

# Detecting premonitory seismicity patterns based on critical point dynamics

G. Zöller<sup>1</sup> and S. Hainzl<sup>1,2</sup>

<sup>1</sup>Institute of Physics, University of Potsdam, Germany

<sup>2</sup>Institute of Earth Sciences, University of Potsdam, Germany

Received: 8 May 2001 – Accepted: 16 July 2001

**Abstract.** We test the hypothesis that critical point dynamics precedes strong earthquakes in a region surrounding the future hypocenter. Therefore, we search systematically for regions obeying critical point dynamics in terms of a growing spatial correlation length (GCL). The question of whether or not these spatial patterns are correlated with future seismicity is crucial for the problem of predictability. The analysis is conducted for earthquakes with  $M \geq 6.5$  in California. As a result, we observe that GCL patterns are correlated with the distribution of future seismicity. In particular, there are clear correlations in some cases, e.g. the 1989 Loma Prieta earthquake and the 1999 Hector Mine earthquake. We claim that the critical point concept can improve the seismic hazard assessment.

## 1 Introduction

Different critical point concepts have been discussed extensively with respect to the predictability of earthquakes (Bufe and Varnes, 1993; Jaumé and Sykes, 1999; Hainzl et al., 1999, 2000; Hainzl and Zöller, 2001). Motivated by damage mechanics and laboratory experiments (Leckie and Hayhurst, 1977; Das and Scholz, 1981), the time-to-failure approaches assume that the preparatory process of a large earthquake is characterized by a highly correlated stress field with a growing correlation length (GCL) and an accelerating energy/moment (AMR) release. In practice, these concepts have been tested by fitting time-to-failure relations to seismicity data. For the AMR model, this relation is

$$(\Sigma\sqrt{E})(t) = A - B(t_f - t)^m, \quad (1)$$

with positive constants  $A$ ,  $B$ ,  $m$ , the time-to-failure  $t_f$ , and the cumulative Benioff strain  $(\Sigma\sqrt{E})(t)$ , where  $E$  is the energy release of an earthquake. In the GCL model, the corre-

lation length  $\xi$  is expected to diverge for  $t \rightarrow t_f$  according to

$$\xi(t) = C(t_f - t)^{-k} \quad (2)$$

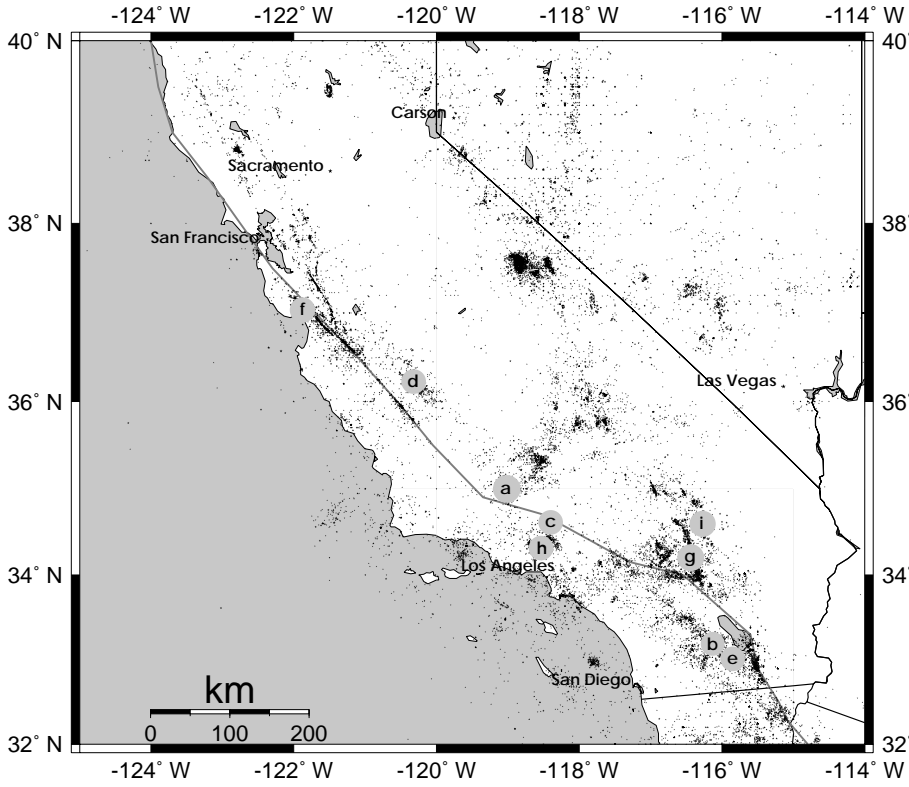
with positive constants  $C$  and  $k$ . Both approaches describe the same underlying mechanism, namely, the critical point dynamics. An important problem is the determination of free parameters, which are in addition to  $A$ ,  $B$ ,  $t_f$ , and  $m$  in Eq. (1), respectively,  $C$ ,  $t_f$ , and  $k$  in Eq. (2), which represent windows for space, time, and magnitude.

The accelerating moment release in terms of cumulative Benioff strain has been documented in several cases, e.g. for California seismicity (Bufe and Varnes, 1993; Bowman et al., 1998; Brehm and Braile, 1998, 1999). The growth of the spatial correlation length has been concluded from variations in the epicenter distribution (Zöller et al., 2001). However, these studies have not been conducted systematically in space and time, i.e. the analysis was restricted to the occurrence time and the epicenter of the largest events. Thus, possible false alarms (critical point behaviour without a subsequent strong earthquake) have not been examined. Therefore, it is an open question whether or not the observed phenomena are unique, i.e. the occurrence of patterns prior to large earthquakes is only meaningful if there is a systematic correlation between these patterns and subsequent earthquakes.

In the present work, we compare patterns based on critical point dynamics in terms of GCL before strong earthquakes with the epicenters of these events, and subsequent intermediate to large earthquakes. By performing a systematic spatial search algorithm, we address the question of spatial correlations. To estimate the significance of the results, the method is also applied to catalogues from an appropriate Poisson process model.

## 2 Data and method

In this section, we present the data and the method to detect spatial correlations between GCL patterns and subsequent



**Fig. 1.** Earthquakes with  $M \geq 3.0$  in California since 1910. Solid circles denote the events with  $M \geq 6.5$  since 1952: circle (a), 1952  $M = 7.5$  Kern County; circle (b), 1968  $M = 6.5$  Borrego Mountain; circle (c), 1971  $M = 6.6$  San Fernando; circle (d), 1983  $M = 6.7$  Coalinga; circle (e), 1987  $M = 6.6$  Superstition Hills; circle (f), 1989  $M = 7.0$  Loma Prieta; circle (g), 1992  $M = 7.3$  Landers; circle (h), 1994  $M = 6.6$  Northridge; and circle (i), 1999  $M = 7.1$  Hector Mine.

seismicity.

We analyze the seismicity in California between the  $32^\circ$  N and  $40^\circ$  N latitude and the  $-125^\circ$  W and  $-114^\circ$  W longitude. The data are taken from the Council of the National Seismic System (CNSS) Worldwide Earthquake Catalogue. The catalogue covers the time span from 1910 to 2000. The distribution of earthquakes is shown in Fig. 1. To account for the completeness of the data, we restrict the analysis to the nine strongest earthquakes with  $M \geq 6.5$  since 1952. Note that completeness of the CNSS catalogue was not achieved until 1940.

For a detailed description of the GCL model, we refer to Zöller et al. (2001). The method is based on a fit of Eq. (2) to the data in a circular space window with radius  $R$  and in a time interval  $(t_0; t_f)$  for earthquakes with magnitudes  $M \geq M_{\text{cut}} = 4.0$ . The exponent  $k$  is set to  $k = 0.4$  according to the result of Zöller et al. (2001). The power law fit is then compared with the fit of a constant and the quality of the power law fit is measured by the curvature value introduced by Bowman et al. (1998),

$$\mathcal{C} = \frac{\text{power law fit root-mean-square error}}{\text{constant fit root-mean-square error}}. \quad (3)$$

Around each epicenter of a strong earthquake, the curvature parameter has been calculated for different values of  $R$  and  $t_0$ . The set of parameters for which  $\mathcal{C}$  is minimal is used for further calculations; i.e. the space window ( $R$ ) and the length of the time interval ( $t_0$ ) are adjusted in order to optimize  $\mathcal{C}$ . The approach of looking at different spatial scales is

based on the observation of Zöller et al. (1998), that the dynamics of a spatially extended system is most clearly visible on intermediate spatial scales between the noisy microscales and the large scales, where the dynamics are hidden due to the averaging. The  $\mathcal{C}$  values are determined on a spatial grid with a resolution of  $0.5^\circ$  in longitude and latitude at nine different times  $t_f^i$ , corresponding to the occurrence times of the nine earthquakes with  $M \geq 6.5$ , denoted with index  $i$ . The result is a function  $\mathcal{C}^i(\mathbf{x})$  for the GCL model, which is compared with the epicenter distribution of the earthquakes with  $M \geq 5.0$  in the time interval  $(t_f^i; t_f^i + 1 \text{ year})$ . This set of epicenters is called the pattern  $Q^i(\mathbf{x})$  for the  $i$ th strongest earthquake. The (arbitrary) magnitude threshold  $M = 5.0$  defining the pattern  $Q^i(\mathbf{x})$  has been introduced, since the premonitory patterns are assumed to be correlated not only with the strongest earthquake, but also with some subsequent main shock activity.

In the next step, the curvature parameter  $\mathcal{C}_{\text{APC}}^i(\mathbf{x})$  is calculated for 100 adjusted Poisson catalogues (APC) in order to derive a measure for the statistical significance of the results. These catalogues are calculated according to the algorithm of Zöller et al. (2001):

1. The CNSS catalogue is declustered using the algorithm of Reasenber (1985);
2. Random epicenters according to the epicenter distribution of the declustered CNSS catalogue are calculated;
3. The earthquake occurrence times are drawn from a Poisson process;

4. The earthquake magnitudes are taken randomly from a probability distribution fulfilling the Gutenberg-Richter law (Gutenberg and Richter, 1956);
5. Aftershocks according to the law of Omori (1894) are added using the algorithm of Reasenber (1985) in the inverse direction.

The resulting earthquake catalogue corresponds to a Poisson process in time with additional aftershock activity. The distributions of the epicenters and the magnitudes are similar to those of the genuine catalogue. Note that only the spatiotemporal correlations of the seismicity are randomized and all other features are preserved. Therefore, the APCs allow one to test for systematic spatiotemporal behaviour.

The likelihood ratio test has been proposed by Gross and Rundle (1998) in order to compare two models with respect to their suitability to describe an observed data set. In this work, the observed data are given by the set  $Q^i(\mathbf{x})$  of epicenters with  $M \geq 5.0$  after the  $i$ th strongest earthquake. Model 1 is defined by the GCL pattern of the original catalogue before the  $i$ th strongest earthquake, i.e. the distribution of curvature parameters  $C^i(\mathbf{x})$  in space. Model 2 is the corresponding pattern  $C^i_{\text{APC}}(\mathbf{x})$  for an APC. For both models, the likelihood function  $L$  is computed with respect to the  $N$  earthquakes, forming the pattern  $Q^i(\mathbf{x})$ :

$$L = \prod_{k=1}^N P(\mathbf{x}_k, C_k). \quad (4)$$

$P(\mathbf{x}_k, C_k)$  is the normalized probability density for an event occurring at the epicenter  $\mathbf{x}_k$  with a premonitory GCL pattern characterized by the curvature parameter  $C_k$ . To apply the likelihood ratio test, we assume Gaussian probability density functions  $P(\mathbf{x}, C) = p_1(\mathbf{x}) \times p_2(C)$  consisting of a two-dimensional Gaussian function  $p_1$  around the spatial grid node  $\mathbf{x}$  with standard deviation  $\sigma_1$  and a (right wing) Gaussian function  $p_2$  depending on the curvature parameter  $C$  with standard deviation  $\sigma_2$ . The value of  $\sigma_1$  is the distance between two adjacent grid nodes and  $\sigma_2 = 0.35$  is an empirical value (Zöller et al., 2001). It should be noted that Eq. (4) must be applied cautiously, since this equation only holds if the  $N$  earthquakes are statistically independent.

The likelihood function is also measured for each of the APCs (model 2). The likelihood ratio  $LR^i = L/L_{\text{APC}}$  of the normalized likelihood functions for model 1 and model 2 is equal to the probability ratio  $p/p_{\text{APC}}$ , where  $p$  denotes the probability that  $Q^i(\mathbf{x})$  arises from the original data (model 1) and  $p_{\text{APC}}$  is the corresponding probability for the APCs (model 2). In the case of  $LR^i > 1$ , the detected GCL patterns in the original catalogue are more correlated with the subsequent occurring intermediate to large earthquakes. In contrast,  $LR^i < 1$  means that the patterns from the random catalogue are correlated with the future seismicity. Due to a rather skewed distribution of  $LR^i$ , the mean value  $\langle LR^i \rangle$  is not an appropriate measure for the spatial correlations. Instead, we use the number  $N_s^i$  of APCs that is a better fit than the original model ( $LR^i < 1$ ) and represents a more robust

**Table 1.** Results of Likelihood Ratio Test.  $N_s$  is the number of adjusted Poisson catalogues, where the GCL patterns are more correlated with main shock activity than for the CNSS catalogue.  $P_{\text{conf}}$  is the probability that nine random numbers (corresponding to the nine strong earthquakes) have a mean value smaller than or equal to  $\langle N_s \rangle$ . The values in the parentheses are the results for  $\langle N_s \rangle$  and  $P_{\text{conf}}$  without the Kern County earthquake

	Earthquake	date	$M$	$N_s$
a.	Kern County	21 Jul 1952	7.5	85
b.	Landers	28 Jun 1992	7.3	34
c.	Hector Mine	16 Oct 1999	7.1	23
d.	Loma Prieta	18 Oct 1989	7.0	16
e.	Coalinga	2 May 1983	6.7	26
f.	Northridge	17 Jan 1994	6.6	62
g.	San Fernando	9 Feb 1971	6.6	16
h.	Superstition Hills	24 Nov 1987	6.6	28
i.	Borrego Mountain	9 Apr 1968	6.5	51
	$\langle N_s \rangle$			38 (32)
	$P_{\text{conf}}$			89% (97%)

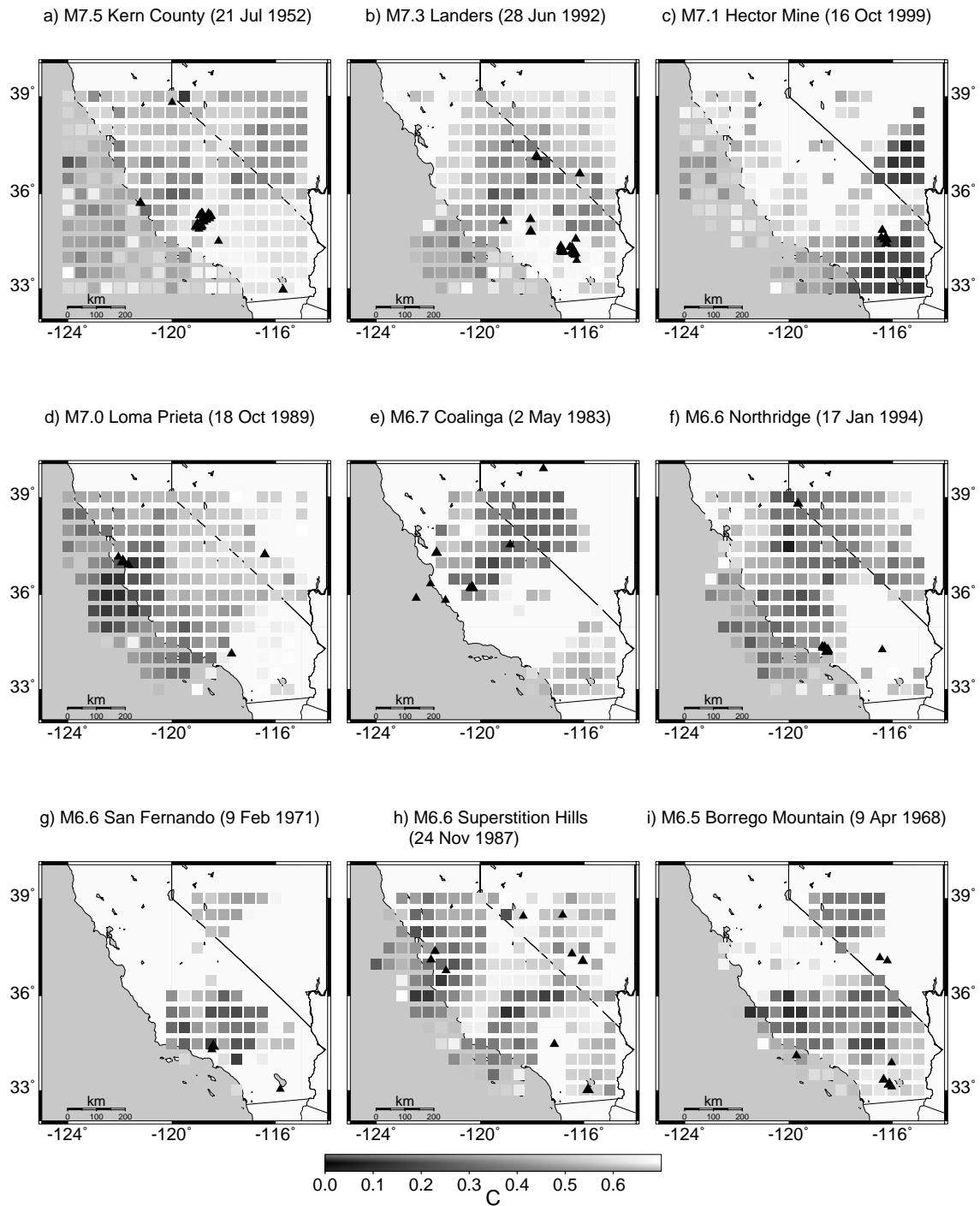
measure. The value of  $N_s^i$  varies between 0 (no APCs fit better than the original model) and 100 (all APCs fit better).

### 3 Results and discussion

Results for the correlation length from Eq. (2) are shown in Fig. 2. The triangles are the earthquakes with  $M \geq 5.0$  occurring during one year after the strong shock with  $M \geq 6.5$  (largest triangle), i.e. the pattern  $Q^i(\mathbf{x})$ . The grey shaded boxes denote the GCL pattern  $C^i(\mathbf{x})$ . Analogously, Fig. 3 is the same for a catalogue from the Poisson process model. Curvature parameters above 0.7 are not shown, since power laws and constant functions are no longer distinguishable.

The likelihood ratio test introduced in Sect. 2 is now applied to compare the patterns  $C^i(\mathbf{x})$  and  $C^i_{\text{APC}}(\mathbf{x})$  with the pattern  $Q^i(\mathbf{x})$ . The quantity  $N_s^i$  ( $0 \leq N_s \leq 100$ ), which is the number of APCs that fit better to  $Q^i(\mathbf{x})$  than the original data, is used as a measure for the predictive power of the GCL pattern in the original catalogue before a certain strong earthquake. Note that we do not introduce alarm conditions using threshold values. The results for  $N_s^i$  are given in Table 1. The confidence level  $p_{\text{conf}}$  in the last row is the probability that nine random numbers (corresponding to the nine strongest earthquakes) have a mean value smaller than or equal to  $\langle N_s \rangle = (1/9) \sum_i N_s^i$ .

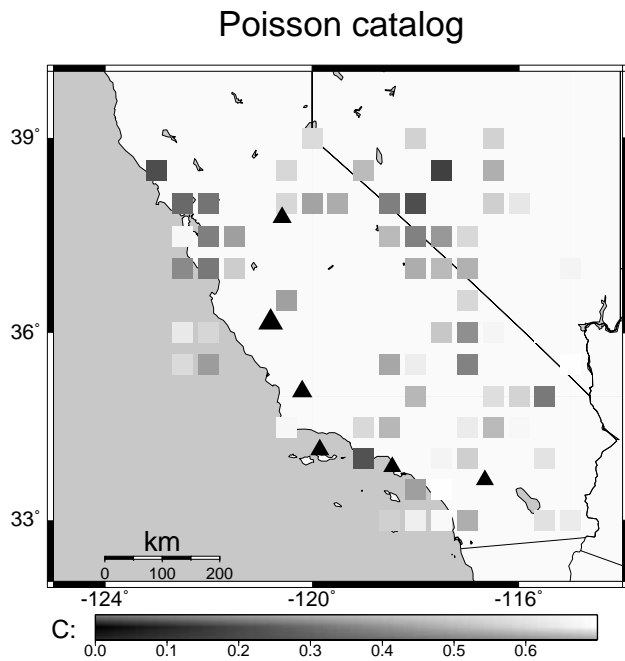
The spatial correlations of the GCL patterns with the future seismicity are clearly visible in some cases, e.g. the Hector Mine, the Loma Prieta, the Coalinga, and the San Fernando earthquakes. The most conspicuous anomaly can be observed prior to the Loma Prieta earthquake in Fig. 2d.



**Fig. 2.** Curvature parameter  $C$  (grey shaded boxes) based on the GCL pattern. The filled triangles are the strong earthquakes (largest triangle) and the earthquakes with  $M \geq 5.0$  until one year after these events.

This is probably due to the fact that there had been no other strong earthquake in the Loma Prieta region since 1910 and consequently, the GCL pattern of this event is not disturbed by the overlapping patterns from the other events. In contrast, the result for the Kern County earthquake is close to a random response. A possible explanation is that the quality and the length of the data may not be sufficient prior to 1952.

As we have checked, the result for the Kern County event can be slightly improved with a magnitude cutoff of  $M_{\text{cut}} = 4.5$  instead of  $M_{\text{cut}} = 4.0$ . The confidence level  $p_{\text{conf}} = 89\%$  for the nine strongest earthquakes is below the typical confidence levels for statistical hypothesis tests, e.g.  $p = 95\%$ . However, if the Kern County earthquake is excluded from the analysis due to a lack of data quality, we obtain  $\langle N_s \rangle = 32$



**Fig. 3.** Curvature parameter  $C$  (grey shaded boxes) with respect to a strong earthquake ( $M = 7.0$ ) for an adjusted Poisson catalogue (APC). The filled triangles are the earthquakes with  $M \geq 5.0$  until one year after these events.

and a resulting probability of  $p_{\text{conf}} = 97\%$ . In this case, the null hypothesis where the results can be reproduced using a realistic Poisson process model without spatiotemporal correlations is rejected with a reasonable high confidence level.

We want to point out that all parameters in our analysis are fixed empirically or determined by the optimization techniques described in the previous section. This is a first order approach which may ignore important information in the data, leading to small significances. Therefore, it is important to determine parameters by physical conditions, e.g. scaling relations such as  $\log R = c_1 + c_2 M$  with constants  $c_1$  and  $c_2$  for the space window  $R$  (Bowman et al., 1998; Zöller et al., 2001) and  $\log T = c_3 + c_4 M$  with constants  $c_3$  and  $c_4$  for the time window  $T$  (Hainzl et al., 2000), as well as search magnitudes should be introduced in order to increase the significances. This would also be a step towards a prediction algorithm, where a spatiotemporal search for anomalies can be conducted. By introducing threshold values in terms of alarm conditions, an analysis by means of error diagrams (Molchan, 1997) could then be carried out. These refinements and extensions are left for future studies.

#### 4 Summary and conclusions

We have tested the hypothesis that spatial anomalies according to the critical point concept for earthquakes occur before strong earthquakes. Therefore, we have used the growing spatial correlation length as an indicator for critical point behaviour. To reduce the number of free parameters, we have

fixed the magnitude cutoff and the critical exponent by values known from the literature. The remaining parameters, namely, space and time windows have been determined systematically by an optimization technique. From a likelihood ratio test in combination with a sophisticated Poisson process model, we have extracted a statistical confidence level.

By applying a search algorithm in space, we find a rough agreement of the predicted regions with future seismicity. Although false alarms and false negatives are present, the original data provide significantly better results than the Poisson process model. The confidence level of 89% is enhanced by excluding the Kern County (1952) earthquake due to a lack of data quality. Further improvements in both the GCL model itself and the statistical test are possible. In particular, it is desirable to map directly probabilities instead of curvature values. This would allow one to compare the present analysis with similar approaches, especially with models based on accelerating energy/moment release.

In conclusion, we have shown that the critical point concept makes a contribution to the improvement of the seismic hazard assessment. Further studies and applications of the methods are promising to increase the significance of the results.

*Acknowledgements.* This work was supported by the Deutsche Forschungsgemeinschaft (SFB 555 and SCH280/13–1). The earthquake data used in this study have been provided by the Northern California Earthquake Data Center (NCEDC) and include contributions of the member networks of the Council of the National Seismic System (CNSS). To produce the figures, we used the GMT system (Wessel and Smith, 1991).

#### References

- Bowman, D. D., Oullion, G., Sammis, C. G., Sornette, A., and Sornette, D.: An observational test of the critical earthquake concept, *J. Geophys. Res.*, 103, 24 359–24 372, 1998.
- Brehm, D. J. and Braile, L. W.: Intermediate-term earthquake prediction using precursory events in the New Madrid Seismic Zone, *Bull. Seismol. Soc. Am.*, 88, 564–580, 1998.
- Brehm, D. J. and Braile, L. W.: Intermediate-term earthquake prediction using the modified time-to-failure method in southern California, *Bull. Seismol. Soc. Am.*, 89, 275–293, 1999.
- Bufe, C. G. and Varnes, D. J.: Predictive modeling of the seismic cycle of the greater San Francisco Bay region, *J. Geophys. Res.*, 98, 9871–9883, 1993.
- Das, S. and Scholz, C. H.: Theory of time-dependent rupture in the Earth, *J. Geophys. Res.*, 86, 6039–6051, 1981.
- Gross, S. and Rundle, J.: A systematic test of time-to-failure analysis, *Geophys. J. Int.*, 133, 57–64, 1998.
- Gutenberg, B. and Richter, C. F.: Earthquake magnitude, intensity, energy and acceleration, *Bull. Seismol. Soc. Am.*, 46, 105–145, 1956.
- Hainzl, S. and Zöller, G.: The role of disorder and stress concentration in nonconservative fault systems, *Physica A*, 294, 67–84, 2001.
- Hainzl, S., Zöller, G., and Kurths, J.: Similar power laws for foreshock and aftershock sequences in a spring-block model for earthquakes, *J. Geophys. Res.*, 104, 7243–7253, 1999.

- Hainzl, S., Zöller, G., Kurths, J., and Zschau, J.: Seismic quiescence as an indicator for large earthquakes in a system of self-organized criticality, *Geophys. Res. Lett.*, 27, 597–600, 2000.
- Jaumé, S. C. and Sykes, L. R.: Evolving towards a critical point: A review of accelerating seismic moment/energy release prior to large and great earthquakes, *Pageoph*, 155, 279–306, 1999.
- Leckie, F. A. and Hayhurst, D. R.: Constitutive equations for creep rupture, *Acta Metall.*, 25, 1059–1070, 1977.
- Molchan, G. M.: Earthquake prediction as a decision-making problem, *Pageoph*, 149, 233–247, 1997.
- Omori, F.: On the aftershocks of earthquakes, *J. Coll. Sci. Imp. Univ. Tokyo*, 7, 111–200, 1894.
- Reasenberg, P.: Second-order moment of central California seismicity, *J. Geophys. Res.*, 90, 5479–5495, 1985.
- Wessel, P. and Smith, W. H. F.: Free software helps map and display data, *Eos Trans., AGU*, 72, 441, 445–446, 1991.
- Zöller, G., Engbert, R., Hainzl, S., and Kurths, J.: Testing for unstable periodic orbits to characterize spatiotemporal dynamics, *Chaos, Solitons and Fractals*, 9, 1429–1438, 1998.
- Zöller, G., Hainzl, S., and Kurths, J.: Observation of growing correlation length as an indicator for critical point behaviour prior to large earthquakes, *J. Geophys. Res.*, 106, 2167–2176, 2001.