deformation from inflation of a dipping finite prolate spheroid in an elastic half-space as a model for volcanic stressing. J. Geophys. Res., **93**, 4249-4257.
- [160] Yeomans, J.M., 1992. Statistical Mechanics of Phase Transitions, Clarendon Press, Oxford, U.K.
- [161] Yu, T.T. and Rundle, J.B., 1995. Inverting for fault zone geometry using genetic algorithms. IUGG Gen. Assembly, **21**, 56.
- [162] Yu, T.T., Fernandez, J., and Rundle, J.B., 1998. Inverting the parameters of an earthquake-ruptured fault with a genetic algorithm. Comp. and Geo., **24**, 173-182.
- [163] Zhou, R., Tajima, F., and Stoffa, P.L., 1995. Application of genetic algorithms to constrain near-source velocity structure for the 1989 Sichuan earthquakes. BSSA, **85**, 590-605.

## Appendix A

### Genetic Algorithm Applications to Geophysical Inverse Problems

#### A.1 Introduction

In the late 1970s, the Long Valley caldera in eastern California began a period of activity unlike any previously recorded, which included two periods of reinflation of the central resurgent dome accompanied by seismicity within and around the caldera itself that continued into the 1990s. The second inflation episode, occurring after several years of relative quiescence, began in 1989, produced a maximum uplift of the resurgent dome of almost 8 cm by 1992, and did not die away to again reach background levels until 1995 [Denlinger and Riley, 1984; Langbein et al., 1987; Rundle and Hill, 1988]. This second inflation period was modeled by Langbein et al. [1995], as two ellipsoidal sources, a primary source located at 5.5 km in depth below the resurgent dome and a secondary source at between 10 and 20 km in depth below the south moat, along with adjacent diking and normal faulting in the northwest of the caldera.

I employ a genetic algorithm (GA) inversion technique to study the possible spherical or ellipsoidal inflation sources in Long Valley caldera, California. I first derive the relationship for the change in volume of a finite ellipsoid in terms of its internal pressure. I then performed extensive synthetic testing on data similar in nature and extent to that used by Langbein in 1995 in order to attempt to quantify the possible error on the GA solutions. Finally, I invert for both a two-

sphere model and a two-ellipsoid model, with and without normal faulting in the northern caldera.

## A.2 Long Valley

### A.2.1 Geology

The Long Valley caldera lies along the eastern Sierra Nevada escarpment, on the western edge of the Basin and Range province, northwest of the town of Bishop, California, and 30 km south of Mono Lake. Elliptical in shape, it extends 32 km in the east-west direction and 17 km north to south, with an average elevation of 2200 m (see Figure A.1). Prior to its formation, volcanism began in the Long Valley area with mafic and andesitic eruptions approximately 3.2 Mbp, continuing through the most recent uplift of the Sierra Nevadas . The caldera itself was formed 0.7 Mbp as a result of the Bishop Tuff eruption. Approximately 600 km<sup>3</sup> of rhyolitic magma was ejected as ash flows from the magma chamber below, which collapsed as a result of the sudden evacuation, forming a depression over two kilometers deep. Rhyolitic volcanism has since been confined within the caldera, including the formation of a 500 m high resurgent dome in the western portion of the caldera within the first 100,000 years following the collapse [Bailey et al., 1976].

Situated as it is along the boundary of the Sierra Nevada mountains and Basin and Range provinces, extensional tectonics play a predominant role in the geologic activity in the Long Valley region. While crustal extension results in normal faulting at the surface, basaltic magmatic intrusion occurs in the lower crust. Volcanic centers are created where the lithosphere is pulled apart. This basal heat source below the Long Valley region is defined by a low-velocity zone in the midcrust. Parallel to the range front, as defined by the regional stress field,

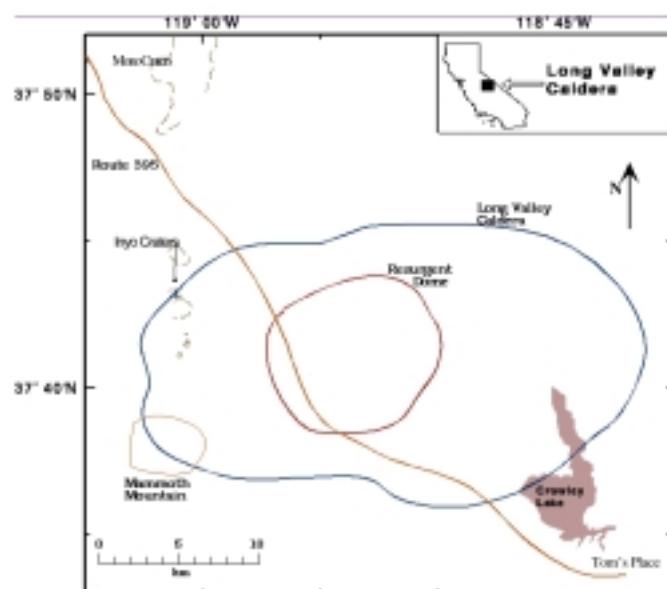


Figure A.1: Long Valley caldera, eastern California.

it replenishes the silicic magma chamber via crustal melting, acting as a common root to the separate magma chambers beneath the caldera, the Inyo Domes and Craters system, and Mono Lake [Hill et al., 1985].

Seismic investigations have delineated several anomalous areas beneath the caldera. In particular, the earthquake sequence beginning in 1980 provided a large number of events for analysis. Ryall and Ryall [1981] confirmed that the presence of a magma chamber in the south central caldera at depths below 7 or 8 kms. Others have imaged additional bodies in either the south, southeast or northwest portions of the caldera at depths as shallow as 4 km, although it is unclear whether some of these are primarily hydrothermal in nature [Sanders and Ryall, 1983; Sanders, 1984; Elbring and Rundle, 1986; Sanders et al., 1995]. Although an eruption on the order of the Bishop Tuff event of  $600 \text{ km}^3$  is unlikely, there is certainly the potential for an event on the order of those that occurred at the Inyo Craters 500 years ago - 0.2 to  $2 \text{ km}^3$  [Hill et al., 1985].

### A.2.2 Deformation

In October of 1978, seismicity resumed in the Long Valley region with the occurrence of an  $M_L = 5.7$  event near the Wheeler Fault south of the caldera. Migrating northward over the next few years, the main activity concentrated in the south moat in May of 1980 with four earthquakes of  $M \geq 6$  and their subsequent aftershocks. In response to this seismicity, the California Division of Highways leveled Highway 395 in 1980, which runs from Toms Place in the south to Lee Vining north of the caldera, and discovered 200-250 mm of uplift had occurred since 1975 [Denlinger and Riley, 1984]. An extensive resurvey was carried out in 1982 and again in 1983, and demonstrated that uplift began sometime after 1975, probably in 1978 or 1979, and continued at an approximately constant rate through 1983 [Castle et al., 1984]. In 1982, a 21-line deformation network

was installed, to be augmented over ensuing years, and was surveyed with an electronic distance meter (EDM) yearly between 1982 and 1986. In addition, a two-color EDM network was installed in 1983. Line length changes between 1983 and 1986 show initial extension as high as 6 ppm/yr in 1983 on lines crossing the resurgent dome [Langbein et al., 1987]. By 1988, the original 11 baselines had been increased to 42 (Figure A.2), and the deformation rate had decreased to background on many of these baselines, with a maximum of only 1 ppm/yr [Langbein, 1989]. By mid 1989, strain rates had decreased to such an extent that several baselines were contracting slightly.

In October of 1989, seismicity increased again in conjunction with seismic activity under Mammoth Mountain, as extension rates across the caldera increased sharply to almost 9 mm/yr, followed by renewed activity under the south moat. The largest event of  $M \cong 4$  occurred in the south moat, where the seismicity was predominantly strike-slip with a mixture of normal events. Following the initial deformation, extension rates slowed to between 3 and 5 ppm/yr through 1991. The extensive monitoring of the area permitted, for the first time, the identification of the temporal sequence of activity. The increase in deformation preceded the south moat seismic activity by 2 to 3 months, and the cumulative moment release from the deformation exceeded that from the seismicity by a factor of 4, indicating that magmatic processes triggered crustal failure, rather than inflation occurring as a result of a regional strain event [Langbein et al., 1993]. Total uplift for the episode beginning in 1989 had reached 8 cm along the Rte. 395 leveling line, and 11 cm along the Rte 203 leveling traverse near Casa Diablo by 1992, before it decreased again to the background levels of 1988 [Langbein et al., 1995].

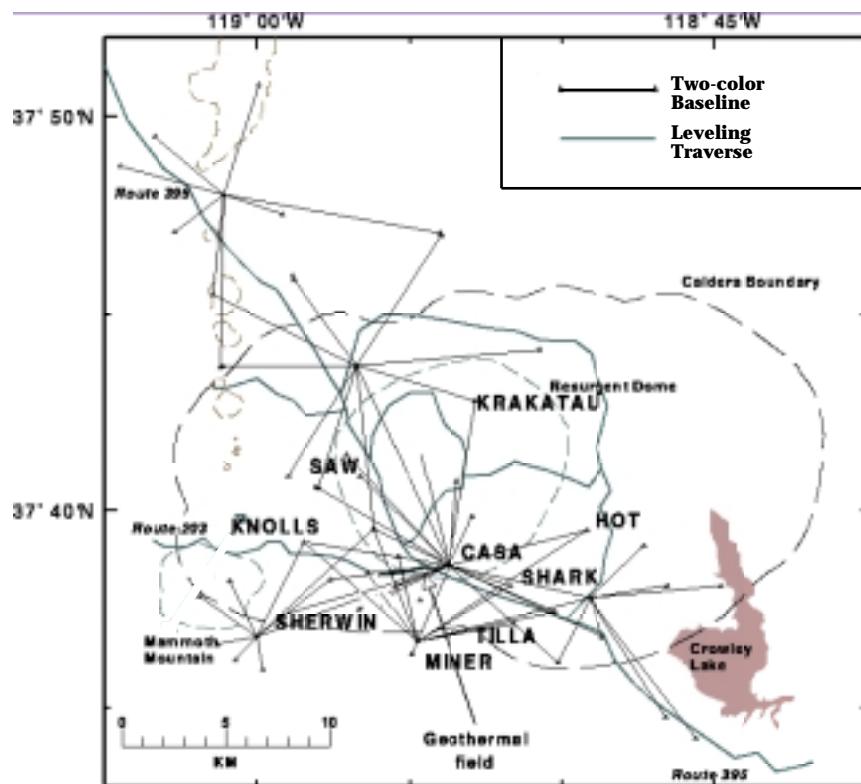


Figure A.2: Two-color EDM & leveling lines at Long Valley caldera (adapted from LVO web site, <http://quake.wr.usgs.gov/QUAKES/geodetic/twocolor/longvalley.html>).

### A.2.3 Modeling

Attempts to model the sources of the deformation in the Long Valley caldera began almost as soon as the activity was detected. Considerable disagreement on these mechanisms existed throughout the early 1980s, which is not surprising when one considers that substantial deformation occurred prior to the initial surveying in 1982, and sufficient temporal or spatial monitoring for accurate modeling did not exist until the mid-1980s. Proposed models have included spherical and ellipsoidal magma chambers, in various quantities and locations, diking in both the south moat and under Mammoth Mountain, and displacement on various fault planes. Additional factors to be considered are the geothermal power plant located 1 km west of Casa Diablo, which went on line in the spring of 1985, and seasonal snow melt on the steep terrain of portions of the caldera.

The large variety in source models for the deformation at Long Valley are a result of the spatial and temporal variation in both the activity below the caldera and the monitoring programs in the area. The second deformation phase, which began in 1989 and was the best well-monitored in both space and time, is the sequence which modeled in this paper. This period of activity initially was modeled with one spherical point source with a volume of  $0.025 \text{ km}^3$  beneath the center of the resurgent dome [Langbein et al., 1993]. More recently, two ellipsoidal point sources, one at 5.5 km beneath the resurgent dome and one between 10 and 20 km deep, in conjunction with normal faulting in the northwest caldera and dike intrusion under Mammoth Mountain, were used to model the deformation between 1989 and 1992 [Langbein et al., 1995]. I have, as a result, introduced both spheroidal and ellipsoidal source models, at various depths and sizes, into this study in an attempt to better understand their features and limitations in a complex geologic area such as Long Valley, California.

#### A.2.4 Data Sources

Both leveling data and two-color EDM data were inverted for this second inflation period. The two-color EDM records measurements with a standard error of  $\sigma' = a^2 + b^2 l^2$ , where  $a = 0.3$  mm,  $b = 0.12$  ppm and  $l$  is the baseline length. Precisions for the entire network (Figure A.2) therefore range from 0.3 to 2.0 mm [Langbein et al., 1993; Langbein et al., 1995].

The leveling network has a standard error,  $\sigma$ , where  $\sigma = \alpha l^{1/2}$ , and  $\alpha = 2.8$  mm/km<sup>1/2</sup>, and  $l$  is again the line length. The entire network contains a total of 105 leveling locations, with a standard error between 2.0 and 4.5 mm [Langbein et al., 1995].

The geothermal power plant near CASA produced subsidence due to ground-water extraction, contaminating adjacent lines. These were removed from the data model, leaving only 25 high quality expansion lines for the inversion.

Both these data arrays were surveyed in 1988 and again in 1992. The synthetic data produced for this model therefore consists of 25 expansion lines and 105 leveling points, with standard errors of 5 and 10 mm., respectively, in an attempt to perform a conservative error analysis on the synthetically modeled data. These form an areal pattern identical in size and extent to that existing in the Long Valley caldera during the period 1988 to 1992.

### A.3 Inversion Methodology

#### A.3.1 The Genetic Algorithm

Many geophysical optimization problems are nonlinear and result in irregular objective functions. Moreover, they often also have a fitness landscape with several local minima. Consequently, local optimization techniques, e.g., linearized matrix inversion, steepest descent, conjugate gradients, etc. can converge pre-

maturely to a local minimum. In addition, the success in obtaining an optimum solution can depend strongly on the choice of the starting parameters. Nonlinear global optimization techniques recently have been incorporated into geophysical inversion problems such as seismological waveform fitting in order to avoid or mitigate these problems.

In general, geophysical inverse problems involve employing large quantities of measured data, in conjunction with an efficient computational algorithm to explore the model space to find the global minimum associated with the optimal model parameters. In a genetic algorithm (GA), the parameters to be inverted for are coded as genes, and a large population of potential solutions for these genes are searched for the optimal solution. Starting with an initial range of models, these algorithms progressively modify the solution by incorporating the evolutionary behavior of biological systems. The fitness of each solution is measured by a quantitative, objective function. The fittest members of each population then are combined using probabilistic transition rules, to form a new offspring population. This procedure is repeated through a large number of generations until the best solution is obtained, based on the fitness measure [Michalewicz, 1992]. It has been demonstrated that those members of the population with a fitness greater than the average fitness of the population itself will increase in number exponentially, effectively accelerating the convergence of the inversion process [Holland, 1975; Goldberg, 1989; Mitchell et al, 1992].

Genetic algorithms are an attractive global search tool suitable for the irregular, multimodal fitness functions typically observed in nonlinear optimization problems in the physical sciences. Because of their initial random and progressively more deterministic sampling of the parameter space, these algorithms offer the possibility of efficiently and relatively rapidly locating the most promising regions of the solution space. Their ability to solve nonlinear, nonlocal optimiza-

tion problems without a priori knowledge of curvature information precludes the need for derivative computations, a particularly important feature because it allows for fast approximate forward modeling where no derivative information is available [Wright, 1991]. Moreover, since genetic algorithms sample the space directly, linearization of the problem is avoided, thus avoiding errors involved in this approximation.

Several recent studies seismic studies have employed GAs to invert for seismic structure [Stoffa and Sen, 1991; Sen and Stoffa, 1992; Stoffa et al., 1994; Boschetti et al., 1996], hypocenter relocation [Billings et al., 1994], seismic phase alignment [Winchester et al., 1993], mantle velocity structure [Curtis et al., 1995; Lomax and Snieder, 1995; Neves et al, 1996], crustal velocity structure [Jin and Madariaga, 1993; Drikkonigen and White, 1995; Zhou et al., 1995; Battacharyya et al., 1999] and fault zone geometry [Yu and Rundle, 1995; Yu et al., 1998]. This program, shown schematically in Figure A.3, employs a random number generator to produce an initial set of 100 potential values for each of the model parameters, which are then coded as genes. These genes in turn combine to form the chromosomes for each member of the initial population of 100 potential solutions to the inversion problem. These members are ranked, from best to worst, according to an external fitness function. The fitness function is obtained from the chi-square fit to the difference between the measured displacements and the corresponding deformations calculated for each potential solution. The members with the lowest chi-square value are the fittest and are selected, based on their relative fitness, to contribute to the next generation, where the genetic operations of crossover and mutation take place.

Crossover and mutation alter the new, fitter population in a process of controlled yet random information exchange, as shown in Figure A.4. First, two randomly mated members of the new population swap genes based on their position

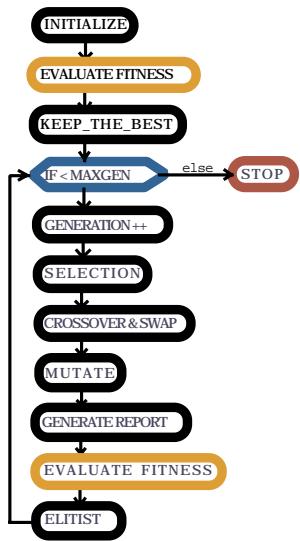


Figure A.3: Schematic of genetic algorithm inversion program.

versus a randomly generated position in the string. These parent chromosomes are recombined, with the left portion of one parent and the right portion of the other parent creating one new offspring of the correct length. The corresponding recombination on the remaining subchromosomes creates a second offspring [Wright, 1991]. This crossover process is repeated for the entire population, then mutation is introduced to ensure against the occasional loss of valuable genetic material. Random alteration of a small number of the genes corresponds to mutation in a GA. For a small percentage of the population, the current gene is replaced by a new, randomly generated value [Michalewicz, 1992]. After completion of both crossover and mutation, the population is reevaluated as above, and the process is repeated over subsequent generations, exploiting information in past generations to search the parameter space with improved performance.

Specific features of the current GA include real-valued genes, an elitist function, and a windowing fitness function. The coding of genes as real parameters, as opposed to the standard binary values of original genetic algorithms, has several advantages for the optimization algorithm. The efficiency of the algorithm is increased as the necessity for converting bit strings to real values in every fitness evaluation is eliminated; there is no loss of precision as a result of a binary representation; and the crossover and mutation techniques can be customized for the specific real parameter representation of the inversion [Wright, 1991]. The CPU intensive nature of the particular fitness function employed for the ellipsoidal modeling and the necessity for high precision calculations made the use of a real-valued GA very attractive. Thus, an individual is represented by an array of real values, which in this case are model parameters.

The elitist function employed in this algorithm ensures that the best member of all populations, current and prior, is not only stored in memory, but is also copied into the current population, where it might otherwise no longer exist. This

a

## CROSSOVER, REAL-VALUED

GENOME 1

A1A1A	B1B1B	C1C1C	D1D1D
-------	-------	-------	-------

GENOME 2

A2A2A	B2B2B	C2C2C	D2D2D
-------	-------	-------	-------

Random location →

OFFSPRING 1

A1A1A	B1B1B	C1C1C	D2D2D
-------	-------	-------	-------

OFFSPRING 2

A2A2A	B2B2B	C2C1C	D1D1D
-------	-------	-------	-------

b

## MUTATION, REAL-VALUED

GENOME

A1A1A	B1B1B	C1C1C	D1D1D
-------	-------	-------	-------

Randomly chosen gene

NEW GENE, RANDOM VALUE



CCCCC

OFFSPRING

A1A1A	B1B1B	CCCCC	D1D1D
-------	-------	-------	-------

Figure A.4: Schematic view of GA operations, a) crossover and b) mutation.

particular genetic algorithm produces new generations synchronously, i.e., the old generation is completely replaced after crossover and mutation take place. However, subsequent to the evaluation of the new generation by the fitness function, the population is checked against the best member of all preceding generations. If that particular member has been lost from the active population, through either crossover or mutation, the program replaces the worst member of the new generation with that best member, so that those features which make it the fittest solution to the inversion problem are not lost from the pool.

In addition, a windowing function was added to the algorithm to prevent search stagnation. Significant tradeoffs between the volcanic source properties (for example, volume and depth) can result in a complex fitness landscape with numerous local minima. This particular fitness landscape generated a population with a small standard deviation of the population fitness after only a relatively few generations. This decreased the selection pressure toward better structures, causing the search to stagnate and even reach premature convergence in several instances. As a result, I created a routine to window the fitness by subtracting the fitness value of the worst member of the previous generation from every new population fitness, prior to selecting the next generation. This ensured that those members with a better relative fitness were included in a greater proportion in the next generation, despite the small absolute difference in their fitness.

Finally, in this particular inversion, the parameter space for the source volume is very large - ranging from 0 to 10,000,000 m<sup>3</sup>. This caused the GA to search the parameter space very slowly, slowing down the convergence. Although most literature on the subject states that both the population size and the method of crossover have little effect on the GA convergence speed, I found that this was not true in an actual geophysical inverse problem with a large search space and significant interaction between the parameters, as seen below. As a result, I im-

plemented a proportional crossover function, where the two parents,  $A$  and  $B$ , are mated together proportionally. A random number,  $\alpha$ , is chosen between 0 and 1.0, and the value of each new parameter for the child,  $C$ , is

$$C = \alpha A + (\alpha - 1)B. \quad (\text{A.1})$$

This proportional crossover allows the GA to search the large parameter space faster, without relying on the relatively slow mutation rate to add new values. In addition, I increased the population size from 100 to 2500 so that there would be better coverage of the parameter space in the initial random sampling process. These two modifications increased the convergence time from more than 500,000 generations to 20,000.

### **A.3.2      The Fitness Function – the forward problem.**

The genetic algorithm works according to selection rules as defined by the laws of evolutionary genetics. The algorithm seeks the “fittest” model, i.e., the model which produces solutions to the given problem which are closest to the observed measurements. In the case of a volcanic source or sources at Long Valley caldera, it is the model whose parameters, when input into the forward source model, produce surface deformations which best match the measured deformations. Given the expected range of values for the model parameters, the algorithm randomly selects a set of models and proceeds to evolve them to produce better, fitter models.

The fitness value,  $FV$ , for any one model, is calculated in the fitness subroutine, which contains the geophysical information specific to this problem. The deformation pattern for the expected volcanic source(s) is calculated in this subroutine and then the value of chi-square is calculated for that deformation pattern as it relates to the measured deformation pattern:

$$\chi^2 = \sum_k \frac{(C_k - E_k)^2}{\sigma_k^2}. \quad (\text{A.2})$$

$C_k$  are the calculated deformations,  $E_k$  are the expected, or measured deformations, and  $\sigma_k$  are the standard deviations for each measurement.

The GA picks the fittest members of the population based upon the maximum fitness value, yet the value of chi-square decreases as the model approaches the correct solution. The following fitness value was chosen:

$$FV = \exp(-\chi^2) \quad (\text{A.3})$$

effectively converting the fitness to a continuously increasing function, such that the genetic algorithm seeks the model with the largest value for  $FV$ .

#### A.4 Spherical vs. Ellipsoidal Functions

A genetic algorithm searches the parameter space for a given problem by applying a forward model, contained in what is called a fitness function, to test the possible solutions based on those parameters. In this problem, the fitness function contains the geophysical information relating the surface deformations to the particular volcanic source model. As the inflation event for the period from 1988 to 1992 at Long Valley was considered to be complex by most researchers, and generally was modeled with two inflation sources, I decided to study an event made up of either two spheroidal or two ellipsoidal source models using the genetic algorithm approach. The half-space is considered to be homogeneous, isotropic and elastic over the relatively short time periods in question.

#### A.4.1 Spherical Source Model

Mogi [1958], in studying various volcanos in Japan and Hawaii, related the surface deformations recorded both before and after volcanic eruptions to the magmatic source below (see Figure A.5). He used the elasticity equations for a small, expanding sphere in an infinite half-space, such that the radius of the sphere is much smaller than its depth, with

$$\delta d = \frac{3r^3 P}{4\mu} \left[ \frac{d}{(f^2 + d^2)^{3/2}} \right] \quad (\text{A.4})$$

$$\delta z = \frac{3r^3 P}{4\mu} \left[ \frac{f}{(f^2 + d^2)^{3/2}} \right] \quad (\text{A.5})$$

where  $\delta d$  equals displacement in the radial direction at the surface,  $\delta z$  equals vertical displacement at the surface,  $r$  is the radius of the sphere,  $d$  is the radial distance at the surface from the center of the sphere,  $P$  equals the change in the hydrostatic pressure in the sphere,  $f = h$  is the depth of the center of the sphere from the surface, and  $\mu L$  is Lame's constant ( $\lambda = \mu$  for Poisson's ratio of  $\nu = 0.25$ ) [Mogi, 1958].

As  $d^2 = x^2 + y^2$ , where  $x$  and  $y$  are the distance in the  $x$  and  $y$  direction from the center of the sphere along the surface of the earth,  $U_{xyz}$  is the displacement in the  $x$ ,  $y$ , or  $z$  direction, and if  $R^2 = x^2 + y^2 + h^2$ , then

$$U_x = (3r^3 P / (4\mu))(x/R^3) \quad (\text{A.6})$$

$$U_y = (3r^3 P / (4\mu))(y/R^3) \quad (\text{A.7})$$

$$U_z = (3r^3 P / (4\mu))(h/R^3). \quad (\text{A.8})$$

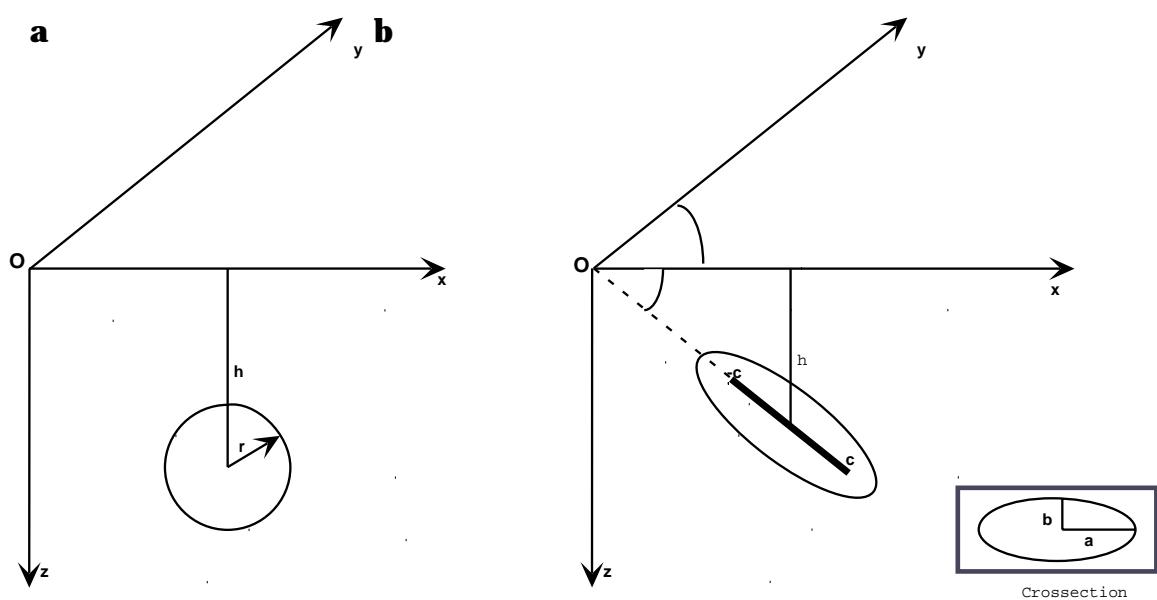


Figure A.5: Volcanic source models for Long Valley caldera. a) Spherical Mogi source and b) finite, prolate ellipsoid.

For the initial fitness function, substitute for pressure,  $P$ , per McTigue, 1987, for a point source at depth greater than twice its radius, such that  $Pr^3 = \mu\Delta V/\pi$ , where  $\Delta V$  is the change in inflation volume. The above equations then reduce to

$$U_x = 3\Delta Vx/(4\pi R^3), \quad (\text{A.9})$$

$$U_y = 3\Delta Vy/(4\pi R^3), \quad (\text{A.10})$$

$$U_z = 3\Delta Vh/(4\pi R^3), \quad (\text{A.11})$$

where  $\Delta V$  = the volume of expansion for the sphere,  $x$  and  $y$  = the distance in  $x$  and  $y$  direction from the source, and  $h$  = depth of the expanding source .

Initially, synthetic data was generated using this function, for measurement locations on the surface of the earth similar in pattern and type to that found at Long Valley caldera. These synthetic surface deformations then were input into the GA to test its ability to invert for the correct four source parameters -  $x$  and  $y$  location, depth and expansion volume. In addition, random gaussian error appropriate to each type of measurement was added to the data to test the accuracy of the GA. Issues such as stability, sensitivity to each parameter, both its type and associated random error, and accuracy were analyzed for this inversion.

Every additional sphere included in the model incorporates four additional parameters to be solved for in the inversion. A two sphere model therefore results in an eight parameter fit. The response of the GA to multiple spherical sources has implications both on the capability of the GA itself to solve for multiple and complex sources, as well as the nature of the problem itself.

#### A.4.2 Ellipsoidal Source Model

In 1986, Davis obtained the expressions for surface deformation due to inflation of an point source ellipsoidal cavity in an elastic half-space from the work of Eshelby in 1957, but using the point force solutions of Mindlin, 1936. In 1988, Yang et al., 1988, derived analytic expressions for a finite, prolate spheroidal source which more accurately model the near field deformations. As the deformation measurements are obtained in the near field, it was the second, finite ellipsoid which was used for the ellipsoidal source model, despite the fact that the numerical integration included in the technique added to the computation time.

The equations obtained by Yang et al., 1988, for the surface deformation due to a finite, prolate ellipsoid, as shown in Figure A.5, can be found in that reference in their entirety. The equations for the intensity pressures for the double forces and dilatational centers, respectively, are

$$P_d = \frac{\pi ab^2}{2\mu c^3} (c^2 - \xi^2) P^\dagger, \quad P_c = \frac{\pi ab^2}{2\mu c^3} [a_1 + b_1(c^2 - \xi^2)], \quad (\text{A.12})$$

where  $a$  is the length of the semimajor axis,  $b$  is the length of the semiminor axis,  $c$  is the distance from the center of the ellipsoid to one of the foci,  $P$  is the intensity of stress at the ellipsoidal surface, and  $a_1$ ,  $b_1$ ,  $P^\dagger$ , and the related  $P^*$  are constants related to the geometry of the ellipsoid as follows:

$$\begin{aligned}
a_1 &= -2b^2 P^\dagger, & b_1 &= 3\frac{b^2}{c^2} P^\dagger + 2(1-2\nu)P^* \\
P^* &= P \frac{B_1 a_{22} - B_2 a_{12}}{a_{11} a_{22} - a_{12} a_{21}}, & P^\dagger &= P \frac{B_2 a_{11} - B_1 a_{21}}{a_{11} a_{22} - a_{12} a_{21}} \\
a_{11} &= Q a^2 I_{aa} + R I_a - 1, & a_{12} &= (Q - R) I_a - 1 \\
a_{21} &= 2R(I_a - 4\pi), & a_{22} &= -16\pi R \\
B_1 &= \frac{3\lambda(Q-R)I_a - (3\lambda-2\mu)}{3\lambda+2\mu}, & B_2 &= \frac{6(\mu-8\pi\lambda R)}{3\lambda+2\mu} \\
Q &= \frac{3}{8\pi(1-\nu)}, & R &= \frac{1-2\nu}{8\pi(1-\nu)} \\
I_a &= -2\pi ab^2 \left[ \frac{2}{ac^2} + \frac{\ln(\frac{a-c}{a+c})}{c^3} \right], & I_{aa} &= -2\pi ab^2 \left[ \frac{2}{3a^3c^2} + \frac{2}{ac^4} + \frac{\ln(\frac{a-c}{a+c})}{c^5} \right]
\end{aligned}$$

The deformation in the z-direction,  $U_z$ , is given as an example:

$$U_z = U_3^* + U_3^\dagger, \quad (\text{A.13})$$

$$\begin{aligned}
U_3^* &= \frac{ab^2}{16c^3\mu[1-\nu]} \{ -\cos\theta[A_1^*r_2 + (3-4\nu)\overline{A}_1^*q_2 - F_1^*q_2] + \sin\theta(B^* + F_2^*) \\
&\quad + 2z\overline{A}_1^* \cos^2\theta \} \quad (\text{A.14})
\end{aligned}$$

$$\begin{aligned}
U_3^\dagger &= \frac{ab^2 P^\dagger}{8c^3\mu[1-\nu]} \{ \cos\theta[-A_1 r_2 + (3-4\nu)\overline{A}_1 q_2 + F_1 q_2] + 4(1-\nu)(1-2\nu)f_3 \\
&\quad + 4(1-\nu)\sin\theta(A_2 + \overline{A}_2) + \sin\theta[A_3 - (3-4\nu)(\overline{A}_3 + 2B) + F_2] \} \quad (\text{A.15})
\end{aligned}$$

and where  $|$  signifies the integration over  $\xi$  between  $c$  and  $-c$ , with coordinate system as shown in Figure A.6, and the other variables are

$$\begin{aligned}
A_1^* &= \frac{a_1}{R_1(R_1+\overline{r}_3)} + b_1 \left[ \ln(R_1 + \overline{r}_3) + \frac{r_3 + \xi}{R_1 + \overline{r}_3} \right] \\
\overline{A}_1^* &= \frac{-a_1}{(R_2(R_2+\overline{q}_3))} - b_1 \left[ \ln(R_2 + \overline{q}_3) + \frac{q_3 - \xi}{R_2 + \overline{q}_3} \right] \\
F_1 &= -2z \sin\theta \left[ \frac{\xi(\xi+C_0)}{R_2^3} + \frac{R_2 + \xi + C_0}{R_2(R_2 + \overline{q}_3)} + 4(1-\nu) \frac{R_2 + \xi}{R_2(R_2 + \overline{q}_3)} \right] \\
F_1^* &= 2z \left\{ \cos\theta q_2 \left[ \frac{a_1(2R_2 + \overline{q}_3)}{R_2^3(R_2 + \overline{q}_3)^2} - b_1 \frac{R_2 + 2\xi}{R_2(R_2 + \overline{q}_3)^2} \right] + \sin\theta \left[ \frac{a_1}{R_2^3} - 2b_1 \frac{R_2 + \xi}{R_2(R_2 + \overline{q}_3)} \right] \right\} \\
F_2 &= -2z \sin\theta \left[ \frac{\xi(\xi+C_0)\overline{q}_3}{R_2^3} + \frac{C_0}{R_2} + (5-4\nu)\overline{A}_1 \right]
\end{aligned}$$

$$\begin{aligned}
F_2^* &= 2z \left\{ \frac{a_1 \overline{y_3}}{R_2^3} - 2b_1 \left[ \sin \theta \overline{A_1} + \cos \theta \frac{q_2(R_2 + \xi)}{R_2(R_2 + \overline{q_3})} \right] \right\} \\
B &= \frac{\xi(\xi + C_0)}{R_2} - \overline{A_2} - C_0 \ln(R_2 + \overline{q_3}) \\
B^* &= \left( \frac{a_1}{R_1} + 2b_1 A_2 \right) + (3 - 4\nu) \left( \frac{a_1}{R_2} + 2b_1 \overline{A_2} \right) \\
f_3 &= \frac{1}{\cos \theta} [q_2 \ln(R_2 + \overline{q_3}) - q_2 \sin \theta \ln(R_2 + \overline{y_3}) + 2y_1 \tan^{-1} \beta] + 2 \sin \theta \overline{A_2} + \\
&q_3 \ln(R_2 + \overline{y_3}) - \xi \\
A_1 &= \frac{\xi}{R_1} + \ln(R_1 + \overline{r_3}), \quad \overline{A_1} = \frac{\xi}{R_2} - \ln(R_2 + \overline{q_3}) \\
A_2 &= R_1 - r_3 \ln(R_1 + \overline{r_3}), \quad \overline{A_2} = R_2 - q_3 \ln(R_2 + \overline{q_3}) \\
A_3 &= \frac{\xi \overline{r_3}}{R_1} + R_1, \quad \overline{A_3} = \frac{\xi \overline{q_3}}{R_2} - R_2
\end{aligned}$$

From Figure A.6,

$$\begin{aligned}
x_1 &= x, \quad x_2 = y - y_0, \quad x_3 = z - z_0, \quad \overline{x_3} = z + z_0 \\
y_1 &= x_1, \quad y_2 = x_2 - \xi_2, \quad y_3 = x_3 - \xi_3, \quad \overline{y_3} = x_3 + \xi_3 \\
r_2 &= x_2 \sin \theta - x_3 \cos \theta, \quad q_2 = x_2 \sin \theta + \overline{x_3} \cos \theta \\
r_3 &= x_2 \cos \theta + x_3 \sin \theta, \quad q_3 = -x_2 \cos \theta + \overline{x_3} \sin \theta \\
\overline{r_3} &= r_3 - \xi, \quad \overline{q_3} = q_3 + \xi \\
R_1 &= (y_1^2 + y_2^2 + y_3^2)^{1/2}, \quad R_2 = (y_1^2 + y_2^2 + \overline{y_3^2})^{1/2} \\
C_0 &= y_0 \cos \theta + z_0 \sin \theta, \quad \beta = \frac{q_2 \cos \theta + (1 + \sin \theta)(R_2 + \overline{q_3})}{y_1 \cos \theta}
\end{aligned}$$

The resulting equation for  $U_{x,y,z}$  is a function of the pressure,  $P$ , the ellipsoidal geometry,  $a$ ,  $b$ , and  $c$ , and the location and orientation of the ellipsoid as specified by  $x$ ,  $y$ , the depth,  $z_0$ , the dip angle  $\theta$ , and the orientation angle,  $\phi$ . By using the original derivation by Eshelby, 1957, and assuming again that the depth is more than twice the semimajor axis,  $a$ , and that  $P$  is constant at the surface of the sphere, a similar substitution to that of McTigue, 1987, can be derived as shown below [Tiampo et al., 2000].

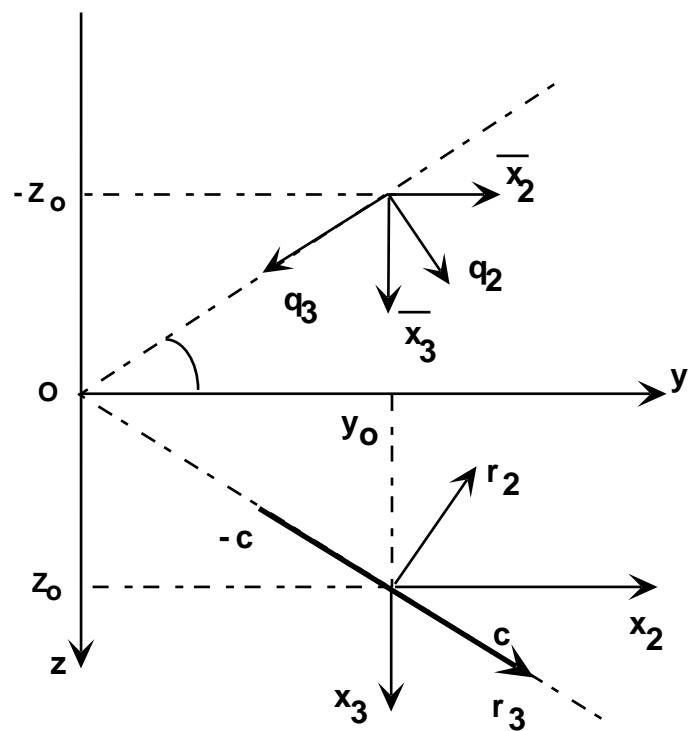


Figure A.6: Coordinate system for ellipsoidal model.

#### A.4.2.1 Volumetric Parameterization

If the strain is designated  $e_{ij}$  and  $p_{ij}$  is stress,  $p_{ij} = \lambda e_{kk} \delta_{ij} + 2\mu e_{ij}$ , then

$$p_{kk} = 3\lambda e_{kk} + 2\mu e_{kk} = 3\kappa e_{kk},$$

where  $\kappa = \lambda + \frac{2}{3}\mu$ .

Simplifying to  $p_{kk} = p$  and  $e_{kk} = e$ , per Eshelby, and remembering that for this problem  $\nu \equiv 0.25$ , i.e.  $\mu = \lambda$ ,

$$p = 5\mu e.$$

Again using Eshelby's notation and letting  $e_{ij}^T$  be the stress-free strain which the ellipsoidal inclusion would undergo if there were no matrix,  $e_{ij}^C$  be the constrained strain of both the inclusion and the matrix, and  $p_{ij}^I$  is the stress inside the inclusion, Eshelby derived:

$$p_{ij}^I = \lambda(e^C - e^T)\delta_{ij} + 2\mu(e_{ij}^C - e_{ij}^T), \text{ so that } p^I = 5\mu(e^C - e^T).$$

But the stress-free strain is related to the constrained strain by the dilatation relation

$$e^C = \frac{(1+\nu)}{3(1-\nu)}e^T,$$

or, with  $\nu = 0.25$ ,

$$e^T = \frac{9}{5}e^C.$$

Therefore,

$$p^I = 5\mu(e^C - \frac{9}{5}e^C) = -4\mu e^C.$$

Remembering that  $P = -p^I$ ,  $P = 4\mu e^C$ , so that

$$e^C = \frac{P}{4\mu}.$$

Dilatation is

$$\frac{\Delta V}{V} = 3e^C,$$

therefore

$$\frac{\Delta V}{V} = \frac{3P}{4\mu}.$$

The volume of a prolate ellipsoid is

$$V = \frac{4\pi ab^2}{3},$$

$$\Delta V = \frac{3P}{4\mu} \left[ \frac{4\pi ab^2}{3} \right] = \frac{Pab^2\pi}{\mu}.$$

This leads to the result that

$$Pab^2 = \frac{\mu\Delta V}{\pi}.$$

As a result of the above derivation, the equations for  $U_{x,y,z}$  can be simplified by substituting the above in for  $Pab^2$  and noting that, if  $\Delta V$  and  $a$  are known,  $b$  can be solved for using the equation for the volume of a prolate ellipsoid noted above, and that  $c$  can be eliminated from the equations as well using the ellipsoidal relation that  $c^2 = a^2 + b^2$ . The final equation for  $U_{x,y,z}$  are a function of seven parameters,  $x$ ,  $y$ ,  $z_0$ ,  $\phi$ ,  $\theta$ ,  $a$  and  $\Delta V$ . The resulting inversion for two ellipsoids results in a fourteen parameter fit [Tiampo et al., 2000].

## A.5 Synthetic Tests and Sensitivity Analyses

Synthetic tests were developed for the two models described above, in order to test the ability of the GA to determine the correct source parameters. An important issue in any inversion is the size and scope of the available data for inversion. This is of particular interest in a genetic algorithm inversion, as larger data sets allow earlier removal of random error, resulting in earlier convergence to the solution. Synthetic tests were done on a four parameter, one sphere model to

o. of Data P t	CPU Time (mins)	Depth Error (percent)	Volume Error (percent)	Location Error (m)
70	4.95	5.28	17.2	404.48
230	10.22	4.68	3.60	82.5
321	19.90	1.89	1.60	63.3
500	22.30	2.1	1.80	20.0
1000	40.23	1.48	2.80	20.0
2000	77.33	0.48	0.40	28.3
4000	210.0	0.52	1.00	0.00

Table A.1: Generations to convergence, data sets of various sizes.

determine the number of generations to convergence for varying sized data sets. The random gaussian error on this synthetic data is approximately 10 mm. Table A.1 shows the generations to convergence for each data set, both with and without random error, in varying surface measurement patterns. Not surprisingly, for additional parameters, larger numbers of data points or increasingly larger numbers of generations are required to reduce the error associated with the gaussian noise to an acceptable amount. In addition, several additional tests were conducted to determine the nature of a two Mogi source model as well as the more complicated ellipsoidal function and those conditions under which the GA behaved optimally.

### A.5.1 Sensitivity Tests

#### A.5.1.1 The Sphere

Sensitivity tests were conducted on a single sphere, located at 10 km in both the x and y direction from the origin, at a depth of 5 km, and with a volume of 0.1 km<sup>3</sup>. Standard errors on the data were 5 and 10 mm, again. In each case, one parameter was varied while the other three remained fixed. Results are shown in Figure A.7. The residual in meters is plotted against the normalized variable in question. Not surprisingly, the x (or y) distance is a steeply varying parameter, with a single, easily defined minimum. This corresponds with the GAs ability to

invariably locate the sphere in the horizontal plane, as demonstrated in Table A.1, regardless of the associated random error or the number of available data points.

The normalized volume has a much more gradual slope to its symmetric residual function. A residual of only 10 mm corresponds to an error in the volume of 15 percent. This, coupled with the axisymmetric minimum associated with the depth parameter, accounts for the larger error associated with the GA inversion for those parameters. It should be noted that the nature of the function itself causes the GA to locate smaller spheres at a shallower depth, but to more often find larger spheres at deeper depths. This feature has implications not only for this inversion technique, but for others performed using Mogi source models.

#### A.5.1.2 The Ellipsoid

A similar sensitivity analysis was performed on the seven parameters for the finite, prolate ellipsoid. One source was used, located at 10 km in the x and y direction again, and again at a depth and volume of 5 km and 0.1 km<sup>3</sup>. The rotation about the x and y axes,  $\phi$ , was 45°, while the dip angle,  $\theta$ , was 30°, and  $a$ , the semimajor axis, was 250 m. The results are shown in Figure A.8. Again, the x (or y) location is steeply varying, although slightly asymmetric. The volume is again a gradual function, with a residual of 10 mm corresponding to an error in the volume of over 30 percent. It should be noted that both  $\phi$  and  $\theta$  have two minima. In the case of  $\phi$ , the second minima is located at such a high residual that, as will be shown later, the GA locates the ellipsoidal orientation with relatively high accuracy. However, in the case of  $\theta$ , the second minima is relatively close, in terms of the residual, to the global minima, causing the GA to occasionally specify a value for the dip angle at five times its actual value.

The depth and semimajor axis sensitivities are interesting functions, both containing important false minimas. Below one-quarter of the normalized depth,

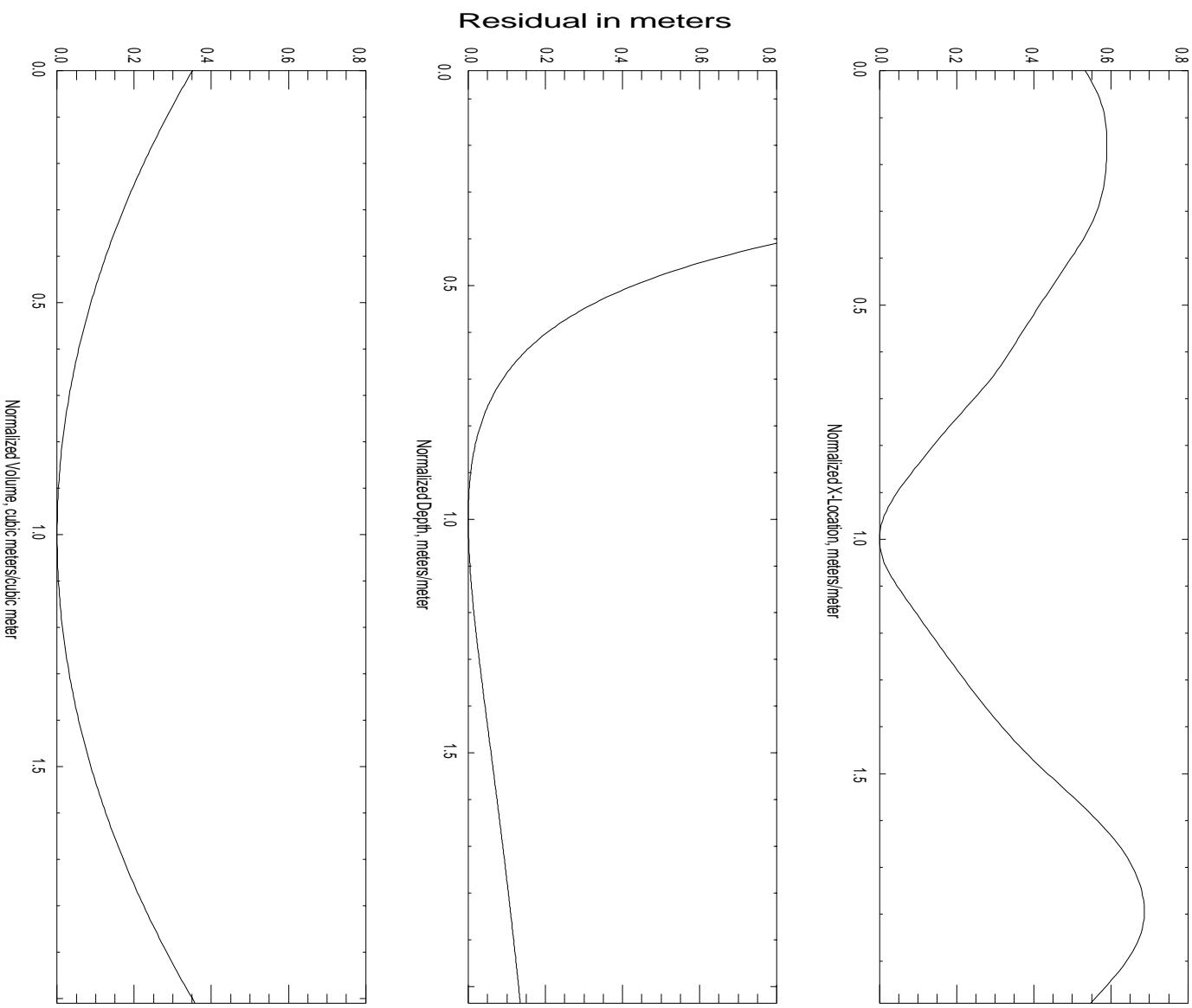


Figure A.7: Sensitivity of fit to Mogi source parameters.

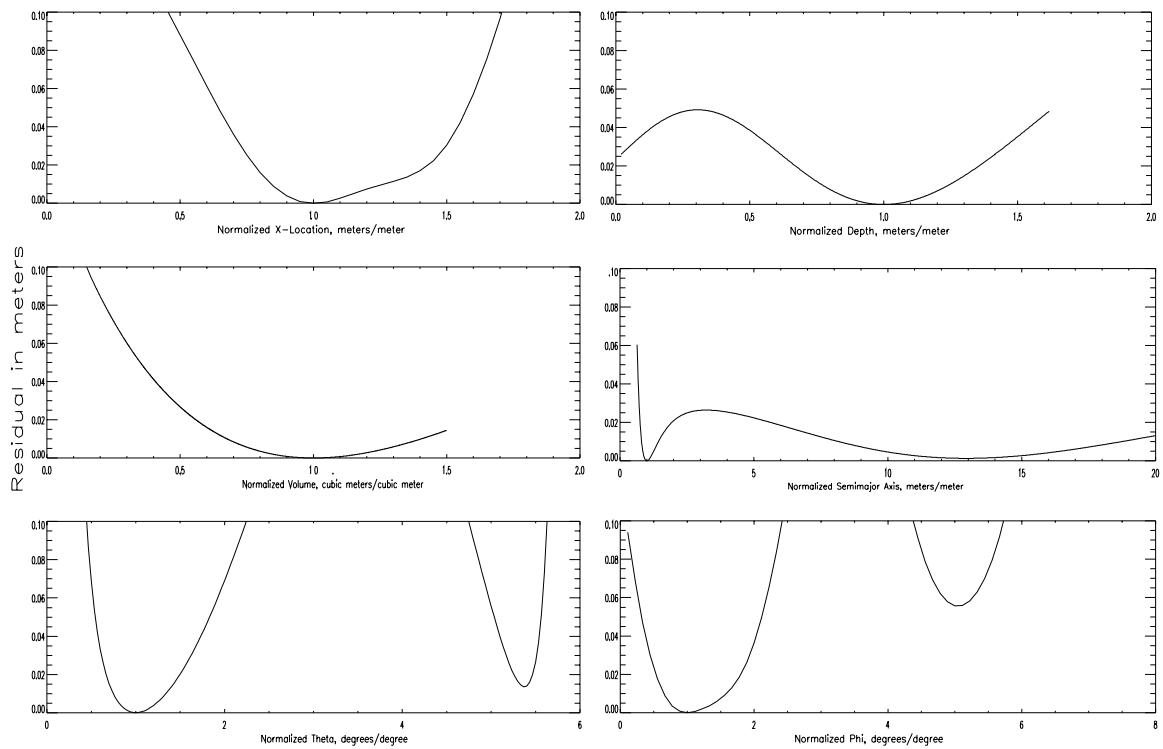


Figure A.8: Sensitivity of fit to ellipsoidal parameters.

the function sinks into a rapidly decreasing local minimum. This feature can be seen in the synthetic tests, where occasionally a solution located at 10 to 20 percent of the actual depth will emerge. Interestingly, these shadow solutions are usually associated with large values of the semimajor axis,  $a$ , which arise from the wide local minimum with very small residuals noticed in that function. Paradoxically, these shallower solutions also tend to have larger volumes. Apparently, these long, narrow ellipses require a larger volume and shallower depth to produce deformation patterns of the appropriate magnitude. This characteristic, one of those inherent in the nonlinear nature of the ellipsoidal source, complicates any evaluation of the inversion.

In a further attempt to evaluate the nonlinear features and interactions involving  $\theta$ ,  $a$ , volume and depth, a single ellipsoid inversion was studied also in order to determine the variation in fitness with these different parameters.

A synthetic model for one ellipsoid was specified, located at 10 km in the x and y directions, a  $\phi$  of  $45^\circ$ , while the dip angle,  $\theta$ , was  $30^\circ$ . The semimajor axis  $a$ , was 250 m, the volume was only  $0.05 \text{ km}^3$ , and the depth was 2500 m. A random gaussian error with a standard deviation slightly more than that in the actual expansion line and leveling data available, 5 mm and 10 mm respectively, was added to the associated deformation for 126 data points, and more than 45 inversions of over 100,000 generations each were run on the data. The GA had little difficulty in finding the location of the ellipsoid - the standard deviation for all runs with  $FV \geq 0.99975$  was only 693 m, clustered around 10 km in each direction, with no outliers. In addition, the mean for  $\phi$ , was  $49^\circ$ , with a standard deviation of only  $6^\circ$ . Again, the spatial pattern of the deformation appears to dominate the inversion. However, if one plots those runs whose  $FV$  was above 0.9997, corresponding to a  $\chi^2$  of 300, upon the assumption that any inversion with a fitness less than that would be considered unacceptable, one can

see the large variation in results for the remaining parameters (Figure A.9). The variation in  $\theta$  is of the least concern - the somewhat bimodal nature of the results are easily identified, and the variation appears to have little impact on the other variables, such as volume, which are of more interest geophysically. However, the plots for volume,  $a$ , and depth are of greater concern. The values of  $a$  with the greatest  $FV$  are those that are close to 250 m, but the corresponding volumes are widely varying. This appears to be a result of the lack of constraint on the depth, in which some of the shallowest sources correspond to large depths and higher values of  $a$ . Clearly, a minimum acceptable value for the depth is required for the inversion process, and any results associated with shallow sources and long semimajor axes must be examined more closely.

## A.6 Inversion Results

Results for the GA inversion for both the two-sphere and two-ellipsoid models are shown in Figures A.10 and A.11. Results for the larger sphere in the two-sphere inversion are remarkably stable. Every run with an  $FV$  greater than 0.9991, whether over 50,000 or 500,000 generations, produces a sphere within 500 m areally for the larger, deeper sphere, and within 500 meters in depth of 9.89 km. The volume,  $V = 0.036 \text{ km}^3$ , displays a similar stability, as it is intrinsically linked, as shown above, with the depth.

It is more difficult to assess the results for the second sphere, as this particular solution has a significantly higher fitness than the rest of the runs. The volume of this smaller source is only  $0.008 \text{ km}^3$ . However, found at 7.3 km in depth, it is similar in location to that found by others, and is coincident with the local seismicity of that period.  $\chi^2$  for this solution is 631.

The two-ellipsoid model again finds both sources in similar locations, but substantially deeper. The solution for the larger source under the resurgent dome

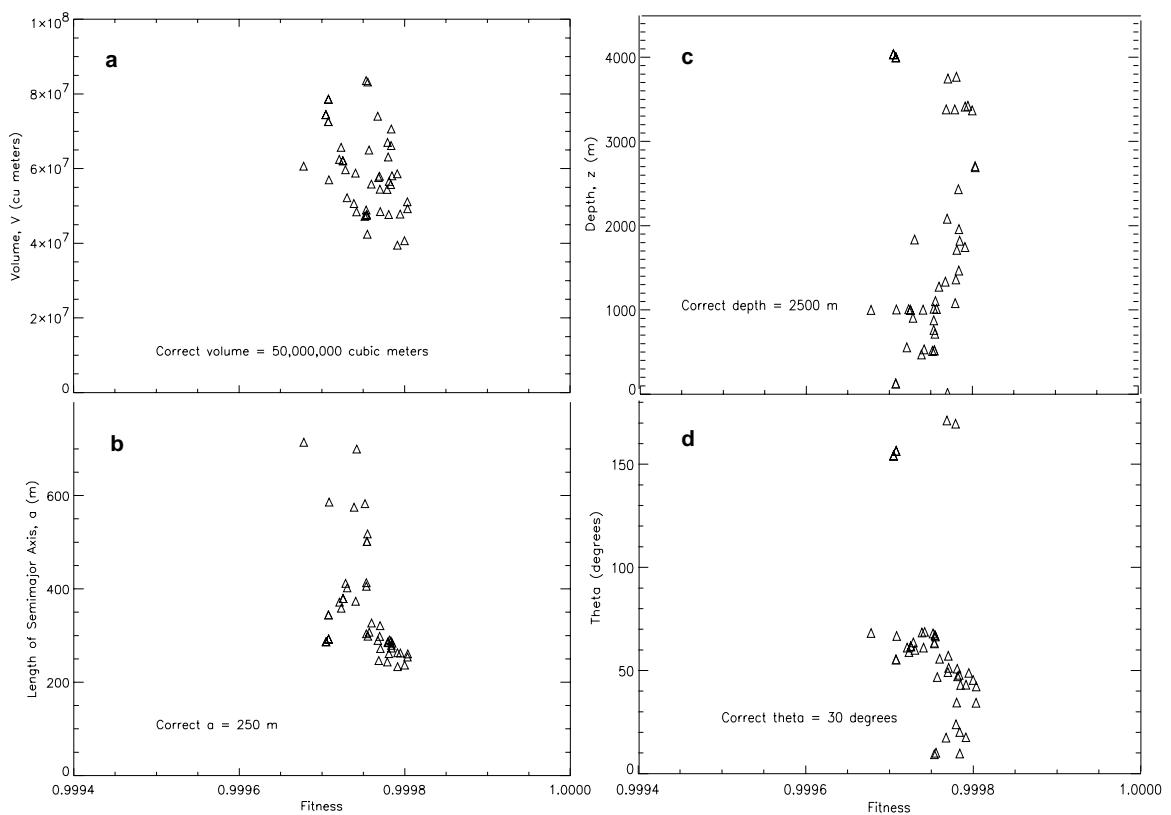


Figure A.9: Variation in fitness with a) volume, b) length of semimajor axis, c) depth, and d) theta.

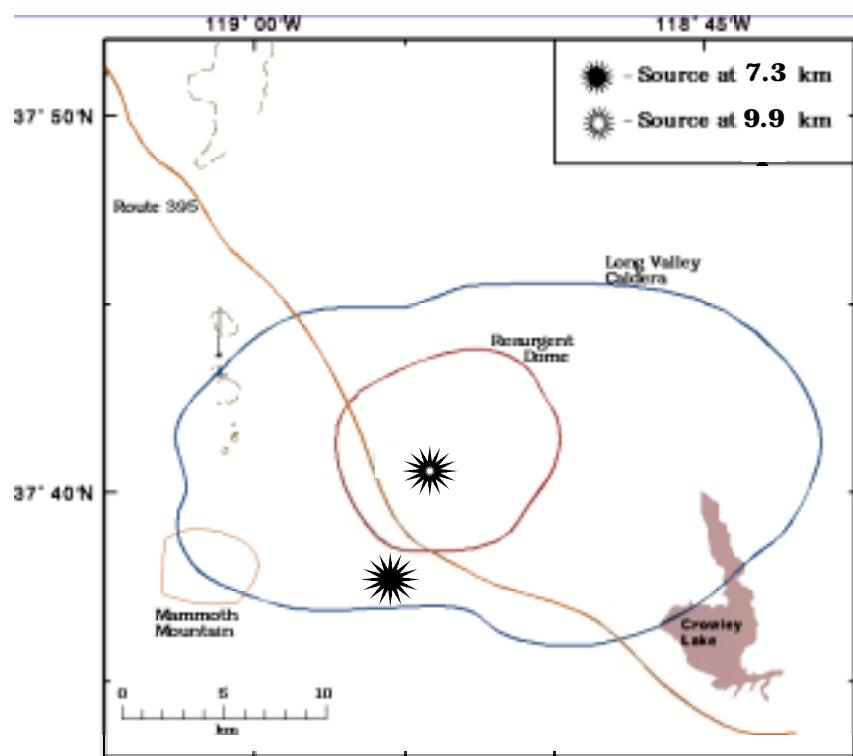


Figure A.10: Location of Mogi point sources as determined by GA inversion.

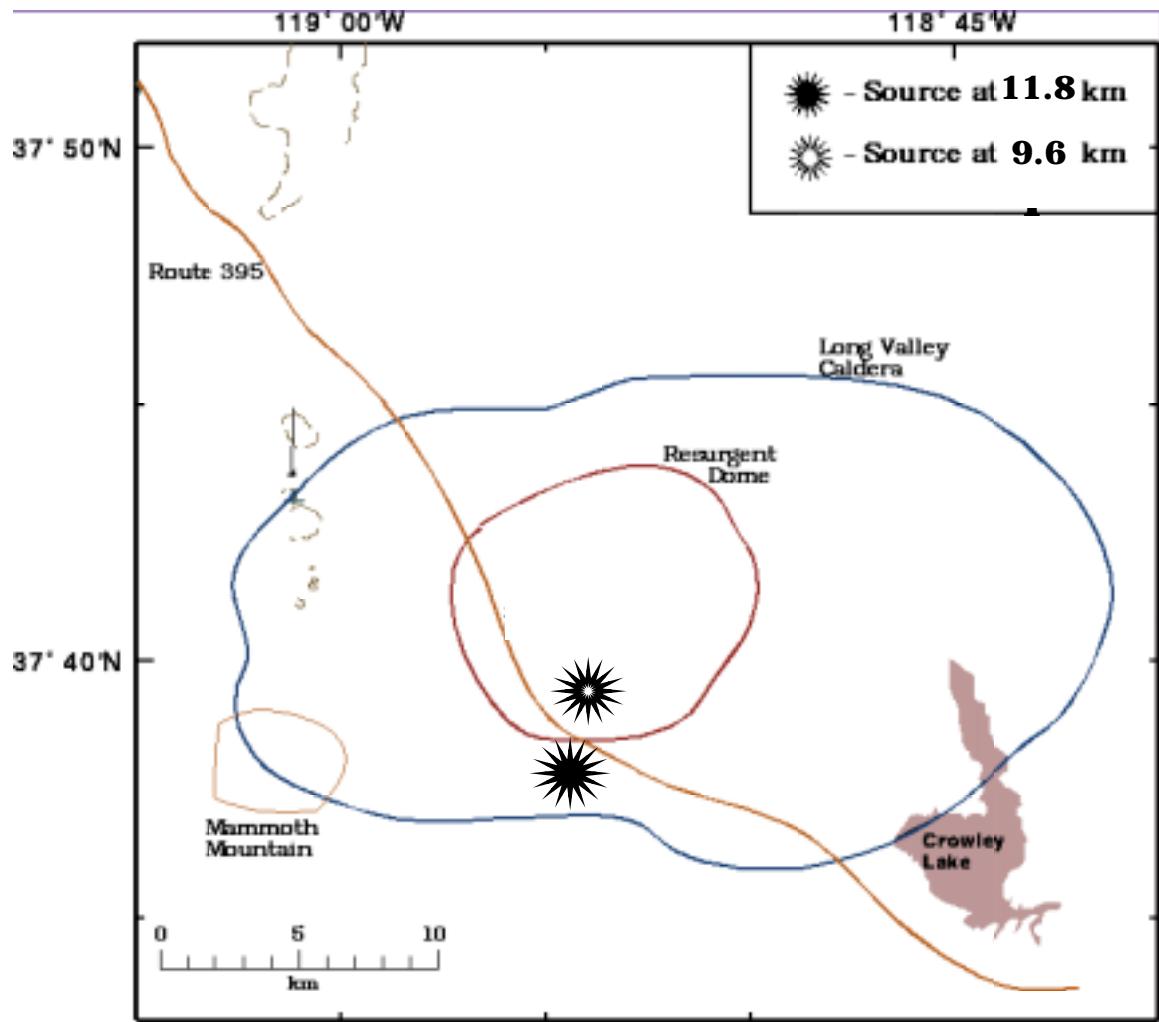


Figure A.11: Location of finite ellipsoid sources as determined by GA inversion.  
 a) Source at 9.6 km:  $\phi = 179^\circ$ ,  $\theta = 109^\circ$ ,  $V=0.037 \text{ km}^3$ ,  $a = 2000 \text{ m}$ . b) Source at 11.8 km:  $\phi = 105^\circ$ ,  $\theta = 1^\circ$ ,  $V=0.002 \text{ km}^3$ ,  $a = 290 \text{ m}$ .

has a volume and location consistent with other inversions, but the ratio of semi-major axis to semiminor axis is approximately 30:1. In addition, while the value for  $\phi$  in the south moat corresponds with the general plane of seismicity at the time, its volume of only  $0.002 \text{ km}^3$  is suspiciously small.  $\chi^2$  for this solution is 807, significantly higher than that of the Mogi source inversion.

### A.7 Conclusions

The fitness values,  $FV_s$  and  $FV_e$ , for both solutions, represent some value of  $\chi^2$  less than optimal.  $\chi^2_s$  is only 631, while  $\chi^2_e$  is worse at 807. This implies that there are additional sources which need to be modeled for this inflation episode, as suggested by Langbein et al, 1995. This is supported by error analyses performed on the Mogi source solution. Error estimation for a GA inversion is complicated by the nonlinear nature of the search, in conjunction with the nonlinear features of the parameter space itself. For this analysis, synthetic tests were repeated for each model, with only those solutions with a fitness value,  $FV$ , greater than an appropriate threshold, used in the final error calculations. As even the size of the initial parameter space chosen for the search affects the final outcome, at least over a finite number of generations, it was necessary that the synthetic tests be conducted using source models similar in nature and extent to that expected at Long Valley. In addition, the stability of each solution, and of particular parameters, was taken into account in the analysis. The correct solution is that with the highest fitness value, yet it can only be accepted as the global minimum, the best solution, if it can be reproduced over multiple runs as a stable solution. Therefore, only stable solutions with the highest  $FV$  were included in the actual error calculations.

For the synthetic tests, the larger, central source was located at 20 km in the x and y directions (north and east) respectively, with the depth at 12.5 km

**Table 2: Error in Sphere Parameters, Two Source Model**

Error in:	Larger Sphere, Depth of 12.5 km Volume = 0.075 km <sup>3</sup>	Smaller Sphere, Depth of 5 km Volume = 0.025 km <sup>3</sup>
X-location, m	57	36
Y-location, m	44	557
Depth, km	0.432	0.584
Volume, km <sup>3</sup>	0.00156	0.00055

Table A.2: Two-sphere error, Mogi source.

and a volume of 0.075 km<sup>3</sup>, both of which are consistent with various studies on Long Valley. The second, south moat source is expected to be smaller, so that the second source was located at 5 km in both the x and y directions, at 5 km depth and a volume of 0.025 km<sup>3</sup>. Over 30 GA inversions were performed on this synthetic data, for different initial conditions and varying total generations. Only those with an *FV* greater than 0.99975, and in which two or more parameter values were repeated together more than once, were included in the final calculations. Table A.2 shows the error on x, y, volume and depth for both spheres. It should be noted that the error on the location in the x and y direction for either source is insignificant in relation to the size of the caldera itself - 557 m in the y-direction for the smaller sphere is less than 2 percent of the size of the caldera.

Table A.2 also shows the error on the volume and depth of both spheres, and it is here that the most important feature of the error analysis is demonstrated. The deeper, yet significantly larger sphere dominates the deformation pattern at the surface of the earth, and, as a result, is much easier for the GA to locate in parameter space. The larger source is located at a depth of within 0.432 km and a volume of 0.00156 km<sup>3</sup>, corresponding to errors of only 3.4 and 2.1 percent

respectively. The smaller sphere, however, has deviations of 11.2 and 2.2 percent in the depth and volume.

The most important feature of this error analysis on the two Mogi source inversion is that it routinely produces  $\chi^2$  values of less than 200, significantly less than those above. This initially led to the conclusion that additional sources, particularly the normal faulting in the northwest of the caldera which was not modeled in this inversion, are necessary to provide a better solution, particularly for the more complicated, fourteen parameter ellipsoidal model.

Attempts at modeling the normal faulting in the north moat, in conjunction with the two ellipsoid model, proved relatively unsuccessful. While the displacements were on the order of those found by Langbein et al., 1997, the fitness values for the complete inversion was no better – on the order of 950 – while the results of the various trials remained unstable. In particular, the recurrence of solutions with large values of the semimajor axis is suspicious, given the large local minimum associated with that parameter, as shown above. It should be noted that this particular problem is not exclusive to the GA inversion and may be responsible for the large semimajor axis found by others [Langbein et al., 1997] in other inversions on the same data. These two issues, coupled with the success of the GA inversions performed on other geophysical problems (see, for example, Bhattacharyya et al., 1999 and Fernandez et al., 1999), leads me to conclude that other source models for the second, smaller source should be investigated.

The major contributions of this work are the following:

- (1) The development of a method for quantifying the error on a GA inversion, a procedure not found in the literature.
- (2) The implementation of windowing and proportional crossover to the GA program.

- (3) The determination that the size of the search space can affect the speed and accuracy of convergence in a GA inversion, and that this effect can be mitigated by varying both the population size and the type of crossover.
- (4) Derivation of the approximate substitution of volume for pressure in an ellipsoid.
- (5) The examination of the sensitivity of fit to the ellipsoid parameters.

#### A.8 Future work

Future work should include, in addition to inverting for other source models, attempts to model gravity change data in the manner of Rundle, 1982, in order to provide better constraints on both intrusion volumes and faulting mechanisms [Fung, 1965]. Finally, a third variation on the model which included one spheroidal and one ellipsoidal source in conjunction with the faulting mechanism should be incorporated into future studies to determine the validity of the significant increase in model complexity associated with the ellipsoidal sources.

## **Appendix B**

### **Publications & Grants**

- (1) NASA Mission to Planet Earth Grant, NGT5-30025, Study and Modeling of the Magmatic Intrusion Source at Long Valley Caldera, CA. 1996-1999.
- (2) Tiampo, K.F., Hofton, M., Rundle, J.B., and Minster, J.B. Inversion for the volcanic source at Long Valley, California, using a genetic algorithm technique. *Eos, Transactions, AGU*, v. 77, n. 46, p. 146, 1996.
- (3) Tiampo, K.F., Rundle, J.B., Hofton, M., Minster, J.B., and Langbein, J.O. Inversion for the volcanic source at Long Valley, California, using multiple data sources and a genetic algorithm technique. *Eos, Transactions, AGU*, v. 78, n. 46, p. 776, 1997.
- (4) Tiampo, K.F., Rundle, J.B., Fernandez, J.M., and Langbein, J.O. Ellipsoidal vs. spherical models for magmatic intrusion sources. *Annales Geophysicae*, v. 16, pt. I, EGS, p. 191, 1998.
- (5) Bhattacharyya, J., Sheehan, A.F., Tiampo, K. Genetic algorithm inversion of regional waveforms for western United States crustal structure. *Eos, Transactions, AGU*, v. 78, n. 46, p. 493, 1998.
- (6) Tiampo, K.F., Rundle, J.B., Gross, S., and Klein, W. Dynamics of seismicity patterns in complex fault systems. *Eos, Transactions, AGU*, v. 79, n. 45, p. 635, 1998.

- (7) Bhattacharyya, J., Sheehan, A.F., Tiampo, K., Rundle, J.B. Using genetic algorithms to model regional waveforms for crustal structure in the western United States. *BSSA*, v. 89, no. 1, Feb. 1999.
- (8) Tiampo, K.F., Rundle, J.B., W. Klein, K. Tiampo and S. Gross. Seismicity pattern dynamics for California fault systems, IUGG 99, Birmingham, England.
- (9) Tiampo, K.F., Rundle, J.B., Fernandez, J.M., and Langbein, J.O. Inversion for multiple volcanic sources using a genetic algorithm technique, IUGG 99, Birmingham, England.
- (10) Fernandez, J., Tiampo, K.F., Jentzsch, G., and Rundle, J.B. On the interpretation of the microgravity and GPS measurements in Mayon volcano, Philippines, IUGG 99, Birmingham, England.
- (11) CRES Graduate Research Fellowship. Pattern Dynamics Analysis of Southern California GPS Data, 1999-2000.
- (12) Rundle, JB, W. Klein, K. Tiampo and S. Gross. Linear pattern dynamics in nonlinear threshold systems, *Phys. Rev. E.*, March, 2000.
- (13) Tiampo, K.F., Rundle, J.B., Fernandez, J.M., and Langbein, J.O. Spherical and ellipsoidal volcanic sources at Long Valley caldera, California, using a genetic algorithm inversion technique, *Journal of Volcanology and Geophysical Research*, accepted for publication, Spring 2000.
- (14) Tiampo, K.F., Rundle, J.B., W. Klein, and Gross, S.J. Detecting systematic space-time variations in seismicity prior to large earthquakes, *Nature*, submitted March 2000.

- (15) Rundle, JB, W. Klein, K. Tiampo and S. Gross. Dynamics of seismicity patterns in systems of earthquake faults, Physics of Earthquakes, AGU Monograph, June, 2000.
- (16) Tiampo, K.F., Rundle, J.B., McGinnis, S., Klein, W., and Gross, S.J. Observation of systematic variations in non-local seismicity patterns from southern California. Physics of Earthquakes, AGU Monograph, June, 2000.
- (17) Tiampo, K.F., Rundle, J.B., McGinnis, S., Klein, W., and Gross, S.J. Nonlocal Space- Time Patterns on the California Fault System. EGS 2000.
- (18) Fernandez, J., Tiampo, K.F., Jentzsch, G., and Rundle, J.B. Elastic-gravitational modeling of magmatic sources, EGS 2000.

## Appendix C

### Seismicity Catalogs for Southern California

#### C.1 SCEC Database, 1932 - 1999

The following is a sample, year 1932 and year 1999 only (SCEC database - <http://www.scec.scec.org>).

Origin		Epicenter Location						
Date	Time	Lon.	Lat.	M	Quality			
1932/01/03	17:57:58.00	32.0333	-115.8333	0.00	3.0	0.00	0	0
1932/01/03	23:23:52.00	37.4500	-119.0000	0.00	2.0	0.00	0	0
1932/01/04	2:02:53.00	37.4500	-119.0000	0.00	2.0	0.00	0	0
1932/01/04	21:30:02.00	33.9000	-117.6500	0.00	2.0	0.00	0	0
1932/01/05	2:37:34.00	33.8667	-118.2833	0.00	1.5	0.00	0	0
1932/01/06	8:08:39.00	33.8833	-117.6333	0.00	2.0	0.00	0	0
1932/01/07	2:29:20.40	33.9000	-118.2000	0.00	1.0	0.00	0	0
1932/01/07	5:39:20.00	34.1833	-117.2833	0.00	1.0	0.00	0	0
1932/01/07	14:55:30.00	32.0333	-115.8333	0.00	3.0	0.00	0	0
1932/01/07	21:25:54.00	34.7500	-118.6667	0.00	2.0	0.00	0	0
1932/01/08	2:34:45.00	32.0333	-115.8333	0.00	4.0	0.00	0	0
1932/01/08	18:16:59.00	36.6667	-121.3333	0.00	4.0	0.00	0	0
1932/01/10	10:44:52.20	34.2500	-117.4167	0.00	3.0	0.00	0	0

1932/01/10 11:25:04.60 34.2667 -117.4333 0.00 2.0 0.00 0 0 C cit  
 1932/01/10 15:39:27.40 33.5333 -116.9333 0.00 3.0 0.00 0 0 B cit  
 1932/01/11 21:08:37.10 34.2500 -117.5833 0.00 2.5 0.00 0 0 D cit  
 1932/01/14 0:51:08.00 34.4333 -119.8667 0.00 3.0 0.00 0 0 C cit  
 1932/01/14 3:16:15.90 36.5833 -117.8167 0.00 3.0 0.00 0 0 B cit  
 1932/01/14 8:55:46.00 34.3667 -118.5833 0.00 3.0 0.00 0 0 C cit  
 1932/01/14 19:35:51.70 33.9167 -118.0333 0.00 3.0 0.00 0 0 A cit  
 1932/01/16 12:17:00.60 33.8833 -117.6333 0.00 1.5 0.00 0 0 C cit  
 1932/01/17 20:06:55.80 33.7167 -117.3000 0.00 2.0 0.00 0 0 C cit  
 1932/01/18 4:29:06.50 34.3000 -118.9167 0.00 2.0 0.00 0 0 D cit  
 1932/01/20 16:29:14.80 33.4667 -118.1167 0.00 2.5 0.00 0 0 B cit  
 1932/01/20 20:21:35.70 33.4167 -118.3333 0.00 3.5 0.00 0 0 B cit  
 1932/01/21 3:05:03.30 34.2167 -117.3833 0.00 1.5 0.00 0 0 B cit  
 1932/01/21 3:07:24.30 34.2167 -117.3833 0.00 3.0 0.00 0 0 B cit  
 1932/01/24 7:28:38.10 34.0167 -118.3000 0.00 2.0 0.00 0 0 B cit  
 1932/01/25 22:57:14.60 37.6167 -118.3167 0.00 3.0 0.00 0 0 B cit  
 1932/01/26 2:37:53.40 34.2333 -117.4167 0.00 3.0 0.00 0 0 B cit  
 1932/01/26 2:38:21.70 34.2333 -117.4167 0.00 3.0 0.00 0 0 B cit  
 1932/01/28 17:17:49.00 32.0333 -115.8333 0.00 4.5 0.00 0 0 C cit  
 1932/01/30 0:48:58.00 34.9167 -120.7333 0.00 3.0 0.00 0 0 C cit  
 1932/01/31 14:29:15.70 33.8833 -118.3167 0.00 3.5 0.00 0 0 B cit  
 1932/02/01 1:04:50.00 32.7833 -118.3333 0.00 3.0 0.00 0 0 C cit  
 1932/02/02 20:51:47.00 36.5667 -120.6000 0.00 3.5 0.00 0 0 C cit  
 1932/02/03 13:30:38.10 34.2833 -117.4833 0.00 2.5 0.00 0 0 B cit  
 1932/02/03 21:04:07.30 36.2500 -117.8000 0.00 2.5 0.00 0 0 B cit  
 1932/02/04 16:02:57.90 34.5500 -119.7333 0.00 3.0 0.00 0 0 C cit  
 1932/02/04 19:38:56.60 37.6667 -118.3833 0.00 3.0 0.00 0 0 C cit

1932/02/05	4:14:45.00	35.8333	-121.4667	0.00	3.5	0.00	0	0	C	cit
1932/02/05	6:46:54.00	35.8333	-121.4667	0.00	3.5	0.00	0	0	C	cit
1932/02/05	16:44:17.90	34.0500	-117.0333	0.00	3.0	0.00	0	0	B	cit
1932/02/06	11:33:32.70	33.8667	-118.2500	0.00	2.5	0.00	0	0	B	cit
1932/02/07	10:08:47.50	33.9333	-118.3833	0.00	3.0	0.00	0	0	B	cit
1932/02/07	21:40:12.10	33.9000	-118.2000	0.00	1.5	0.00	0	0	C	cit
1932/02/09	12:52:42.00	32.6333	-115.8000	0.00	3.0	0.00	0	0	C	cit
1932/02/11	2:57:15.00	33.9000	-118.2833	0.00	2.5	0.00	0	0	B	cit
1932/02/11	8:32:49.90	37.5167	-119.0000	0.00	2.5	0.00	0	0	C	cit
1932/02/11	23:11:20.00	34.4167	-116.8500	0.00	4.0	0.00	0	0	B	cit
1932/02/12	2:23:27.00	32.5000	-115.6667	0.00	3.5	0.00	0	0	C	cit
1932/02/12	2:30:21.00	32.5000	-115.6667	0.00	4.0	0.00	0	0	C	cit
1932/02/12	4:34:38.40	33.7833	-117.5500	0.00	2.0	0.00	0	0	B	cit
1932/02/14	17:55:06.80	34.9667	-119.0000	0.00	2.5	0.00	0	0	C	cit
1932/02/16	12:21:49.50	34.3667	-116.7833	0.00	2.5	0.00	0	0	C	cit
1932/02/17	6:46:04.00	34.5500	-119.7833	0.00	2.0	0.00	0	0	C	cit
1932/02/19	0:46:14.70	34.3833	-116.8667	0.00	2.0	0.00	0	0	C	cit
1932/02/19	5:39:25.00	34.5667	-119.8333	0.00	2.0	0.00	0	0	C	cit
1932/02/19	5:53:16.00	34.4167	-116.7500	0.00	2.5	0.00	0	0	C	cit
1932/02/19	6:39:43.00	34.5667	-119.8333	0.00	2.0	0.00	0	0	C	cit
1932/02/20	6:15:49.00	35.1333	-120.6667	0.00	3.0	0.00	0	0	C	cit
1932/02/21	4:41:15.00	36.3333	-120.5833	0.00	3.5	0.00	0	0	C	cit
1932/02/21	9:26:49.60	34.5500	-118.1500	0.00	2.5	0.00	0	0	B	cit
1932/02/21	12:20:28.00	32.5167	-115.6167	0.00	3.0	0.00	0	0	C	cit
1932/02/21	14:03:23.00	32.5167	-115.6167	0.00	3.0	0.00	0	0	C	cit
1932/02/23	20:24:49.40	36.2500	-117.8000	0.00	1.0	0.00	0	0	C	cit
1932/02/26	2:26:38.80	32.6000	-115.6167	0.00	3.5	0.00	0	0	B	cit

1932/02/26	3:43:19.80	32.6000 -115.6167	0.00 3.5 0.00	0 0	B cit
1932/03/01	7:10:56.00	33.4333 -116.8833	0.00 2.5 0.00	0 0	B cit
1932/03/06	1:02:32.00	36.2500 -120.5000	0.00 3.0 0.00	0 0	C cit
1932/03/06	2:38:27.60	34.3333 -118.4000	0.00 2.0 0.00	0 0	B cit
1932/03/06	21:00:17.10	37.0500 -117.5500	0.00 2.0 0.00	0 0	C cit
1932/03/07	11:31:00.00	36.3667 -120.5000	0.00 3.0 0.00	0 0	C cit
1932/03/07	12:34:27.10	34.3333 -118.4000	0.00 2.0 0.00	0 0	B cit
1932/03/09	7:02:26.90	34.1333 -117.9167	0.00 1.0 0.00	0 0	C cit
1932/03/13	5:51:41.80	32.4167 -116.3333	0.00 3.5 0.00	0 0	C cit
1932/03/13	13:37:28.90	33.8000 -117.3833	0.00 2.0 0.00	0 0	C cit
1932/03/13	23:09:23.70	34.4333 -120.1833	0.00 3.5 0.00	0 0	B cit
1932/03/21	10:45:00.00	33.9000 -118.2833	0.00 3.0 0.00	0 0	B cit
1932/03/21	23:11:57.00	33.8000 -118.5833	0.00 2.5 0.00	0 0	C cit
1932/03/23	0:20:14.10	35.6000 -116.9667	0.00 4.0 0.00	0 0	B cit
1932/03/24	8:17:20.00	33.2167 -116.1500	0.00 2.5 0.00	0 0	B cit
1932/03/25	6:57:31.90	34.2667 -118.7833	0.00 2.5 0.00	0 0	B cit
1932/03/26	4:42:47.00	33.2000 -116.0833	0.00 2.5 0.00	0 0	C cit
1932/03/26	6:57:44.00	33.0333 -115.9667	0.00 3.0 0.00	0 0	C cit
1932/03/27	9:00:17.00	33.6333 -116.7500	0.00 2.5 0.00	0 0	C cit
1932/03/29	3:16:03.30	34.3500 -117.0000	0.00 3.0 0.00	0 0	C cit
1932/03/31	5:00:19.60	36.2167 -117.9167	0.00 3.0 0.00	0 0	C cit
1932/03/31	18:42:10.00	34.8333 -119.3333	0.00 2.5 0.00	0 0	D cit
1932/03/31	23:46:15.10	34.1500 -117.3833	0.00 2.0 0.00	0 0	C cit
1932/04/01	3:04:10.60	36.2167 -117.9167	0.00 2.5 0.00	0 0	C cit
1932/04/01	7:06:35.60	34.1833 -116.3167	0.00 2.5 0.00	0 0	C cit
1932/04/02	2:21:44.00	32.3667 -116.5000	0.00 3.0 0.00	0 0	D cit
1932/04/06	9:02:33.90	34.3333 -118.4000	0.00 2.5 0.00	0 0	B cit

1932/04/06	14:57:45.10	33.2333	-116.7333	0.00	3.0	0.00	0	0	B	cit
1932/04/06	14:58:22.90	33.2333	-116.7333	0.00	2.5	0.00	0	0	B	cit
1932/04/06	22:42:00.00	34.5833	-120.7500	0.00	3.0	0.00	0	0	C	cit
1932/04/09	17:55:11.00	34.8167	-116.7000	0.00	3.0	0.00	0	0	C	cit
1932/04/12	10:08:47.30	34.0667	-117.0000	0.00	2.0	0.00	0	0	C	cit
1932/04/12	11:18:58.50	34.0833	-118.0667	0.00	2.5	0.00	0	0	B	cit

---

1999/08/27	02:17:38.63	35.72	-118.48	5.50	1.50
1999/08/27	02:20:28.67	35.76	-118.46	1.32	1.35
1999/08/27	03:51:33.71	36.01	-117.81	1.14	1.17
1999/08/27	04:12:29.54	36.25	-120.82	9.10	1.21
1999/08/27	04:52:12.06	35.74	-118.47	4.94	1.20
1999/08/27	06:13:18.79	34.96	-117.03	7.61	1.91
1999/08/27	09:50:30.16	33.45	-116.57	6.00	1.60
1999/08/27	10:05:19.00	34.48	-116.51	3.20	1.26
1999/08/27	10:48:04.09	35.47	-118.45	8.43	2.01
1999/08/27	13:26:12.05	33.47	-116.43	15.67	0.70
1999/08/27	13:46:36.12	35.97	-117.62	2.26	1.26
1999/08/27	14:18:43.57	34.34	-116.47	2.71	1.51
1999/08/27	14:48:48.52	34.33	-116.47	5.98	1.46
1999/08/27	17:04:29.56	35.81	-118.47	9.77	2.30
1999/08/27	17:44:06.59	35.72	-118.48	5.55	1.31
1999/08/27	18:20:53.30	34.35	-116.84	0.00	1.76
1999/08/27	19:31:53.31	35.93	-119.27	6.00	2.28
1999/08/27	19:34:15.35	32.59	-116.15	6.86	1.76
1999/08/27	20:54:42.01	34.39	-118.60	9.64	1.76
1999/08/27	21:03:22.12	32.83	-116.07	10.71	1.91

1999/08/27 21:08:32.20 35.70 -118.49 8.21 1.51  
1999/08/27 22:35:21.00 34.24 -118.42 6.00 1.54  
1999/08/28 00:25:17.70 32.85 -116.17 3.93 2.22  
1999/08/28 00:26:49.33 33.98 -117.18 18.26 1.95  
1999/08/28 00:32:15.79 32.85 -116.17 4.61 2.10  
1999/08/28 00:38:06.14 32.84 -116.17 4.05 2.03  
1999/08/28 00:48:46.70 34.02 -116.62 8.07 2.95  
1999/08/28 03:11:13.87 35.45 -118.42 6.00 1.47  
1999/08/28 03:58:37.96 34.37 -116.47 3.53 1.63  
1999/08/28 06:22:35.85 35.67 -118.09 10.78 1.25  
1999/08/28 07:00:55.90 33.30 -116.76 11.10 2.78  
1999/08/28 07:02:18.34 33.29 -116.75 12.19 2.90  
1999/08/28 07:04:56.11 33.30 -116.76 6.00 1.86  
1999/08/28 07:38:46.00 36.01 -117.72 1.01 1.88  
1999/08/28 09:20:40.38 34.41 -116.48 3.42 1.77  
1999/08/28 09:53:02.63 34.37 -116.47 2.67 1.97  
1999/08/28 09:55:58.02 36.12 -120.13 8.38 1.50  
1999/08/28 10:45:55.87 36.04 -117.67 1.83 1.21  
1999/08/28 10:59:23.21 32.70 -115.54 0.00 1.30  
1999/08/28 11:13:02.93 36.03 -117.84 2.91 0.88  
1999/08/28 12:45:44.10 34.84 -118.94 10.86 1.30  
1999/08/28 13:20:39.26 34.32 -116.84 6.30 1.30  
1999/08/28 13:30:33.75 33.84 -118.48 8.73 2.40  
1999/08/28 13:38:46.17 35.72 -118.47 4.67 1.51  
1999/08/28 15:09:20.00 34.24 -118.64 10.52 0.94  
1999/08/28 15:32:50.20 36.03 -117.83 1.49 1.31  
1999/08/28 15:37:20.82 36.03 -117.83 2.78 2.33

1999/08/28 15:43:09.41 36.03 -117.83 2.02 1.79  
1999/08/28 15:45:40.89 36.02 -117.88 5.17 1.33  
1999/08/28 15:47:58.01 34.09 -117.33 16.37 1.89  
1999/08/28 16:48:53.44 36.03 -117.83 2.98 1.78  
1999/08/28 16:49:49.85 36.03 -117.84 2.59 1.23  
1999/08/28 17:06:07.42 36.03 -117.83 3.14 1.51  
1999/08/28 18:57:46.21 33.72 -116.73 17.57 1.67  
1999/08/28 19:27:41.28 34.37 -116.49 0.00 1.76  
1999/08/28 20:06:16.57 36.01 -117.81 1.33 1.10  
1999/08/28 22:46:52.82 34.05 -116.31 9.47 1.31  
1999/08/28 23:47:01.83 35.76 -118.46 0.98 1.52  
1999/08/29 01:32:36.57 33.49 -116.46 6.00 1.50  
1999/08/29 03:39:27.20 33.27 -116.01 0.21 2.12  
1999/08/29 04:02:23.95 34.00 -116.35 0.01 1.62  
1999/08/29 04:37:27.54 34.51 -118.64 21.34 1.25  
1999/08/29 05:59:07.42 36.05 -120.60 4.85 1.74  
1999/08/29 06:36:33.22 35.81 -121.30 7.98 1.35  
1999/08/29 07:07:35.37 35.32 -117.65 8.64 1.29  
1999/08/29 07:44:04.13 36.01 -117.81 1.03 1.01  
1999/08/29 07:44:26.16 36.01 -117.81 2.56 2.10  
1999/08/29 07:44:38.52 36.01 -117.81 2.34 1.51  
1999/08/29 08:18:25.61 35.75 -118.48 3.32 2.20  
1999/08/29 08:31:57.65 33.33 -116.89 17.90 1.18  
1999/08/29 08:57:36.65 36.03 -117.84 6.00 1.09  
1999/08/29 10:55:17.56 36.05 -120.60 3.96 1.05  
1999/08/29 13:30:04.42 35.30 -117.63 8.28 0.80  
1999/08/29 14:22:58.03 36.03 -117.83 3.40 1.18

1999/08/29 14:25:06.71 36.02 -117.84 3.40 0.67  
1999/08/29 15:59:44.19 34.14 -116.81 0.88 0.41  
1999/08/29 17:10:52.00 34.62 -116.55 8.76 1.62  
1999/08/29 17:28:01.29 35.72 -118.47 4.25 2.32  
1999/08/29 17:37:40.48 36.03 -117.85 2.93 1.14  
1999/08/29 17:43:48.48 36.22 -120.84 5.68 1.17  
1999/08/29 17:47:44.02 34.19 -116.82 6.00 1.17  
1999/08/29 18:02:58.28 36.01 -117.81 1.26 1.29  
1999/08/29 19:22:20.51 35.74 -118.47 4.05 1.51  
1999/08/29 20:57:23.13 36.03 -117.77 1.75 1.20  
1999/08/29 23:03:58.12 33.71 -116.83 18.07 1.21  
1999/08/29 23:49:28.43 34.30 -118.48 13.30 1.21  
1999/08/30 02:02:01.64 34.19 -118.52 19.47 1.95  
1999/08/30 04:30:44.98 33.39 -116.80 6.00 1.51  
1999/08/30 05:29:38.62 33.48 -116.50 6.00 1.84  
1999/08/30 07:50:59.93 35.73 -118.47 4.69 2.16  
1999/08/30 07:51:50.99 35.73 -118.47 4.94 1.64  
1999/08/30 08:56:35.63 35.73 -118.47 4.51 1.83  
1999/08/30 09:01:30.17 33.34 -116.38 6.00 2.01  
1999/08/30 12:27:32.78 33.50 -116.49 6.00 2.53  
1999/08/30 14:54:55.49 36.17 -118.04 6.00 1.51  
1999/08/30 16:15:32.66 32.87 -117.72 6.00 2.20  
1999/08/30 20:36:43.95 33.23 -117.25 18.60 1.73  
1999/08/30 21:41:34.00 34.30 -118.44 6.00 1.76  
1999/08/30 23:17:57.76 35.74 -118.47 4.57 1.71

## C.2 Catalog of Large Events, M > 6, 1812-1992.

Source - <http://www.gps.caltech.edu/~jishu>.

Source Parameters of Magnitude 6 and Larger Earthquakes in Southern California

Origin Year	Epicenter Date	D, LQ Time Lon. Lat. km	Location or Felt Area Magnitude( <i>s</i> )
1812 12 08	1500 34.37 117.65	0.0	a Wrightwood.....
1812 12 21	1900 34.20 119.90	0.0	c Santa Barbara Channel.....
1852 11 29	2000 32.50 115.00	0.0	c Volcano Lake, B.C.....
1855 07 11	0415 34.10 118.10	0.0	c Los Angeles region.....
1857 01 09	0812 35.30 119.80	0.0	a Great Fort Tejon Earthquake.....
1858 12 16	1000 34.00 117.50	0.0	b San Bernardino region.....
1862 05 27	2000 32.70 117.20	0.0	c San Diego Region.....
1872 03 26	1030 36.70 118.10	0.0	a Owens Valley.....
1872 03 26	1406 36.90 118.20	0.0	c Owens Valley.....
1872 04 03	1215 37.00 118.20	0.0	c Owens Valley.....
1875 11 15	2230 32.50 115.50	0.0	d Imperial Valley to Colorado River delta
1883 09 05	1230 34.20 119.90	0.0	c Santa Barbara Channel.....
1885 04 12	0405 36.57 120.65	0.0	c Southern Diablo Range.....
1890 02 09	1206 33.40 116.30	0.0	c San Jacinto or Elsinore fault region(?)
1890 04 24	1136 36.90 121.60	0.0	a Pajaro Gap.....
1892 02 24	0720 32.55 115.63	0.0	b Laguna Salada, B.C.....
1892 05 28	1115 33.20 116.20	0.0	c San Jacinto or Elsinore fault region(?)
1894 07 30	0512 34.30 117.60	0.0	c Lytle Creek region.....

1897 06 20 2014 37.00 121.50 0.0 a Gilroy.....  
 1899 07 22 2032 34.30 117.50 0.0 b Lytle Creek region.....  
 1899 12 25 1225 33.80 117.00 0.0 a San Jacinto and Hemet..... 6.4  
 1901 03 03 0745 36.00 120.50 0.0 x Parkfield..... 6.4  
 1906 04 19 0030 32.90 115.50 0.0 x Imperial Valley..... 6.2  
 1907 09 20 0154 34.20 117.10 0.0 x San Bernardino region.....  
 1908 11 04 0837 36.00 117.00 0.0 x Death Valley region.....  
 1910 05 15 1547 33.70 117.40 0.0 x Glen Ivy Hot Springs.....  
 1915 06 23 0359 32.80 115.50 0.0 x Imperial Valley.....  
 1915 06 23 0456 32.80 115.50 0.0 x Imperial Valley.....  
 1915 11 21 0013 32.00 115.00 9.0 x Volcano Lake, B.C..... 6.6  
 1916 10 23 0244 34.90 118.90 0.0 x Tejon Pass region.....  
 1916 11 10 0911 36.00 117.00 0.0 x South of Death Valley..... 5.5  
 1918 04 21 2232 33.80 117.00 7.0 B San Jacinto..... 6.8 6.8  
 1922 03 10 1121 36.00 120.50 0.0 x Parkfield..... 6.3  
 1923 07 23 0730 34.00 117.30 12.0 A San Bernardino region..... 6.3 6.3  
 1925 06 29 1442 34.30 119.80 0.0 x Santa Barbara.....  
 1927 11 04 1350 34.35 120.90 10.0 a Southwest of Lompoc..... 7.0 6.6  
 1933 03 11 0154 33.66 117.97 13.0 A Long Beach..... 6.4 6.4  
 1934 06 08 0447 35.80 120.33 0.0 B Parkfield..... 6.4 6.1  
 1934 12 30 1352 32.25 115.50 10.0 D Laguna Salada, B.C..... 6.5 6.5  
 1934 12 31 1845 32.00 114.75 12.0 D Colorado River delta..... 7.0 7.1  
 1937 03 25 1649 33.47 116.42 3.0 B Buck Ridge..... 5.6 6.0  
 1940 05 19 0436 32.87 115.48 7.0 B Imperial Valley..... 7.2 6.8 6.9  
 1942 10 21 1622 33.00 116.03 7.0 B Fish Creek Mountains..... 6.3 6.6  
 1946 03 15 1349 35.73 118.05 22.0 A Walker Pass..... 6.0  
 1947 04 10 1558 34.98 116.55 5.0 A Manix..... 6.5 6.5

1948 12 04 2343 33.92 116.48 12.0 A Desert Hot Springs..... 5.9 6.0  
1952 07 21 1152 35.00 119.02 16.0 A Kern County earthquake..... 7.8 7.5  
1952 07 21 1205 35.00 119.00 0.0 D Kern County..... 6.4  
1952 07 23 0038 35.37 118.58 0.0 A Kern County..... 6.1  
1952 07 29 0703 35.38 118.85 0.0 A Bakersfield..... 6.1  
1952 11 22 0746 35.77 121.15 10.0 B Bryson..... 6.0  
1954 03 19 0954 33.29 116.18 9.0 B Arroyo Salada..... 6.1 6.4  
1966 06 28 0426 35.92 120.53 18.6 B Parkfield..... 6.4 5.8 6.1 5.6  
1968 04 09 0228 33.19 116.13 10.0 B Borrego Mountain..... 6.8 6.7 6.5  
1971 02 09 1400 34.41 118.40 8.4 B San Fernando..... 6.7 6.6  
1979 10 15 2316 32.61 115.32 12.3 B Imperial Valley..... 6.4 6.4  
1980 06 09 0328 32.19 115.08 5.0 D Victoria, B.C..... 6.4 6.1  
1983 05 02 2342 36.25 120.26 9.5 C Coalinga..... 6.5 6.2 6.2 6.3  
1987 11 24 0154 33.08 115.78 8.0 A Elmore Ranch fault.... 6.2 5.7 6.2 6.2  
1987 11 24 1315 33.01 115.84 10.0 A Superstition Hills..... 6.6 5.9 6.6 6.6  
1992 04 23 0450 33.96 116.32 12.4 A Joshua Tree..... 6.3 5.6 6.1 6.1  
1992 06 28 1157 34.20 116.44 8.0 A Landers..... 7.5 7.3 7.4  
1992 06 28 1505 34.20 116.83 5.3 C Big Bear..... 6.6 6.0 6.5 6.4  
1994 01 17 1230 34.21 118.54 18.4 A Northridge..... 6.7 6.7

### C.3 Catalog of Large Events, M > 5, 1933 - 1994.

Source - <http://www.gps.caltech.edu/~jishu>.

Table A1. Source Parameters of M  $\geq 5$  Earthquakes in Southern California  
(32deg-37deg)

Origin	Epicenter	D, LQ	Location or Felt Area			
Year	Date	Time	Lon.	Lat.	km	Magnitude(s)
1933	03 11	0154	117.97	33.66	13.0	A Long.Beach..... 6.4 6.4
1933	10 02	0910	118.13	33.78	0.0	A ..... 5.4
1934	06 05	2148	120.33	35.80	0.0	B Parkfield..... 5.0
1934	06 08	0430	120.33	35.80	0.0	B Parkfield..... 5.0
1934	06 08	0447	120.33	35.80	0.0	B Parkfield..... 6.4 6.1
1934	11 25	0818	116.67	32.08	0.0	C ..... 5.0
1934	12 24	1626	120.48	35.93	0.0	B ..... 5.0
1934	12 30	1352	115.50	32.25	10.0	D Laguna.Salada,,B.C..... 6.5 6.5
1934	12 31	1845	114.75	32.00	12.0	D Colorado.River.Delta..... 7.0 7.1
1935	10 24	1448	116.80	34.10	0.0	A ..... 5.1
1935	12 20	0745	115.50	33.17	0.0	C felt.in.Imperial.Valley..... 5.2
1937	03 25	1649	116.41	33.47	3.0	B Buck.Ridge..... 5.6 6.0
1938	05 31	0834	117.51	33.70	10.0	B ..... 5.2
1938	09 27	1223	120.90	36.30	0.0	C felt.in.San.Benito.&.Mont.Count. 5.0
1939	05 01	2353	117.50	32.00	0.0	B felt.in.San.Diego..... 5.0
1939	05 04	2044	114.82	35.97	0.0	A near.Boulder.City..... 5.0
1939	06 24	1301	121.00	36.40	0.0	C damage.in.San.Benito.County..... 5.5
1939	06 24	1627	117.50	32.00	0.0	B felt.in.San.Diego..... 5.0

1940 03 10 1801 115.00 37.00 0.0 D Nevada..... 5.0  
 1940 05 18 0503 116.30 34.08 0.0 A felt.over.a wide.area..... 5.3  
 1940 05 19 0436 115.48 32.87 7.0 A Imperial.Valley..... 7.2 7.0 6.9  
 1941 07 01 0750 119.58 34.37 0.0 A Carpenteria..... 5.5  
 1941 09 21 1953 118.93 34.87 0.0 A Gorman..... 5.1  
 1942 03 03 0103 115.75 34.00 0.0 C felt.at.Twentynine.Palms..... 5.0  
 1942 05 23 1547 115.98 32.98 0.0 C felt.in.Imperial.Valley..... 5.1  
 1942 10 21 1622 116.03 32.97 7.0 B Fish.Creek.Mountains..... 6.3 6.6  
 1942 10 22 0150 115.72 33.23 0.0 C Salton.Sea..... 6.1 5.5  
 1943 08 29 0345 116.97 34.27 0.0 C San.Bernardino.Mountains..... 5.3  
 1943 12 22 1550 115.80 34.33 0.0 B Bullion.Mountains,Mojave.Desert 5.3  
 1944 06 12 1045 116.72 33.98 10.0 A San.Gorgonio.Pass..... 5.0  
 1944 06 12 1116 116.71 33.99 10.0 A San.Gorgonio.Pass..... 5.2  
 1945 04 01 2343 120.02 34.00 0.0 C Santa.Rosa.Island..... 5.1  
 1945 08 15 1756 116.13 33.22 0.0 B felt.widely..... 5.7  
 1946 01 08 1854 115.83 33.00 0.0 C ..... 5.4  
 1946 03 15 1321 117.99 35.75 0.0 A felt.in.Pasadena..... 5.5  
 1946 03 15 1349 118.05 35.73 22.0 A Walker.Pass..... 6.0 6.0  
 1946 07 18 1427 115.98 34.53 0.0 B felt.in.Mojave.Desert..... 5.5  
 1947 04 10 1558 116.55 34.98 5.0 A Manix..... 6.5 6.5  
 1947 07 24 2210 116.50 34.02 0.0 A ..... 5.3  
 1947 11 18 2159 119.45 33.27 0.0 B San.Nicolas.Island..... 5.0  
 1948 02 24 0815 118.55 32.50 0.0 C felt.at.San.Diego..... 5.3  
 1948 12 04 2343 116.48 33.92 12.0 A Desert.Hot.Springs..... 6.2 6.3  
 1949 03 09 1228 121.48 36.02 0.0 B VII.at.Hollister..... 5.2  
 1949 05 02 1125 115.68 34.02 0.0 C Pinto.Mountain..... 5.8  
 1949 11 04 2042 116.55 32.20 0.0 C Guadalupe,B.C..... 5.7

1950 07 28 1750 115.57 33.12 0.0 C northern.Imperial.Valley..... 5.4  
 1950 07 29 1436 115.57 33.12 0.0 C Calipatria..... 5.5  
 1951 01 24 0717 115.73 32.98 0.0 B near.Calipatria..... 5.8  
 1951 07 29 1053 121.18 36.58 0.0 B Hollister..... 5.0  
 1951 12 26 0046 118.35 32.82 0.0 B San.Clemente.Island..... 5.9  
 1952 07 21 1152 119.02 35.00 16.0 A Great.Kern.County..... 7.8 7.5 7.5  
 1952 07 29 0703 118.85 35.38 0.0 A Bakersfield..... 6.3 6.1  
 1952 08 23 1009 118.20 34.52 13.1 A Acton..... 5.1  
 1952 11 22 0746 121.15 35.77 10.0 B Bryson..... 6.0  
 1953 06 14 0417 115.72 32.95 0.0 B Imperial..... 5.5  
 1954 01 12 2333 119.02 35.00 0.0 A Wheeler.Ridge..... 5.6  
 1954 01 27 1419 118.63 35.15 0.0 B Tehachapi..... 5.0  
 1954 02 01 0423 115.30 32.30 0.0 D felt.in.Imperial.Valley..... 5.2  
 1954 02 01 0432 115.30 32.30 0.0 D felt.in.Imperial.Valley..... 5.6  
 1954 03 19 0954 116.18 33.29 9.0 B Arroyo.Salada..... 6.1 6.4  
 1954 04 25 2033 121.68 36.93 0.0 B Watsonville..... 5.3  
 1954 05 23 2352 118.98 34.98 0.0 A ..... 5.1  
 1955 04 25 1043 115.00 32.33 0.0 D ..... 5.2  
 1955 11 02 1940 120.92 36.00 0.0 A ..... 5.2  
 1955 12 17 0607 115.50 33.00 0.0 B ..... 5.2  
 1956 11 16 0323 120.47 35.95 0.0 B felt.in.King.City..... 5.0  
 1957 04 25 2157 115.81 33.22 0.3 B felt.in.Boulevard..... 5.2  
 1957 05 26 1559 116.00 33.23 15.1 A felt.in.Imperial.Valley..... 5.0  
 1958 12 01 0321 115.75 32.25 0.0 B ..... 5.8  
 1959 03 02 2357 121.58 36.98 0.0 B ..... 5.3  
 1960 01 20 0325 121.43 36.78 0.0 B Hollister..... 5.0  
 1961 01 28 0812 118.05 35.78 5.5 A Kernville..... 5.3

1961 04 09 0723 121.30 36.68 0.0 A Hollister..... 5.6  
 1961 10 19 0509 117.76 35.83 2.0 A China.Lake..... 5.4  
 1962 09 16 0536 118.04 35.75 3.8 B Walker.Pass..... 5.0  
 1962 10 29 0242 116.87 34.33 8.6 B Big.Bear..... 5.0  
 1963 09 14 1946 121.60 36.89 3.4 A Chittenden-Watsonville..... 5.4  
 1963 09 23 1441 116.93 33.71 16.5 B Hemet..... 5.1  
 1965 09 25 1743 116.50 34.71 10.6 B Camp.Cady.Ranch..... 5.2  
 1965 09 26 0700 116.03 34.71 8.3 C Kelso..... 5.1  
 1966 06 28 0426 120.53 35.92 18.6 B Parkfield..... 6.4 5.8 6.4 5.6  
 1968 04 09 0228 116.13 33.19 10.0 B Borrego.Mountain..... 6.8 6.7 6.5  
 1968 07 05 0045 119.70 34.12 5.9 B Santa.Barbara.Channel..... 5.3  
 1969 01 23 2301 116.04 33.89 17.7 B Indio..... 5.0  
 1969 04 28 2320 116.35 33.34 12.0 B Coyote.Creek..... 5.6 5.8  
 1969 10 22 2251 121.62 34.58 10.0 B Lompoc..... 5.4  
 1969 10 24 0829 119.19 33.29 10.0 B ..... 5.1  
 1969 10 27 1059 121.34 36.90 10.0 B Hollister..... 5.0  
 1969 11 05 1754 121.44 34.61 10.0 B Avila.Beach..... 5.6  
 1970 09 12 1430 117.54 34.27 8.0 A Lytle.Creek..... 5.3 5.2  
 1971 02 09 1400 118.40 34.41 8.4 B San.Fernando..... 6.5 6.7 6.6  
 1971 09 30 2246 115.82 33.03 8.0 B Superstition..... 5.0  
 1972 02 24 1556 121.18 36.62 8.0 C Bear.Valley..... 5.0  
 1973 02 21 1445 119.04 34.06 8.0 B Point.Mugu..... 5.2 5.7 5.5 5.3  
 1973 08 06 2329 119.48 33.99 16.9 A ..... 5.0  
 1974 11 28 2301 121.61 36.90 7.0 D Hollister..... 5.0  
 1975 06 01 0138 116.50 34.52 4.5 A Galway.Lake..... 5.0  
 1976 01 10 1258 115.47 32.08 12.3 C ..... 5.0  
 1976 11 04 1041 115.62 33.13 0.0 X ..... 5.0

1978 05 05 2103 115.30 32.21 6.0 C ..... 5.2  
 1978 08 13 2254 119.68 34.40 12.7 B Santa.Barbara..... 5.8 5.1  
 1979 01 01 2314 118.69 33.95 9.6 B Malibu..... 5.2  
 1979 03 15 2107 116.44 34.33 2.5 A Homestead.Valley..... 5.5 5.3  
 1979 10 15 2316 115.32 32.61 12.3 B Imperial.Valley..... 6.4 6.4  
 1980 02 25 1047 116.51 33.50 13.6 A Anza..... 4.9 5.5  
 1980 06 09 0328 115.08 32.19 5.0 D Victoria,.B.C..... 6.4 6.4 6.1  
 1981 04 26 1209 115.63 33.10 3.8 C Westmorland..... 6.0 5.9 5.7  
 1981 09 04 1550 119.10 33.66 6.0 C Santa.Barbara.Island..... 5.8 5.5  
 1982 10 25 2226 120.40 36.29 6.0 C New.Idria..... 5.4 5.5  
 1983 05 02 2342 120.26 36.25 9.5 C Coalinga..... 6.5 6.2 6.2 6.3  
 1983 08 29 1010 122.46 35.91 0.0 D ..... 5.2  
 1985 08 04 1201 120.05 36.15 6.0 C North.Kettleman.Hills..... 5.9 6.1 5.8  
 1986 07 08 0920 116.61 34.00 10.4 A North.Palm.Springs..... 6.0 6.1 5.6  
 1986 07 13 1347 117.86 32.98 6.0 C Oceanside..... 5.3 5.4  
 1987 02 07 0345 115.30 32.39 6.0 C ..... 5.4  
 1987 10 01 1442 118.08 34.05 14.6 A Whittier.Narrows..... 6.0 5.9  
 1987 10 04 1059 118.10 34.06 13.3 A Whittier.Narrows..... 5.3  
 1987 11 24 0154 115.78 33.08 8.0 A Elmore.Ranch.fault..... 6.2 5.7 6.2 6.2  
 1987 11 24 1315 115.84 33.01 10.0 A Superstition.Hills.fault. 6.6 5.9 6.6 6.6  
 1988 06 10 2306 118.74 34.94 6.8 A Tejon.Ranch..... 5.4  
 1988 12 03 1138 118.14 34.15 15.4 A Pasadena..... 4.9 5.0  
 1989 01 19 0653 118.63 33.92 7.5 A Malibu..... 5.0  
 1990 02 28 2343 117.70 34.14 6.0 A Upland..... 5.5 5.3  
 1991 06 28 1443 118.00 34.26 12.5 A Sierra.Madre..... 5.6 5.4  
 1992 04 23 0450 116.32 33.96 12.4 A Joshua.Tree..... 6.3 5.6 6.1 6.1  
 1992 06 28 1157 116.44 34.20 8.0 A Landers..... 7.5 7.3 7.4

1992 06 28 1505 116.83 34.20	5.3	C	Big Bear.....	6.6	6.0	6.5	6.4
1992 07 05 2118 116.32 34.58	0.1	A .....		5.4			
1992 07 11 1814 118.07 35.21	10.7	A .....		5.1	5.7		
1992 11 27 1600 116.90 34.34	1.5	A .....		5.4			
1993 05 28 0447 119.10 35.15	21.4	A	Wheeler.Ridge.....	5.2			
1993 08 21 0146 116.32 34.03	9.1	A .....		5.0			
1994 01 17 1230 118.54 34.21	18.4	A	Northridge.....	6.7	6.7		
1994 06 16 1624 116.40 34.27	3.4	A .....		5.0			
1995 06 26 0840 118.67 34.39	13.3	A .....		5.0			
1995 08 17 2239 117.66 35.78	5.6	A .....		5.4			
1995 09 20 2327 117.63 35.76	5.4	A	China Lake - Ridgecrest area.....	5.5			
1995 09 20 2327 117.64 35.76	5.4	A	China Lake - Ridgecrest area.....	5.8			

LAST UPDATE: Sept. 16, 1997.

## Appendix D

### KL Decomposition Programs

The programs for the KL decomposition can be broken down into three major categories - the idl programs which sort and bin the catalog data into locations and time series, the fortran program which performs the eigenvalue decomposition, outputting eigenvectors and principal component time series, and the idl programs which normalized those eigenvectors, outputting them to data files which GMT can use for mapping, and pc time series. I also included a sample of the GMT program which creates the eigenvector maps.

#### D.1 Data Sorting Programs

```
; read_all.pro - idl program, reads in the catalog data
allrecords = dblarr(13,3198)
strall=strarr(3198)
temp=dblarr(13)
stemp=""
openr,88,"..//newcit.ca"
for n = 0L, 33197L do begin
readf,88,format='(I4,I2,I2,I4,I2,F5.2,F9.4,F10.4,F7.2,F4.2,F5.2,I3,I7,A2),
&temp,stemp
strall(n)=strcompress(stemp,/remove_all)
```

```
allrecords(*,n)=temp
```

```
endfor
```

```
close,88
```

```
end
```

```
; record_all.pro - idl program, sorts catalog data into necessary information
```

```
m=0L
```

```
all98=dblarr(8,3198)
```

```
for n = 0L, 3197L do begin
```

```
if (strall[n] eq 'A' or strall[n] eq 'B') then begin
```

```
all98[* ,m]=allrecords[0:7,n]
```

```
;print,all98[* ,m]
```

```
m=m+1
```

```
endif
```

```
endfor
```

```
print,m
```

```
end
```

```
; lon_all.pro - takes catalog data from above, calculates time
```

```
; steps, then figures out how many in the locations of interest
```

```
; at each time.
```

```
;restore,"idlall3.dat"
```

```

;help
;total1=dblarr(8,245889)
;total1=alldata
p = m - 1
alldata=dblarr(m)
alldata=all98[*,0:p]
alldata(2,*)=alldata(2,*)-1.0
day=dblarr(1,m)
day(0)=alldata(2,0)+(alldata(3,0)+(alldata(4,0)+alldata(5,0)/60.)/60.)/24.
;print,day(0)
day(*)=alldata(2,*)+(alldata(3,*)+(alldata(4,*)+alldata(5,*)/60.)/60.)/24.
month = dblarr(m)
FOR k = 0L, p DO BEGIN
  IF (alldata(1,k) EQ 9.) then n=30.
  IF (alldata(1,k) EQ 4.) then n=30.
  IF (alldata(1,k) EQ 6.) then n=30.
  IF (alldata(1,k) EQ 11.) then n=30.
  IF (alldata(1,k) EQ 2. ) then n=28.
  IF (alldata(1,k) EQ 1.) then n=31.
  IF (alldata(1,k) EQ 3.) then n=31.
  IF (alldata(1,k) EQ 5.) then n=31.
  IF (alldata(1,k) EQ 7.) then n=31.
  IF (alldata(1,k) EQ 8.) then n=31.
  IF (alldata(1,k) EQ 10.) then n=31.
  IF (alldata(1,k) EQ 12.) then n=31.
  alldata(1,k)=alldata(1,k)-1.0
  month(k)=alldata(1,k)+day(k)/n

```

```

ENDFOR

p1=p-10

print,month(p1:p)

yr=dblarr(1,m)

yr(*)=alldata(0,*)+month(*)/12.

print,yr(0:10)

print,yr(p-10:p)

yr=(double(round(yr*1000.)))/1000.

print,yr(0:10)

print,yr(p-10:p)

s = 0

corrall=dblarr(3,m)

corrall(0,*)=yr(*)

;print,corrall(0)

print,corrall(0,p)

lat=dblarr(1,m)

lat(*)=alldata(6,*)

lon=dblarr(1,m)

lon(*)=alldata(7,*)

print,lat(0),lon(0)

print,lat(p),lon(p)

corrall(1,*)=lat(*)

corrall(2,*)=lon(*)

;print,corr1977

openr,33,"filenames.dat"

lat=32.5

count=indgen(37)

```

```
names=strarr(37)
readf,33,names
close,33
print,names
for n = 0, 36 do begin
;print,n,names(n)
openw,11,names(n)
lat1=lat + 0.1
B=where(corrall(1,*) ge lat and corrall(1,*) lt lat1,C)
if (C gt 0) then values=corrall(*,B)
count(n)=C
printf,11,C
if (C gt 0) then printf,11,values
lat=lat1
close,11
endfor
openr,33,"filenames.dat"
openr,22,"longnames.dat"
lonnames=strarr(61)
names=strarr(37)
loncount = indgen(2257)
newnames=strarr(2257)
openw,55,"newnames.dat"
readf,33,names
readf,22,lonnames
close,33
close,22
```

```

;print,names
;print,lonnames
i = 0
for n = 0, 36 do begin
  lon = -121.
  ;print,n,names(n)
  openr,11,names(n)
  readf,11,z
  if (z gt 0) then lonvalues=dblarr(3,z)
  if (z lt 0) then readf,11,lonvalues
  close,11
  for j = 0, 60 do begin
    lon1=lon + 0.1
    lonlat=names(n)+lonnames(j)
    openw,44,lonlat
    printf,55,lonlat
    if (count(n) eq 0) then E = 0
    if (count(n) gt 0) then D=where(lonvalues(2,*) ge lon and lonvalues(2,*) lt
    lon1,E)
    if (E eq 0) then begin
      printf,44,E
    endif
    if (count(n) gt 0 and E gt 0) then begin
      values1=lonvalues(0,D)
      ;      print,"val=",values1
      values1=values1(uniq(values1))
      ;      print,"unq val=",values1
    end
  end
end

```

```

s=s+1

G=where(values1 gt 0,H)

;      print,"H=",H
;      print,"uval+",values1
printf,44,H
printf,44,values1

endif

close,44

lon = lon1

i = i + 1

endfor

lat=lat1

close,11

endfor

close,55

print,s

;save,filename="idl_all.dat",/all

end

; series_all.pro - creates time series and puts them into a matrix
; for the eof calculations.

series = dindgen(1,66500L)
seriesname = strarr(2257)
count1=0

```

```

;print,series[0,0:10]
series = 1932. + series/1000.
openw,88,"series.dat"
printf,88,series
close,88

;print,series[0,0:10]
;print,series[0,66440:66449]
timemat=fltarr(s,66500)
openr,55,"newnames.dat"
readf,55,newnames
close,55

r = 0
openw,9,"matloc_all.dat"
for q = 0L, 2256 do begin
  newnamefile = newnames(q)
;print,newnamefile
  openr,11,newnamefile
  readf,11,count1
  print,q,count1
  loc = indgen(3)
  if (count1 gt 0) then begin
    ; Don't remove the mean, the eof program does that
    ;      mean1 = -count1/66500.
    timemat(r,*) = 0.0
    ;      seriesname1=dblarr(1,66500)
    seriesvalues = dblarr(count1)
    readf,11,seriesvalues
  end
end

```

```

;      print,seriesvalues

c = count1 - 1

FOR k = 0L, 66449L DO BEGIN

FOR n = 0L, c DO BEGIN

IF (abs(series(k)-seriesvalues(n)) lt 0.0001) then begin

timemat(r,k) = 1.0

;      print,series(k)

;      print,seriesvalues(n)

;      print,timemat(r,k)

ENDIF

ENDFOR

ENDFOR

newlat=strmid(newnames(q),3,3)

newlon=strmid(newnames(q),9,4)

loc(0)=r

loc(1)=newlat

loc(2)=newlon

printf,9,loc

r = r + 1

endif

close,11

endfor

close,9

save,filename="idl_all.dat",/all

end

```

## D.2 KL Decomposition

This program is modified from a fortran program kindly provided to me by Dr. Frank Evans, PAOS, University of Colorado at Boulder. Rather than covariance, it computes the correlation between sites, but is substantially unchanged otherwise.

```
WRITE (*,*) 'Time column to transfer to output (or 0 for none) :'
```

```
READ (*,*) TIMECOL
```

```
WRITE (*,*) 'Significant eigenvalues: ',
```

```
.    'total fraction of variance and max number :'
```

```
READ (*,*) FSIGNIF, NSIGNIF
```

```
WRITE (*,*) 'Covariance output file name (or NONE) : '
```

```
READ (*,'(A)') OUTFILE1
```

```
WRITE (*,*) 'Eigenvector output file name (or NONE) : '
```

```
READ (*,'(A)') OUTFILE2
```

```
WRITE (*,*) 'Time series output file name (or NONE) : '
```

```
READ (*,'(A)') OUTFILE3
```

```
IF (FSIGNIF .GT. 1.0)
```

```
.    STOP 'Significance fraction must be less than 1'
```

```

MCOL = SCOL+NCOL-1
IF (NCOL .GT. MAXCOL) STOP 'EOFANAL: MAXCOL exceeded. '
IF (MCOL .GT. MAXCOL+100) STOP 'EOFANAL: MAXCOL ex-
ceeded. '

```

C        Zero the moment arrays

```

DO I1 = 1, NCOL
  MOM1(I1) = 0.0D0
DO I2 = 1, NCOL
  MOM2(I1,I2) = 0.0D0
ENDDO
ENDDO

```

C        Open input file

```
OPEN (UNIT=1, FILE=INFILE, STATUS='OLD')
```

C        Read in the data and compute the moments as we go

C        The first and second absolute moments are accumulated in

C        REAL\*8 arrays so that there is enough numerical precision.

NPTS = 0.

DONE = .FALSE.

DO WHILE (.NOT. DONE)

```
  READ (1,* ,IOSTAT=IERR) (DATACOL(J), J=1, MCOL)
```

```

IF (IERR .GT. 0) THEN
    WRITE (*,*) 'Error reading input file'
    STOP
ELSE IF (IERR .LT. 0) THEN
    DONE = .TRUE.
ELSE
    NPTS = NPTS + 1
    DO I1 = 1, NCOL
        MOM1(I1) = MOM1(I1) + DBLE(DATACOL(SCOL-1+I1))
        DO I2 = I1, NCOL
            MOM2(I1,I2) = MOM2(I1,I2) +
                DBLE(DATACOL(SCOL-1+I1))**
&DBLE(DATACOL(SCOL-1+I2))
        ENDDO
    ENDDO
    ENDIF
    ENDDO
    CLOSE (1)

C      Calculate the covariance matrix
    DO I1 = 1, NCOL
        DO I2 = I1, NCOL
            COV(I1,I2)=MOM2(I1,I2)/DBLE(NPTS)-
&MOM1(I1)*MOM1(I2)/DBLE(NPTS)**2
            COV(I2,I1) = COV(I1,I2)
        ENDDO
        GRID(I1) = FLOAT(I1)
    ENDDO

```

C Calculate frequency (rep = 1/freq) and standard deviation

```
DO I1 = 1, NCOL
    REP(I1) = DBLE(NPTS)/MOM1(I1)
    SD(I1) = SQRT(COV(I1,I1))
ENDDO
```

C Calculate correlation matrix & period correlation matrix

```
DO I1 = 1, NCOL
    DO I2 = I1, NCOL
        CF(I1,I2) = DBLE(COV(I1,I2))/(SD(I1)*SD(I2))
        CF(I2,I1) = CF(I1,I2)
        CFT(I1,I2) = CF(I1,I2)*REP(I1)*REP(I2)
        CFT(I2,I1) = CFT(I1,I2)
    ENDDO
ENDDO
```

C Calculate eigenfunctions for correlation matrix

```
IF (OUTFILE1 .NE. 'NONE') THEN
    CALL OUTPUT_FIELD (OUTFILE1, MAXCOL,NCOL,NCOL,GRID,
&GRID,CF)
ENDIF
WRITE (*,*) 'Done computing correlation matrix.'
```

C OUTPUT REP AND SD, FIND OUT WHAT'S WRONG!

```
OPEN (UNIT=77, FILE='debug.dat', STATUS='UNKNOWN')
```

```
C WRITE (77,'(A)') '!      MEAN          REP          SD'
WRITE (77,'(A)') '!      REP          SD'
```

```

DO I1 = 1, NCOL
    WRITE (77,'(1X,I3,1X,I3,1X,E14.7')I1,NCOL,MOM1(I1)
    WRITE (77,'(1X,E14.7,1X,E14.7')REP(I1), SD(I1)
C      WRITE (77,'(1X,E14.7,1X,E14.7,1X,E14.7')MOM1(I1), REP(I1),
SD(I1)
ENDDO
CLOSE (77)

C      Find eigenvalues and eigenvectors
CALL SEIGEN (MAXCOL, NCOL, CF, EVAL, WORK, EVEC,
IERR)
IF (IERR .NE. 0) THEN
    WRITE (*,*) 'SEIGEN failed.'
    STOP
ENDIF

C      Output eigenvalues and eigenvectors
VAR = 0
DO I = NCOL, 1, -1
    VAR = VAR + EVAL(I)
ENDDO
ICUT = NCOL+1
S = 0
DO I = NCOL, 1, -1
    S = S + EVAL(I)
    IF (S .LT. FSIGNIF*VAR) ICUT = I

```

```

ENDDO

ICUT = NCOL - ICUT + 2
WRITE (*,*) ICUT, ' significant eigenvalues.'
ICUT = MIN(ICUT,NSIGNIF)

IF (OUTFILE2 .NE. 'NONE') THEN
  OPEN (UNIT=2, FILE=OUTFILE2, STATUS='UNKNOWN')
  WRITE (2,'(A)') '! Eigenvalues vect(N) vect(N-1) ...'
  DO I2 = 1, NCOL
    WRITE (2,'(1X,I4,1X,E14.7,1X,30(1X,F9.6))')
    .      I2, EVAL(I2), (EVECT(I2,NCOL-I1+1), I1=1,ICUT)
  ENDDO
  CLOSE (2)
ENDIF

```

C Project the data onto the eigenvectors (EOFs) to  
C find the EOF amplitude time series.

```

IF (OUTFILE3 .NE. 'NONE') THEN
  OPEN (UNIT=1, FILE=INFILE, STATUS='OLD')
  OPEN (UNIT=3, FILE=OUTFILE3, STATUS='UNKNOWN')

```

NPTS = 0

DONE = .FALSE.

DO WHILE (.NOT. DONE)

READ (1,\* ,IOSTAT=IERR) (DATACOL(J), J=1, MCOL)

```

IF (IERR .GT. 0) THEN
  STOP 'Error reading input file'
ELSE IF (IERR .LT. 0) THEN
  DONE = .TRUE.
ELSE
  NPTS = NPTS + 1
  DO I1 = 1, ICUT
    SUM = 0.0
    DO I2 = 1, NCOL
      SUM = SUM + EVECT(I2,NCOL-I1+1)*DATACOL(SCOL-1+I2)
    ENDDO
    VECT(I1) = SUM
  ENDDO
  IF (TIMECOL .GT. 0) THEN
    WRITE (3,'(1X,F9.3,50(1X,F9.3))')
    .   DATACOL(TIMECOL), (VECT(I),I=1,ICUT)
  ELSE
    WRITE (3,'(1X,I6,50(1X,F9.3))') NPTS, (VECT(I),I=1,ICUT)
  ENDIF
  ENDIF
ENDDO
CLOSE (1)
CLOSE (3)
ENDIF
END

```

SUBROUTINE OUTPUT\_FIELD (FILENAME, NP,NX,NY, XGRID,YGRID,  
FIELD)

C Outputs a gridded 2D dataset in text format.

C Parameters (all are input):

C FILENAME name of output gridded file

C NP physical array size

C NX number of grid points in X dimension

C NY number of grid points in Y dimension

C XGRID X grid locations

C YGRID Y grid locations

C FIELD 2D array of field values at grid points

INTEGER NX, NY

REAL XGRID(NX), YGRID(NY), FIELD(NP,\*)

CHARACTER\*(\*) FILENAME

INTEGER I, J, K, N

OPEN (UNIT=9, FILE=FILENAME, STATUS='UNKNOWN')

WRITE (9, '(1X,I4,1X,I4)') NX, NY

DO 80 K = 1, NX, 6

N = MIN(NX-K,5)

WRITE (9, '(6(1X,E12.5),:)') (XGRID(K+I), I=0,N)

80 CONTINUE

DO 90 K = 1, NY, 6

```

N = MIN(NY-K,5)
WRITE (9, '(6(1X,E12.5),:)) (YGRID(K+J), J=0,N)
90    CONTINUE
      DO 110 J = 1, NY
          DO 100 K = 1, NX, 6
              N = MIN(NX-K,5)
              WRITE (9, '(6(1X,E12.5),:)) (FIELD(K+I,J), I=0,N)
100    CONTINUE
110    CONTINUE
      CLOSE (9)

      RETURN
      END

```

```

SUBROUTINE SEIGEN(NM,N,A,D,E,Z,IERR)
C
      INTEGER I,J,K,L,M,N,II,JP1,NM,MML
      REAL*8 A(NM,N)
      REAL*4 D(N),E(N),Z(NM,N)
      REAL*4 B,C,F,G,H,P,R,S,HH,SCALE,MACHEP,FOO,BAR
      C      REAL*8 DSQRT,DABS,DSIGN

```

C

C THIS SUBROUTINE IS A MERGER OF THE EISPACK SUBROUTINES TRED2 AND

C TQL2, WHICH ARE THEMSELVES TRANSLATED FROM THE ALGOL PROCEDURES.

C NUM. MATH. 11, 181-195(1968) BY MARTIN, REINSCH, AND WILKINSON.

C HANDBOOK FOR AUTO. COMP., VOL.II-LINEAR ALGEBRA, 212-226(1971).

C MATRIX EIGENSYSTEM ROUTINES - EISPACK GUIDE, 30(1974).

C

C THIS SUBROUTINE FINDS THE EIGENVALUES AND EIGENVECTORS OF A

C FULL SYMMETRIC REAL MATRIX.

C

C THE WORK IS DONE IN TWO PARTS. THE FIRST (CORRESPONDING TO

C TRED2) REDUCES THE FULL SYMMETRIC MATRIX TO A SYMMETRIC

C TRIDIAGONAL MATRIX USING AND ACCUMULATING ORTHOGONAL SIMILARITY

C TRANSFORMATIONS. THE SECOND (TQL2) FINDS THE EIGENVALUES AND

C EIGENVECTORS OF THE SYMMETRIC TRIDIAGONAL MATRIX BY THE QL METHOD.

C

C ON INPUT:

C

C NM MUST BE SET TO THE ROW DIMENSION OF TWO-DIMENSIONAL

C ARRAY PARAMETERS AS DECLARED IN THE CALLING PROGRAM

C DIMENSION STATEMENT;

C

C N IS THE ORDER OF THE MATRIX;

C

C A CONTAINS THE REAL SYMMETRIC INPUT MATRIX. ONLY THE

C LOWER TRIANGLE OF THE MATRIX NEED BE SUPPLIED.

C

C E IS A REAL VECTOR FOR INTERMEDIATE STORAGE.

C

C ON OUTPUT:

C

C D CONTAINS THE EIGENVALUES IN ASCENDING ORDER. IF AN

C ERROR EXIT IS MADE, THE EIGENVALUES ARE CORRECT BUT

C UNORDERED FOR INDICES 1,2,...,IERR-1;

C

C Z CONTAINS ORTHONORMAL EIGENVECTORS OF THE SYMMETRIC

C TRIDIAGONAL (OR FULL) MATRIX. IF AN ERROR EXIT IS MADE,

C Z CONTAINS THE EIGENVECTORS ASSOCIATED WITH

THE STORED

```

C      EIGENVALUES;
C
C      IERR IS SET TO
C      ZERO      FOR NORMAL RETURN,
C      J      IF THE J-TH EIGENVALUE HAS NOT BEEN
C              DETERMINED AFTER 30 ITERATIONS.
C
C      A AND Z MAY COINCIDE. IF DISTINCT, A IS UNALTERED.
C
C      QUESTIONS AND COMMENTS SHOULD BE DIRECTED TO B.
```

S. GARBOW,

```

C      APPLIED MATHEMATICS DIVISION, ARGONNE NATIONAL
LABORATORY
```

C

C

C

DO 100 I = 1, N

C

DO 100 J = 1, I

Z(I,J) = A(I,J)

100 CONTINUE

C

IF (N .EQ. 1) GO TO 320

C ::::::: FOR I=N STEP -1 UNTIL 2 DO - :::::::

DO 300 II = 2, N

I = N + 2 - II

```

L = I - 1

H = 0.0D0

SCALE = 0.0D0

IF (L .LT. 2) GO TO 130

C      :::::::::::: SCALE ROW (ALGOL TOL THEN NOT NEEDED) ::::::::::::

DO 120 K = 1, L

120  SCALE = SCALE + ABS(Z(I,K))

C

IF (SCALE .NE. 0.0D0) GO TO 140

130  E(I) = Z(I,L)

GO TO 290

C

140  DO 150 K = 1, L

      Z(I,K) = Z(I,K) / SCALE

      H = H + Z(I,K) * Z(I,K)

150  CONTINUE

C

F = Z(I,L)

G = -SIGN(SQRT(H),F)

E(I) = SCALE * G

H = H - F * G

Z(I,L) = F - G

F = 0.0D0

C

DO 240 J = 1, L

      Z(J,I) = Z(I,J) / (SCALE * H)

      G = 0.0D0

```

C ::::::: FORM ELEMENT OF A\*U :::::::

DO 180 K = 1, J

180 G = G + Z(J,K) \* Z(I,K)

C

JP1 = J + 1

IF (L .LT. JP1) GO TO 220

C

DO 200 K = JP1, L

200 G = G + Z(K,J) \* Z(I,K)

C ::::::: FORM ELEMENT OF P :::::::

220 E(J) = G / H

F = F + E(J) \* Z(I,J)

240 CONTINUE

C

HH = F / (H + H)

C ::::::: FORM REDUCED A :::::::

DO 260 J = 1, L

F = Z(I,J)

G = E(J) - HH \* F

E(J) = G

C

DO 260 K = 1, J

Z(J,K) = Z(J,K) - F \* E(K) - G \* Z(I,K)

260 CONTINUE

C

DO 280 K = 1, L

280 Z(I,K) = SCALE \* Z(I,K)

C

290 D(I) = H

300 CONTINUE

C

320 D(1) = 0.0D0

E(1) = 0.0D0

C :::::::::: ACCUMULATION OF TRANSFORMATION MATRICES

:::::::::::

DO 500 I = 1, N

L = I - 1

IF (D(I) .EQ. 0.0D0) GO TO 380

C

DO 360 J = 1, L

G = 0.0D0

C

DO 340 K = 1, L

340 G = G + Z(I,K) \* Z(K,J)

C

DO 360 K = 1, L

Z(K,J) = Z(K,J) - G \* Z(K,I)

360 CONTINUE

C

380 D(I) = Z(I,I)

Z(I,I) = 1.0D0

IF (L .LT. 1) GO TO 500

C

DO 400 J = 1, L

Z(I,J) = 0.0D0  
Z(J,I) = 0.0D0  
400 CONTINUE  
C  
500 CONTINUE  
C  
C ::::::: LAST CARD OF TRED2 :::::::  
C  
C -----  
C  
C ::::::: MACHEP IS A MACHINE DEPENDENT PARAMETER  
SPECIFYING  
C THE RELATIVE PRECISION OF FLOATING POINT  
ARITHMETIC.  
C IT IS THE SMALLEST POSITIVE NUMBER SUCH THAT  
1.+MACHEP=  
FOO=1.00001  
591 CONTINUE  
BAR=FOO  
FOO = SQRT(FOO)  
IF (FOO.NE.1.0D0) GOTO 591  
MACHEP=BAR-1.0D0  
C  
IERR = 0  
IF (N .EQ. 1) GO TO 1001  
C  
DO1100 I = 2, N

1100 E(I-1) = E(I)

C

F = 0.0D0

B = 0.0D0

E(N) = 0.0D0

C

DO1240 L = 1, N

J = 0

H = MACHEP \* (ABS(D(L)) + ABS(E(L)))

IF (B .LT. H) B = H

C :::::::::: LOOK FOR SMALL SUB-DIAGONAL ELEMENT :::::::::::::

DO1110 M = L, N

IF (ABS(E(M)) .LE. B) GO TO1120

C :::::::::: E(N) IS ALWAYS ZERO, SO THERE IS NO EXIT

C :::::::::: THROUGH THE BOTTOM OF THE LOOP ::::::::::::

1110 CONTINUE

C

1120 IF (M .EQ. L) GO TO1220

1130 IF (J .EQ. 30) GO TO 1000

J = J + 1

C :::::::::: FORM SHIFT ::::::::::::

P = (D(L+1) - D(L)) / (2.0D0 \* E(L))

R = SQRT(P\*P+1.0D0)

H = D(L) - E(L) / (P + SIGN(R,P))

C

DO1140 I = L, N

1140 D(I) = D(I) - H

C

$$F = F + H$$

C ::::::: QL TRANSFORMATION :::::::

$$P = D(M)$$

$$C = 1.0D0$$

$$S = 0.0D0$$

$$MML = M - L$$

C ::::::: FOR I=M-1 STEP -1 UNTIL L DO - :::::::

$$DO1200 II = 1, MML$$

$$I = M - II$$

$$G = C * E(I)$$

$$H = C * P$$

$$IF (ABS(P) .LT. ABS(E(I))) GO TO1150$$

$$C = E(I) / P$$

$$R = SQRT(C*C+1.0D0)$$

$$E(I+1) = S * P * R$$

$$S = C / R$$

$$C = 1.0D0 / R$$

GO TO1160

$$1150 \quad C = P / E(I)$$

$$R = SQRT(C*C+1.0D0)$$

$$E(I+1) = S * E(I) * R$$

$$S = 1.0D0 / R$$

$$C = C * S$$

$$1160 \quad P = C * D(I) - S * G$$

$$D(I+1) = H + S * (C * G + S * D(I))$$

C ::::::: FORM VECTOR :::::::

DO1180 K = 1, N

$$H = Z(K,I+1)$$

$$Z(K,I+1) = S * Z(K,I) + C * H$$

$$Z(K,I) = C * Z(K,I) - S * H$$

1180 CONTINUE

C

1200 CONTINUE

C

$$E(L) = S * P$$

$$D(L) = C * P$$

IF (ABS(E(L)) .GT. B) GO TO1130

1220 D(L) = D(L) + F

1240 CONTINUE

C :::::: ORDER EIGENVALUES AND EIGENVECTORS :::::::

DO1300 II = 2, N

$$I = II - 1$$

$$K = I$$

$$P = D(I)$$

C

DO1260 J = II, N

IF (D(J) .GE. P) GO TO1260

$$K = J$$

$$P = D(J)$$

1260 CONTINUE

C

IF (K .EQ. I) GO TO1300

$$D(K) = D(I)$$

D(I) = P

C

DO1280 J = 1, N

P = Z(J,I)

Z(J,I) = Z(J,K)

Z(J,K) = P

1280 CONTINUE

C

1300 CONTINUE

C

GO TO 1001

C ::::::: SET ERROR - NO CONVERGENCE TO AN

C EIGENVALUE AFTER 30 ITERATIONS :::::::

1000 IERR = L

1001 RETURN

C ::::::: LAST CARD OF TQL2 :::::::

C

C ::::::: LAST CARD OF SEIGEN :::::::

END

### D.3 Ouput programs

```
; out_gmt2.pro - idl program to take eigenvector information and output
; eigenvalue plots and boxfiles with locations for GMT use in making maps.

openr,22,"eof_all.dat"
eof=dblarr(27,1329)
readf,22,eof
close,22
openr,11,"matloc_all.dat"
loc=fltarr(3,1329)
readf,11,loc
close,11
;sta=loc(0, *)
lat=0.05+loc(1,*)/10.
lon=-0.05-loc(2,*)/10.
m=dblarr(2)

!p.font=0
set_plot,'ps'
w=dblarr(1329)
w=eof[1,*]
s=reverse(sort(w))
w=w(s)
x=findgen(1329)
device,filename="lambda_all.ps",/demi,/bkman,/landscape
```

```

plot,x[0:20],w[0:20],title="Eigenvalues for EOF Decomposition of Southern
California Seismicity",ytitle="Percent of Variance",xtitle="Mode"
device,/close
device,filename="lambda_logall.ps",/demi,/bkman,/landscape
plot,x[0:1328],w[0:1328],title="Eigenvalues for EOF Decomposition of South-
ern California Seismicity",xrange=[0.5,2000],yrange=[0.00001,1.0],ytitle="Percent
of Variance",xtitle="Mode",/xlog,/ylog,/xstyle,/ystyle
device,/close
set_plot,'x'

eof1=dblarr(3,1329)
eof1a=dblarr(1329)
eof1a=eof[2,*]
m(0)=max(eof1a)
m(1)=abs(min(eof1a))
norm=max(m)
eof1a=(eof1a)/norm
eof1[0,*]=lon
eof1[1,*]=lat
eof1[2,*]=eof1a
openw,11,"box.1"
printf,11,eof1
close,11

eof2=dblarr(3,1329)
eof2a=dblarr(1329)

```

```

eof2a=eof[3,*]
m(0)=max(eof2a)
m(1)=abs(min(eof2a))
norm=max(m)
eof2a=(eof2a)/norm
eof2[0,*]=lon
eof2[1,*]=lat
eof2[2,*]=eof2a
openw,11,"box.2"
printf,11,eof2
close,11

```

```

eof3=dblarr(3,1329)
eof3a=dblarr(1329)
eof3a=eof[4,*]
m(0)=max(eof3a)
m(1)=abs(min(eof3a))
norm=max(m)
eof3a=(eof3a)/norm
eof3[0,*]=lon
eof3[1,*]=lat
eof3[2,*]=eof3a
openw,11,"box.3"
printf,11,eof3
close,11

```

eof4=dblarr(3,1329)

```

eof4a=dblarr(1329)
eof4a=eof[5,*]
m(0)=max(eof4a)
m(1)=abs(min(eof4a))
norm=max(m)
eof4a=(eof4a)/norm
eof4[0,*]=lon
eof4[1,*]=lat
eof4[2,*]=eof4a
openw,11,"box.4"
printf,11,eof4
close,11

```

```

eof5=dblarr(3,1329)
eof5a=dblarr(1329)
eof5a=eof[6,*]
m(0)=max(eof5a)
m(1)=abs(min(eof5a))
norm=max(m)
eof5a=(eof5a)/norm
eof5[0,*]=lon
eof5[1,*]=lat
eof5[2,*]=eof5a
openw,11,"box.5"
printf,11,eof5
close,11

```

```

eof6=dblarr(3,1329)
eof6a=dblarr(1329)
eof6a=eof[7,*]
m(0)=max(eof6a)
m(1)=abs(min(eof6a))
norm=max(m)
eof6a=(eof6a)/norm
eof6[0,*]=lon
eof6[1,*]=lat
eof6[2,*]=eof6a
openw,11,"box.6"
printf,11,eof6
close,11

```

```

eof7=dblarr(3,1329)
eof7a=dblarr(1329)
eof7a=eof[8,*]
m(0)=max(eof7a)
m(1)=abs(min(eof7a))
norm=max(m)
eof7a=(eof7a)/norm
eof7[0,*]=lon
eof7[1,*]=lat
eof7[2,*]=eof7a
openw,11,"box.7"
printf,11,eof7
close,11

```

```

eof8=dblarr(3,1329)
eof8a=dblarr(1329)
eof8a=eof[9,*]
m(0)=max(eof8a)
m(1)=abs(min(eof8a))
norm=max(m)
eof8a=(eof8a)/norm
eof8[0,*]=lon
eof8[1,*]=lat
eof8[2,*]=eof8a
openw,11,"box.8"
printf,11,eof8
close,11

```

```

eof9=dblarr(3,1329)
eof9a=dblarr(1329)
eof9a=eof[10,*]
m(0)=max(eof9a)
m(1)=abs(min(eof9a))
norm=max(m)
eof9a=(eof9a)/norm
eof9[0,*]=lon
eof9[1,*]=lat
eof9[2,*]=eof9a
openw,11,"box.9"
printf,11,eof9

```

```
close,11
```

```
eof10=dblarr(3,1329)  
eof10a=dblarr(1329)  
eof10a=eof[11,*]  
m(0)=max(eof10a)  
m(1)=abs(min(eof10a))  
norm=max(m)  
eof10a=(eof10a)/norm  
eof10[0,*]=lon  
eof10[1,*]=lat  
eof10[2,*]=eof10a  
openw,11,"box.10"  
printf,11,eof10  
close,11
```

```
END
```

```
; pca2.pro - idl program to plot pc time series.
```

```
pca3=fltarr(26,66500)  
yr=dindgen(66500)
```

```

yr=1932.+yr/1000.

openr,11,"pca_all.dat"
readf,11,pca3
close,11

set_plot,'ps'
!p.font=0

device,file="pca1.ps",/bkman,/demi,/landscape
plot,yr,xtitle="Time",ytitle="PC Amplitude",title="Southern California Seis-
micity, PC1",pca3[1,*],/xstyle,yrange=[-2.,4]
device,/close

device,file="pca2.ps",/bkman,/demi,/landscape
plot,yr,xtitle="Time",ytitle="PC Amplitude",title="Southern California Seis-
micity, PC2",pca3[2,*],/xstyle,yrange=[-2.,4]
device,/close

device,file="pca3.ps",/bkman,/demi,/landscape
plot,yr,xtitle="Time",ytitle="PC Amplitude",title="Southern California Seis-
micity, PC3",pca3[3,*],/xstyle,yrange=[-2.,4]
device,/close

device,file="pca4.ps",/bkman,/demi,/landscape
plot,yr,xtitle="Time",ytitle="PC Amplitude",title="Southern California Seis-
micity, PC4",pca3[4,*],/xstyle,yrange=[-2.,4]
device,/close

```

```
device,file="pca5.ps",/bkman,/demi,/landscape  
plot,yr,xtitle="Time",ytitle="PC Amplitude",title="Southern California Seis-  
micity, PC5",pca3[5,*],/xstyle,yrange=[-2.,4]
```

```
device,/close
```

```
device,file="pca6.ps",/bkman,/demi,/landscape  
plot,yr,xtitle="Time",ytitle="PC Amplitude",title="Southern California Seis-  
micity, PC6",pca3[6,*],/xstyle,yrange=[-2.,4]
```

```
device,/close
```

```
device,file="pca7.ps",/bkman,/demi,/landscape  
plot,yr,xtitle="Time",ytitle="PC Amplitude",title="Southern California Seis-  
micity, PC7",pca3[7,*],/xstyle,yrange=[-2.,4]
```

```
device,/close
```

```
device,file="pca8.ps",/bkman,/demi,/landscape  
plot,yr,xtitle="Time",ytitle="PC Amplitude",title="Southern California Seis-  
micity, PC8",pca3[8,*],/xstyle,yrange=[-2.,4]
```

```
device,/close
```

```
device,file="pca9.ps",/bkman,/demi,/landscape  
plot,yr,xtitle="Time",ytitle="PC Amplitude",title="Southern California Seis-  
micity, PC9",pca3[9,*],/xstyle,yrange=[-2.,4]
```

```
device,/close
```

```
device,file="pca10.ps",/bkman,/demi,/landscape
```

```

plot,yr,xtitle="Time",ytitle="PC Amplitude",title="Southern California Seis-
micity, PC10",pca3[10,*],/xstyle,yrange=[-2.,4]
device,/close

set _plot,'x'

end

#!/bin/csh
# This is my first GMT plot
# eof1.map - executable GMT program for plotting first eigenvector in
southern California.

set range="-121.1/-114.9/32/37."
set proj="M4.5"

psbasemap -R${range} -J${proj} -F0/0/0 -P \
-B1f1:"CBF1 for Southern California Seismicity, 1932-1998":WSen \
-V -K \
> eof1.ps

```

```

psxy box.1 -J${proj} -R${range} -Cpca3.cpt -P -M -K -O -N \
-Ss0.065 -V >> eof1.ps

#psxy box.test -J${proj} -R${range} -Cpca3.cpt -P -M -K -O -N \
#      -Ss0.05 -V >> eof1.ps

psxy calif-fault.gmtlin -J${proj} -R${range} -P -M -K -O \
-W6/128/128/128 -V >> eof1.ps

pscoast -R${range} -Dh -J${proj} -P -O \
-S220/220/255 -Ia -Na -W6 -V -K \
>> eof1.ps

pstext -R${range} -J${proj} -G0/0/0 -O label.txt -V -O \
>> eof1.ps

#psscale -Cpca3.cpt -D2.25/5.25/4.5/0.15h \
#      -V -O >> eof1.ps

ghostview eof1.ps&

```

## Appendix E

### Pattern Dynamics (PDPC) Programs

Here I have included the idl programs which take the seismicity data, in the form of its eigenvectors, and compute  $\hat{\mathbf{S}}$ . The first Plot\_L2\_Probability, creates a seismicity data file, the second Plot\_L2\_Seis, calculates the probabilities for various years differences, and the third file, Plot\_L2\_Deltas, outputs the actual  $\Delta\hat{\mathbf{S}}$  for plotting purposes. Again, I have included a sample GMT plotting program.

```
; ; Plot_L2_Probability
;      Creates data file for plotting relative probabilities

num_iter = 0

file_eig = ''
print, ''
print, 'Enter Eigenvector File Name'
read, file_eig
```

```
file_data = ''
print, ''
print, 'Enter Seismicity File'
read, file_data

openr, 1, file_eig
readf, 1, num_eigen

SEIS1 = fltarr(num_eigen)
SEISMICITY      = fltarr(num_eigen)
SEIS2 = fltarr(num_eigen)

YEAR = 1

openr, 2, file_data
readf, 2, num_years

SEIS_FILE = fltarr(num_years, num_eigen)
YEAR_FILE = fltarr(num_years)
BETA_FILE = fltarr(num_years, num_eigen)

for year_index = 0L, num_years-1 do begin

    readf, 2, YEAR
    readf, 2, SEIS1
```

```

SEISMICITY[*] = SEIS1

YEAR_FILE(year_index) = long(YEAR)
SEIS_FILE(year_index, *) = SEISMICITY(*)

endfor

close, 2

MEAN_SEIS      = fltarr(num_years)
SDEV_SEIS      = fltarr(num_years)

nm_eig = float(num_eigen)
sq_nm_eig = sqrt(nm_eig)

;      Now accumulate seismicity
;

for year_index = 1L, num_years-1 do begin

    SEIS_FILE(year_index, *) = SEIS_FILE(year_index, *) +
    &SEIS_FILE(year_index-1, *)

endfor

;

;      Normalize each year

```

;

```

for year_index = 1L, num_years-1 do begin
    SEIS_FILE(year_index, *) = SEIS_FILE(year_index,*)/float(year_index+1)
endfor

```

```

for year_index = 0L, num_years-1 do begin

```

; Here I normalize the seismicity function so that  
; the total power  
; (L2 Norm) in the power spectrum is always the same.  
;

```
SEISMICITY(*) = SEIS_FILE(year_index, *)
```

```
nm_eig = float(num_eigen)
```

```

for j = 0L, num_eigen-1 do begin
    MEAN_SEIS(year_index) = MEAN_SEIS(year_index)
&+ SEISMICITY(j)
endfor

```

```
MEAN_SEIS(year_index) = MEAN_SEIS(year_index)/num_eigen
```

```

for i=0L, num_eigen-1 do begin
    SEIS2(i) = (SEISMICITY(i) - MEAN_SEIS(year_index))
endfor

```

```

for j = 0L, num_eigen-1 do begin
    SDEV_SEIS(year_index) = SDEV_SEIS(year_index)
    + SEIS2(j)*SEIS2(j)
endfor

```

SDEV\_SEIS(year\_index) = SDEV\_SEIS(year\_index)/num\_eigen

SDEV\_SEIS(year\_index) = sqrt(SDEV\_SEIS(year\_index))

```

for i=0L, num_eigen-1 do begin
    SEISMICITY(i) = SEIS2(i)/SDEV_SEIS(year_index)
endfor

```

SEIS\_FILE(year\_index, \*) = SEISMICITY(\*)

endfor

save, filename="seis.dat",/all

```
end  
  
;  
; Plot_L2_Pickup  
; For doing different years with the same seisfile  
;  
  
restore, filename="seis.dat"  
  
final_name=strarr(27)  
base_name=strarr(27)  
PROB_MATRIX = dblarr(27,num_eigen)  
num = dblarr(2)  
  
; Enter plot loop  
  
v = 0  
resp = 'y'  
  
base_year = 1  
final_year = 1
```

```

while (resp eq 'y') do begin

    print, 'What Base Year for Plot?'
    read, base_year
    base_year = long(base_year)

    print, 'What Final Year for Plot?'
    read, final_year
    final_year = long(final_year)

    year_zero = YEAR_FILE(0)
    base_year_index = base_year - year_zero
    final_year_index = final_year - year_zero
    year_diff_index = final_year - base_year

    PROB = fltarr(num_eigen)

;      Make up the pattern vector
;

    PATTERN = fltarr(num_eigen)

        for i=0L, num_eigen-1 do begin
            PATTERN(i) = SEIS_FILE(final_year_index,i) -
SEIS_FILE(base_year_index,i)

```

endfor

for i=0L, num\_eigen-1 do begin

PROB(i) = PATTERN(i)\*PATTERN(i)

endfor

PROB\_SUM = 0.0

for i=0L, num\_eigen-1 do begin

PROB\_SUM = (PROB\_SUM + PROB(i))

endfor

PROB(\*) = (PROB(\*) - PROB\_SUM/num\_eigen)

PROB\_MATRIX[v,\*] = PROB

final\_name[v] = string(final\_year)

base\_name[v] = string(base\_year)

; print,v,final\_name[v],base\_name[v]

v = v + 1

print, 'Continue with another plot? (y/n)'

read, resp

endwhile

num(0)=max(PROB\_MATRIX)

num(1)=abs(min(PROB\_MATRIX))

norm=max(num)\*0.05

```

output_box=fltarr(3,num_eigen)

openr,11,"boxfile"
loc=fltarr(3,num_eigen)
readf,11,loc
close,11
output_box[1,*]=0.05+loc[1,*]
output_box[0,*]=0.05+loc[2,*]

for i = 0, v-1 do begin
  boxname = strcompress('box.' + base_name[i] + final_name[i] +
'_prob',/remove_all)
  output_box[2,*]=PROB_MATRIX[i,*]/norm

for s = 0L, num_eigen-1 do begin
  if (output_box[2,s] gt 1.0) then output_box[2,s] = 1.0
  if (output_box[2,s] lt 0.0) then output_box[2,s] = 0.0
endfor

f=where(output_box[2,*] gt 0.0,C)
loc1=fltarr(3,C)
loc1=output_box(*,f)

openw,11,boxname

```

```
printf,11,loc1
close,11

print,boxname,v
r = 0.0
hist = fltarr(1000)
for k = 0, 999 do begin
    f = where(output_box[2,*] ge r and output_box[2,*] lt (r +
0.001),C)
    hist(k) = C
    r = r + 0.001
endfor

endfor

end

;

;      Plot_L2_DeltaS
;      Calculates and outputs  $\Delta\hat{S}$ .
;
```

```
restore, filename="seis.dat"

final_name=strarr(27)
base_name=strarr(27)
PROB_MATRIX = dblarr(27,num_eigen)
num = dblarr(2)
print,'Remember to use less than 27 different years!!'

; Enter plot loop

v = 0
resp = 'y'

base_year = 1
final_year = 1

while (resp eq 'y') do begin

    print, 'What Initial Year for Plot?'
    read, base_year
    base_year = long(base_year)

    print, 'What Final Year for Plot?'
    read, final_year
    final_year = long(final_year)
```

```

year_zero = YEAR_FILE(0)
base_year_index = base_year - year_zero
final_year_index = final_year - year_zero
year_diff_index = final_year - base_year

PROB = fltarr(num_eigen)

;      Make up the pattern vector
;

PATTERN = fltarr(num_eigen)

for i=0L, num_eigen-1 do begin
    PATTERN(i) = SEIS_FILE(final_year_index,i) -
SEIS_FILE(base_year_index,i)
endfor

;

for i=0L, num_eigen-1 do begin
    PROB(i) = PATTERN(i)*PATTERN(i)
endfor

;

PROB_SUM = 0.0
for i=0L, num_eigen-1 do begin
    PROB_SUM = (PROB_SUM + PROB(i))
endfor

;
PROB(*) = (PROB(*) - PROB_SUM/num_eigen)

```

PROB\_MATRIX[v,\*] = PATTERN

```

final_name[v] = string(final_year)
base_name[v] = string(base_year)
;      print,v,final_name[v],base_name[v]

```

v = v + 1

print, 'Continue with another plot? (y/n)'

read, resp

endwhile

num(0)=max(PROB\_MATRIX)

num(1)=abs(min(PROB\_MATRIX))

norm=max(num)

output\_box=fltarr(3,num\_eigen)

openr,11,"boxfile"

loc=fltarr(3,num\_eigen)

readf,11,loc

close,11

output\_box[1,\*]=0.05+loc[1,\*]

output\_box[0,\*]=0.05+loc[2,\*]

for i = 0, v-1 do begin

```

boxname = strcompress('box.' + base_name[i] + final_name[i] +
'_deltas',/remove_all)

output_box[2,*]=PROB_MATRIX[i,*]/norm

;      for s = 0L, num_eigen-1 do begin
;          if (output_box[2,s] gt 1.0) then output_box[2,s] = 1.0
;          if (output_box[2,s] lt 0.0) then output_box[2,s] = 0.0
;      endfor

;      f=where(output_box[2,*] gt 0.0,C)
;      loc1=fltarr(3,C)
loc1=output_box

openw,11,boxname
printf,11,loc1
close,11

print,boxname,v
r = 0.0
hist = fltarr(1000)
for k = 0, 999 do begin
    f = where(output_box[2,*] ge r and output_box[2,*] lt (r +
0.001),C)
    hist(k) = C
    r = r + 0.001
endfor

```

```

endfor

end

#!/bin/csh
# This is my first GMT plot

set range="-123./-114.9/32/38.5"
set proj="M4.5"

psbasemap -R${range} -J${proj} -F0/0/0 -P \
-B1f1:"PDPC, 1984-1991":WSen \
-V -K \
> beta_19841991_prob.ps

psxy box.19841991_prob -J${proj} -R${range} -Cfour.cpt -P -M -K -O -N \
-Ss0.05 -V >> beta_19841991_prob.ps

psxy box.actual8491 -J${proj} -R${range} -Cpca_actual.cpt -P -M -K -O \
-N \
-Si0.075 -V >> beta_19841991_prob.ps

```

```

psxy calif-fault.gmtlin -J${proj} -R${range} -P -M -K -O \
-W6/128/128/128 -V >> beta_19841991_prob.ps

pscoast -R${range} -Dh -J${proj} -P -O \
-Ia -Na -W6 -V -K \
>> beta_19841991_prob.ps

pstext -R${range} -J${proj} -G0/0/0 -O label.txt -V -K -O \
>> beta_19841991_prob.ps

psscale -Cfour.cpt -D2.25/5.25/4.5/0.15h \
-V -O >> beta_19841991_prob.ps

ghostview beta_19841991_prob.ps&

psbasemap -R${range} -J${proj} -F0/0/0 -P \
-B1f1:"PDPC, 1991-1979":WSen \
-V -K \
> beta_19791991_prob.ps

psxy box.19791991_prob -J${proj} -R${range} -Cfour.cpt -P -M -K -O -N \
-Ss0.05 -V >> beta_19791991_prob.ps

psxy box.actual7991 -J${proj} -R${range} -Cpca_actual.cpt -P -M -K -O \
-N \

```

```
-Si0.075 -V >> beta_19791991_prob.ps

psxy calif-fault.gmtlin -J${proj} -R${range} -P -M -K -O \
-W6/128/128/128 -V >> beta_19791991_prob.ps

pscoast -R${range} -Dh -J${proj} -P -O \
-Ia -Na -W6 -V -K \
>> beta_19791991_prob.ps

pstext -R${range} -J${proj} -G0/0/0 -O label.txt -V -K -O \
>> beta_19791991_prob.ps

psscale -Cfour.cpt -D2.25/5.25/4.5/0.15h \
-V -O >> beta_19791991_prob.ps

ghostview beta_19791991_prob.ps&
```

## Appendix F

### Genetic Algorithm Programs

The following are a listing of the nine programs necessary to run the ellipsoidal genetic algorithm inversion. I have not included copies as they total over 120 pages in total. The programs themselves are available via ftp from richter.colorado.edu.

The main program is inversion.c, followed by fit.c, the actual fitness function for both the ellipse and the normal fault. In addition, I have included ran1.c, a gaussian random number generator written from Numerical Recipes in C, 2nd edition. Finally, the four input files are gadata.txt, fit1.in, error.in, and motion.in are included, followed by a sample output file, galog.txt.

Inversion.c is a simple genetic algorithm implementation where the evaluation function takes positive values only and the fitness of an individual is the same as the value of the objective function.

The random value generator used is ran1.c, with the initial negative value seed supplied by the user with the following semantics: "a.out seed".

Current options include onept or twopt crossover and a windowing function. The window function should be used if the standard deviation of the fitness becomes too small and the search stagnates. The window function narrows the difference between successive generations and increases the pressure to convergence.

In order to find the optimal crossover and mutation rates for a particular problem, as well as number of generations, I recommend creating a set of synthetic data (you have to write a forward model anyway, for the fitness function), and testing the algorithm with various rates until you get good convergence for your inversion model. A good rule-of-thumb is to set the mutation rate equal to  $1/n$ , where  $n$  is the number of parameters.

The fitness function, fit.c, is for a two-ellipse inversion with one normal fault. The inversions finds the fourteen ellipse parameters, seven per source, and the normal fault displacement.

Input files include gadata.txt, which gives the range of acceptable search parameters for the fitness function, fit1.in, which currently gives the location of the data to be inverted for, and motion.in, which gives the actual values at those locations (in my case deformation). The current fitness function performs a chisquare fit using an error file of the standard deviations to the data, error.in. Our Sun SPARC 10 compiles the a.out using gcc with the -lm option. Fit.c and ran1.c must also be in the current directory.