Bulletin of the Seismological Society of America, Vol. 96, No. 6, pp. 2465–2471, December 2006, doi: 10.1785/0120060073

Short Note

The Impact of the Spatial Uniform Distribution of Seismicity on Probabilistic Seismic-Hazard Estimation

by Céline Beauval, Sebastian Hainzl, and Frank Scherbaum

Abstract The first step in the estimation of probabilistic seismic hazard in a region commonly consists of the definition and characterization of the relevant seismic sources. Because in low-seismicity regions seismicity is often rather diffuse and faults are difficult to identify, large areal source zones are mostly used. The corresponding hypothesis is that seismicity is uniformly distributed inside each areal seismic source zone. In this study, the impact of this hypothesis on the probabilistic hazard estimation is quantified through the generation of synthetic spatial seismicity distributions. Fractal seismicity distributions are generated inside a given source zone and probabilistic hazard is computed for a set of sites located inside this zone. In our study, the impact of the spatial seismicity distribution is defined as the deviation from the hazard value obtained for a spatially uniform seismicity distribution. From the generation of a large number of synthetic distributions, the correlation between the fractal dimension D and the impact is derived. The results show that the assumption of spatially uniform seismicity tends to bias the hazard to higher values. The correlation can be used to determine the systematic biases and uncertainties for hazard estimations in real cases, where the fractal dimension has been determined. We apply the technique in Germany (Cologne area) and in France (Alps).

Introduction

The first step toward the establishment of a seismic building code in a country is often the evaluation of probabilistic seismic hazard. The method for estimating probabilistic hazard was initiated more than 30 years ago (Cornell, 1968; McGuire, 1976). At a site, estimating hazard in probabilistic terms consists of estimating ground-motion levels that refer to given probabilities of being exceeded at least once over given time periods. Since the 1970s, several variants of the method have been proposed (e.g., Bender and Perkins, 1993; Lapajne *et al.*, 2003) and many aspects of the probabilistic computation have been studied and improved, such as taking into account long-term (e.g., Cramer *et al.*, 2002) and short-term time dependence (Beauval *et al.*, 2006a) in the seismicity models, distributing seismicity in space using kernel smoothing rather than delineating areal seismic source zones (e.g., Frankel, 1995; Cao *et al.*, 1996; Beauval *et al.*, 2006b), including site effects in the probabilistic framework (e.g., Cramer, 2003; Bazzurro and Cornell, 2004) and capturing epistemic uncertainty in inputs through logic trees (e.g., Cramer, 2001). Nonetheless, most probabilistic hazard studies in low-seismicity countries are still based on using seismic zones. The hypothesis dealing with the spatial characterization of seismicity that is still widely used is that seismicity is uniformly distributed inside large areal source zones. Indeed, because low-seismicity regions usually display diffuse seismicity and active faults are very difficult to identify, large areal source zones are defined according to different geophysical and geological criteria (e.g., Autran *et al.*, 1998, for a seismotectonic zoning of the French territory; or Leydecker and Aichele, 1998, for Germany).

Seismicity is a classical example of a complex phenomenon that can be quantified using fractal concepts (Turcotte, 1997). In particular, fault networks and epicenter distributions are known to have fractal properties (Goltz, 1998). Thus, a natural way to analyze the spatial distribution of seismicity is to determine the fractal dimension (D-value). This D-value is an extension of the Euclidean dimension and measures the degree of clustering of earthquakes. In a two-dimensional space, D can be a decimal number and ranges from 0 (point) to 2.0 (uniform distribution in space). This study aims at characterizing the bias in probabilistic hazard estimates resulting from the incomplete knowledge of the degree of clustering of the "true" seismicity distribution. The fractal dimension considered in this study is the correlation dimension (Grassberger and Procaccia, 1983). We first establish a correlation between D and the associated uncertainty on hazard through the generation of synthetic seismicity distributions. Then we apply this approach in two regions of France and Germany and deduce, from the D-value estimations, the corresponding uncertainty bounds on probabilistic hazard estimations.

Synthetic Source Zone: Which Impact for Which D?

The aim is to compare on a grid of sites the hazard estimated from the "true" seismicity distribution to the hazard computed from a uniform smoothing of the seismicity in the source zone, and to link the difference in the hazard estimates to the fractal dimension of the "true" distribution. A quadratic seismic source zone is considered (320 km side length, 5 km unit cell) and the hazard is estimated for a grid of sites located inside the source zone (Fig. 1).

Generation of Synthetic Distributions

Synthetic seismicity distributions are generated over the source zone, increasing the clustering of the seismicity from a line ($D \approx 1.0$) to a uniform distribution over the area ($D \approx 2.0$). Surfaces with a fractal dimension D are generated according to the detailed description in Turcotte (1997):

1. Fourier transformation of a $2L \times 2L$ matrix of random numbers taken from a Gaussian probability distribution.
2. Division of the Fourier coefficients by the radial wavenumber to the power of $4 - D$.
3. Application of the inverse Fourier transform.

To avoid periodicity in the data, the grid is subsequently restricted to size $L \times L$. The fractal surface in the three-dimensional space is then transformed into a two-dimensional data set of fractal dimension $D - 1$ by determining the contours corresponding to the height $z = 0.5\ (z_{\max} - z_{\min})$. To get a continuous probability distribution, we smooth the data set by a Gaussian filter with standard deviation σ and normalize the distribution. Using this probability density function to distribute earthquakes yields theoretically a fractal epicentral distribution of dimension $D_{\text{theory}} = D - 1$. In our computation, $L = 64$ and $\sigma = 1$. Note that the results are not dependent on these values.

Computation of Fractal Dimension

The D-value of each seismicity distribution is estimated, showing that the calculated value is within $D_{\text{theory}} \pm 0.1$. The D-value is computed by estimating the slope of the linear part of the correlation integral (Grassberger and Procaccia, 1983; Goltz, 1998). In a logarithmic scale, the number of pairs of events separated by a distance smaller than r is plotted as a function of r. The curve is established from 3000 events generated over the zone according to the spatial density probability and D is computed over the interval 5–30 km. Because the number of events is high, all the curves display a clear linear part in this interval.

Quantification of the Impact on Hazard of the Uniform Hypothesis

When areal source zones are used in a probabilistic study, the seismicity inside the source zone is gathered and a recurrence curve is determined from this subcatalog of events (evaluation of parameters a and b, i.e., seismicity rate and b-value slope of the recurrence curve). Then this modeled seismicity is uniformly distributed over the source zone. For this purpose the source zone is divided into unit zones; a recurrence curve with the same b-value but with a seismicity rate proportional to the surface of the unit is attributed to each unit. The consequence of such uniform smoothing is that the seismotectonic zoning usually controls the distribution of the final hazard estimates.

A uniform distribution of seismicity over the source zone (i.e., $D = 2.0$) results in identical acceleration values inside the source zone. Distributing the seismicity in a nonhomogenous manner obviously leads to a nonhomogenous estimation of hazard at the sites; the closer the site is to the high seismicity densities the higher is the hazard estimated at this site. The impact on probabilistic hazard is defined as the difference between the acceleration calculated for a spatially uniform seismicity A_{unif} and the estimated one A, normalized by the uniform value and expressed in percentage:

$$I = \frac{A_{\text{unif}} - A}{A_{\text{unif}}} \cdot 100 \tag{1}$$

Therefore, positive impacts correspond to sites where the uniform distribution of seismicity results in an increase of hazard. Note that for very low values of A, the impact is fixed to a 100% value, but as will be shown in the following, this happens only for the lowest values of D (highly clustered seismicity).

Probabilistic Computation

The probabilistic seismic hazard is estimated according to the classical methodology and earthquakes are assumed to follow a Poissonian process in time (Cornell, 1968; McGuire, 1976). An acceleration determined for a return period of, for example, 475 yr has a probability of 10% of being exceeded at least once over a time period of 50 years. Once the spatial density probability distribution is obtained, the seismicity rate is distributed over the source zone. The overall seismicity rate is fixed to 100 events with $M \geq 3.0$ per year. For each unit, the truncated Gutenberg–Richter magnitude recurrence curve is modeled with a slope $b = 1.0$ and a maximum magnitude fixed to M 6.0. Furthermore, the minimum magnitude considered in the probabilistic computation is M 4.5 and the ground-motion predictions of the attenuation relationship are truncated at $+3\sigma$ above the median. The attenuation ground-motion model used here is the Berge-Thierry *et al.* (2003) relationship (the one best adapted to France). Of course, any attenuation relationship

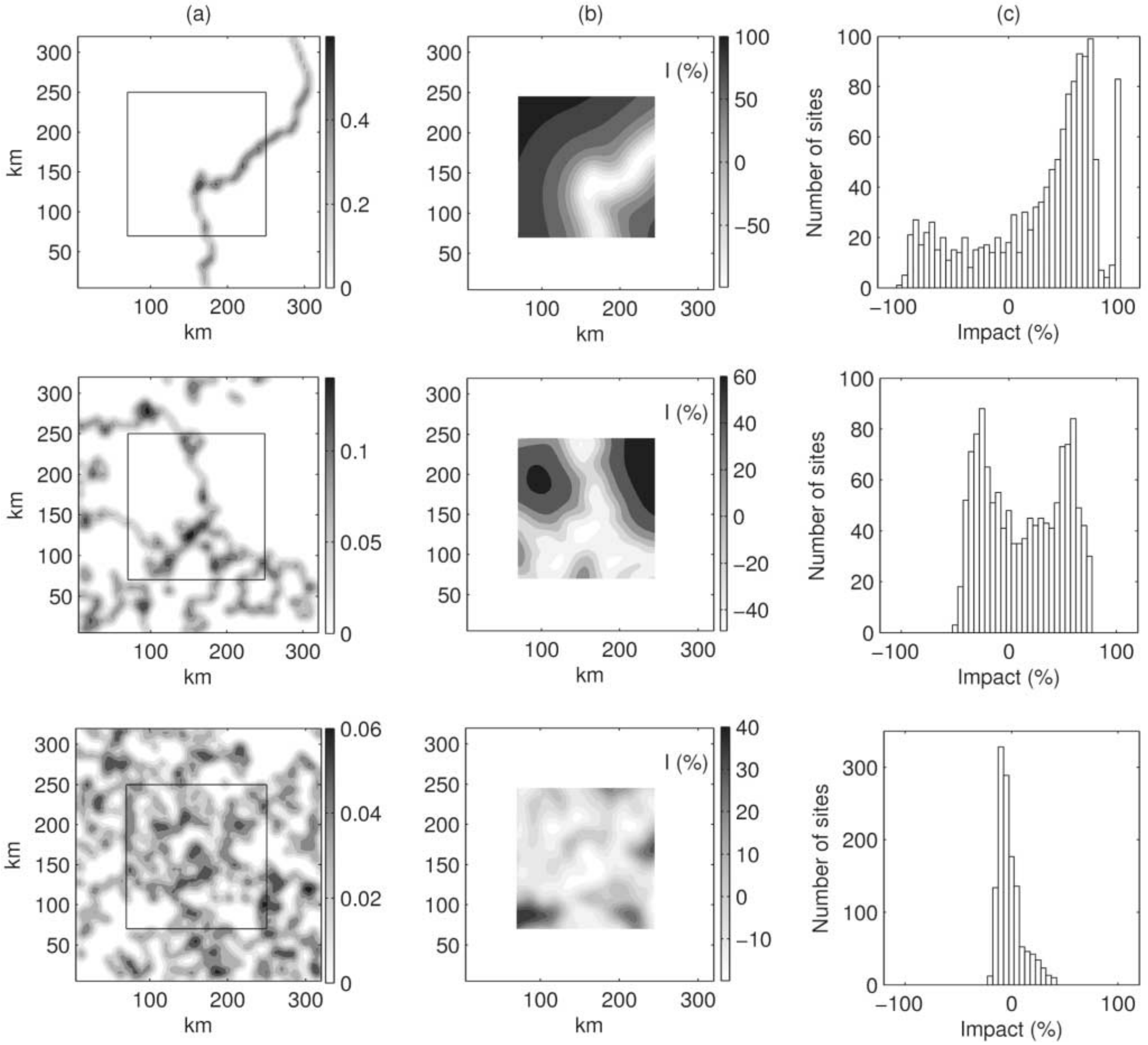

Figure 1. Upper row: $D = 1.05$; middle row: $D = 1.45$, bottom row: $D = 1.76$. (a) Synthetic seismicity distribution, number of $M \geq 3$ per year in 5×5 km^2, square: grid of sites (180 km × 180 km, every 5 km). (b) Impact of uniform hypothesis on probabilistic hazard ($T = 475$ yr), see equation (1). (c) Corresponding distribution of impacts.

could have been used because we deal here with synthetic data. When the seismicity is distributed uniformly, the resulting peak ground acceleration (PGA) is 0.29*g* at 475 yr, which corresponds to sites of highest seismic hazard in countries like France or Germany. However, the results of this study do not depend on absolute values because impacts correspond to a normalized difference with respect to the uniform hazard value.

Results for Example Fractal Distributions

Examples of impact estimation are displayed in Figure 1 for three synthetic seismicity distributions characterized by three different *D*-values (1.05, 1.45, and 1.76, left column). Corresponding impacts on hazard are displayed at the site locations (center column); sites are selected far enough from the border to avoid boundary effect in the hazard estimation. As expected, the spatial pattern of the impact estimates is strongly linked to the spatial distribution of the seismicity. The impact distribution (Fig. 1c) can be considered as the uncertainty distribution for hazard, which characterizes the error resulting from the assumption of uniform seismicity for the complete source zone in an integral way. In this example, a *D*-value of 1.05 leads to impacts on hazard taking values between −100% and +100%. A *D*-value of 1.45

leads to a narrower distribution of impacts: values between -50% and $+75\%$. This distribution is rather bimodal; there are two categories of sites, either close/far to the high-seismicity densities. A *D*-value of 1.76 leads to an even narrower impact distribution, with values between -25% and 45%. As expected, the closer the "true" distribution is to a uniform distribution, the lower are the impacts of the uniform assumption in space. The aim here is to quantify such impacts.

Impact on Probabilistic Hazard Versus *D*-Value of Source Zones

The idea here is to correlate the *D*-value of the source zone to values quantifying the impact of assuming uniform spatial distributions on hazard. Probabilistic hazard is computed for three return periods: 475 yr (the return period of interest for conventional building regulation), 10^4 and 10^5 yr (return periods of interest for special sites such as nuclear power plants).

For a fixed theoretical *D*-value, different spatial patterns and slightly different computed *D*-values may result. Impacts are determined for 30 runs per theoretical *D*-value (corresponding to 30 values in the interval $[D_{\text{theory}} - 0.1\ D_{\text{theory}} + 0.1]$). To characterize each distribution, the percentiles 15% and 85% are selected. Figure 2a displays calculated percentiles versus calculated *D*-values for the return period 475 yr. Subsequently, for each percentile, mean values are determined within a 0.1 interval. The results confirm that the impact distribution narrows and percentiles tend to zero when the *D*-value increases. If the real seismicity embedded in an areal source zone is distributed along a line ($D < 1.2$), the uniform assumption in space implies an increase of the hazard of up to 70% for the 85% percentile (or a decrease of up to 40% for the 15% percentile). Whereas if the seismicity is more diffuse and characterized by a fractal *D*-value of 1.6, then the uniform assumption leads to an increase of hazard values of up to 25%, again for the 85% percentile.

Considering longer return periods (10^4 and 10^5 yr), similar results are observed, with a narrowing of the impact distribution around zero with increasing values of *D* (Fig. 2b). Medians are also displayed for the three return periods (Fig. 2c) showing that whatever the clustering of the seismicity, sites where the uniform distribution of seismicity results in an increase of hazard are more numerous than sites where it results in a decrease. Thus, the uniform distribution of seismicity within a large areal source zone tends to lead to an overestimation of hazard. Note that an increase of the return period leads to an even stronger overestimation: the percentiles shift to higher values.

Application to Two Regions: According to *D*, What Is the Uncertainty on Hazard?

Now we want to quantify the impact of using homogeneous source zones in real cases. For that, we evaluate in the following the *D*-value for two exemplary regions: the Alps at the border between France and Italy, and the Lower Rhine Embayment close to Cologne, Germany.

In the Alps, *D*-values are computed from the instrumental LDG catalog (homogeneous magnitude M_L; Laboratoire de Détection et de Géophysique, Bruyères-le-Châtel, Nicolas *et al.*, 1998). All earthquakes of $M \geq 3.0$ are taken into account during the period 1975–1999. For the Lower Rhine Area, *D*-values are computed from the instrumental part of the German catalog (Leydecker, 2004). All earthquakes of $M \geq 2.5$ during the period 1975–2004 are included. In both cases, other combinations of minimum magnitude of completeness and time periods have been tested,

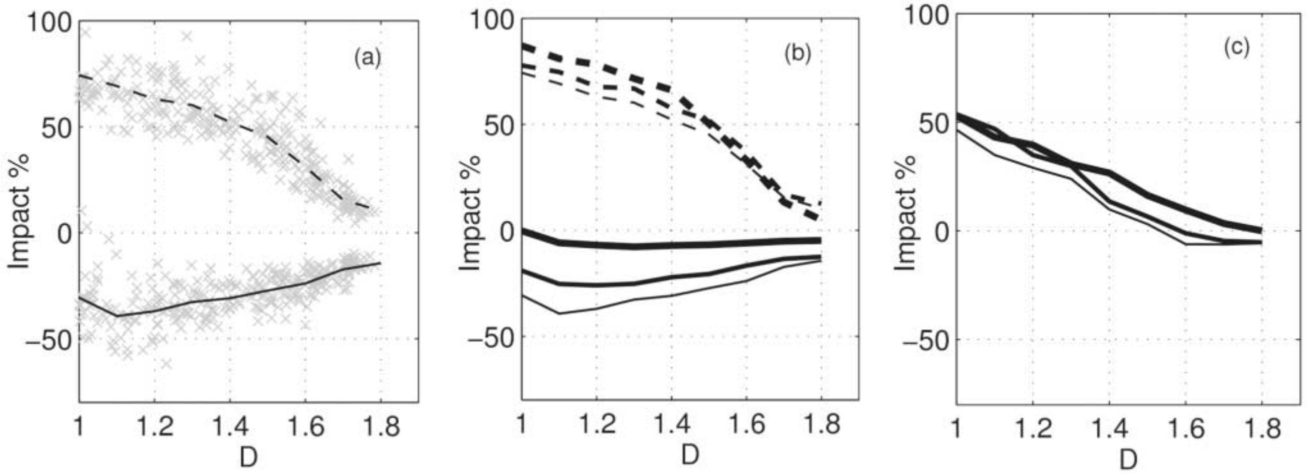

Figure 2. Impacts of the spatial uniform distribution of seismicity on probabilistic hazard. (a) Results at 475 yr; crosses, percentiles 15% and 85% of distributions from 30 runs per *D*-value; curves, mean percentiles computed over 0.1 interval. (b) Percentiles 15% and 85% at 475, 10^4, and 10^5 yr; the thicker the line the longer the return period. (c) Medians of impact distributions.

showing that the *D*-value estimates are rather stable. As depth determinations bear large uncertainties, distances are estimated in two dimensions only.

A spatial mapping of the *D*-value is performed. A grid of 0.1° × 0.1° is defined over the study region, and for each grid point, the *D*-value is computed using all earthquakes falling inside a circle with a fixed radius (70 km). The distance range used is 8–50 km; the minimum distance must be higher than the uncertainty on epicentral location. Results (Figs. 3 and 4) are displayed only if the number of events used is higher than 40 and if the misfit on the *D*-value, ΔD, is lower than 0.1 (least square, norm of residuals). In the Lower Rhine area, the *D*-value ranges between 1.2 and 1.4, whereas in the Alps the *D*-values range between 1.2 and 1.6. Note that a previous study by Sue *et al.* (2002) computed the *D*-value inside two seismic arcs located in the Western Alps. They used a different seismicity catalog and different magnitude ranges; in two dimensions they obtained *D*-values of 1.38 ± 0.05 and 1.4 ± 0.04, which are within the interval obtained in this study.

Seismotectonic zonings used in seismic-hazard studies both in Germany (Leydecker and Aichele, 1998) and in France (Autran *et al.*, 1998) are superimposed on the *D*-value mapping. The Lower Rhine area is roughly characterized by a *D*-value between 1.2 and 1.4; based on the correlation established from synthetics, such fractal dimension indicates an overestimation of hazard inside the source of up to 70% at the 85% percentile, at 475 yr (Fig. 5). In France, considering for example the source zone located between 44.2–45° latitude and 6.3–7.0° longitude, the *D*-value reflects a more diffuse seismicity ($1.4 \leq D \leq 1.6$). This *D*-value indicates up to a 50% increase of hazard (at the 85% percentile) for a site located inside the source zone, when assuming a uniform distribution of seismicity inside the source zone.

Note that the *D*-values rely on instrumental earthquakes only (i.e., recent time periods and predominance of low magnitudes). We have tried to include historical earthquakes in the fractal analysis, but doing this leads to too-low numbers of events to establish correlation integrals. Indeed, taking into account longer time periods (earthquakes prior to the seventies) implies increasing the minimum magnitude for completeness reasons, thus decreasing strongly the total number of events. Therefore, we cannot prove the time stability and magnitude independence of the spatial distribution. However, this assumption seems to be justified by numerous observations showing that fault networks (and therefore earthquakes locations) are self-similar (Turcotte, 1997).

Conclusions

In all probabilistic seismic-hazard studies performed in low-seismicity regions, most seismic sources identified are areal zones, because of the large difficulties in identifying active faults. Recurrence inside the source zone is usually modeled by a Gutenberg–Richter relation calculated from the subcatalog of the zone, and the seismicity rates are uniformly distributed over the source zone for the probabilistic computation. In this study, we show that the impact of the latter assumption can be estimated from the *D*-value of the region of interest. For that, we generate fractal spatial seismicity distributions changing from highly clustered to diffuse characteristics. For each distribution, the impact on hazard of assuming a uniform distribution of the seismicity is

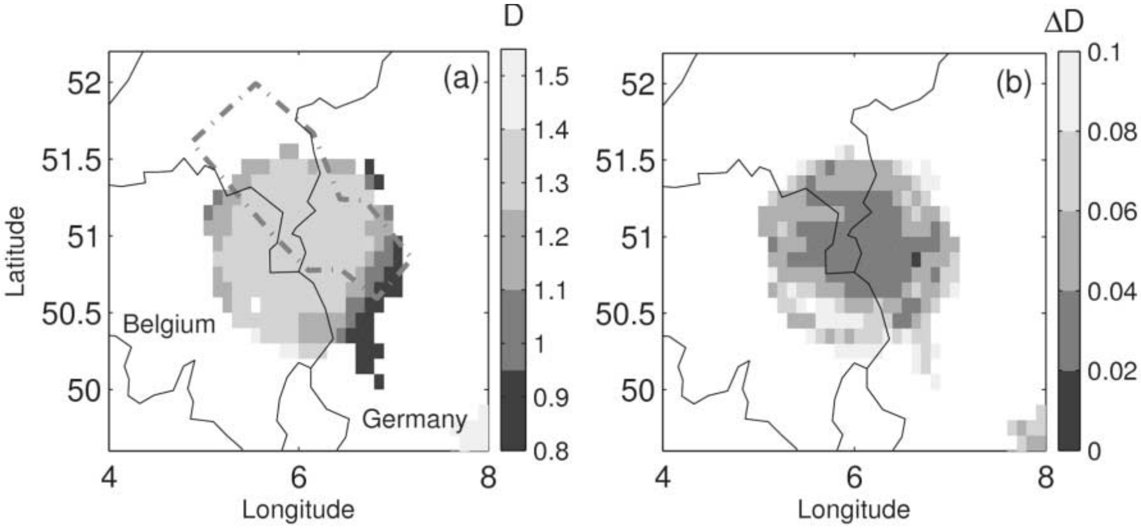

Figure 3. Application in the Lower Rhine Embayment. (a) Mapping of *D*: for each grid point, all earthquakes $M \geq 2.5$ closer than 70 km to the grid point are taken into account, *D* is computed from the correlation integral over the range 8–50 km; dashed lines define the seismotectonic source zone “Lower Rhine area” (zoning by Leydecker and Aichele, 1998). (b) Misfit (norm of residuals, least square), only *D*-values with $\Delta D < 0.1$ are displayed.

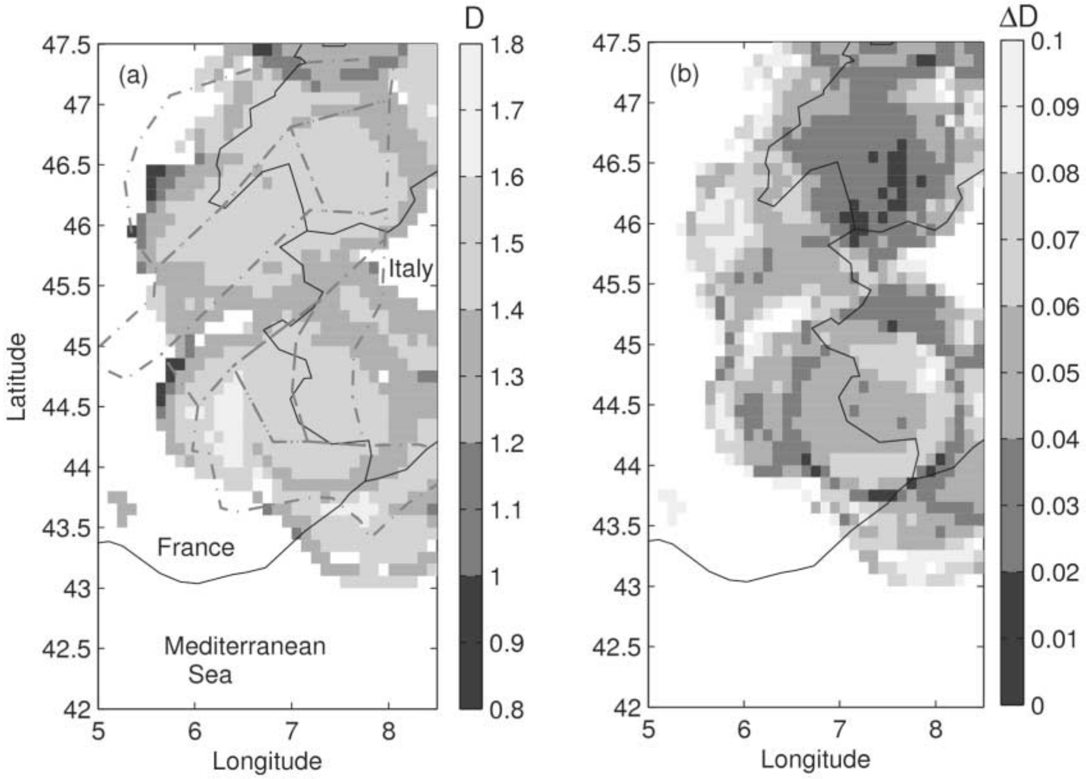

Figure 4. Application in the Alps. (a) Mapping of *D*: for each grid point, all earthquakes $M \geq 3.0$ closer than 70 km to the grid point are taken into account, *D* is computed from the correlation integral over the range 8–50 km; dashed lines define the seismotectonic source zones (zoning by Autran *et al.*, 1998). (b) Misfit, only *D*-values with $\Delta D < 0.1$ are displayed.

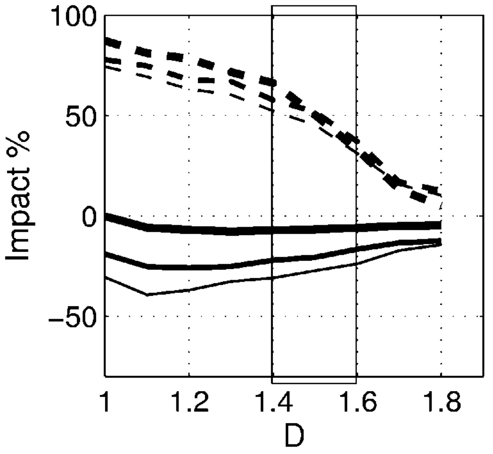

Figure 5. Deduction of the impact of the uniform distribution of seismicity on the probabilistic hazard. Example for a source zone displaying *D*-values around 1.4–1.6. Lines: percentiles 15% and 85% at 475, 10^4, and 10^5 yr; the thicker the line the longer the return period (see legend of Fig. 2).

quantified by comparing the “true” hazard with the uniform hazard value. The results show that, if minimum and maximum bounds for the fractal *D*-value can be estimated for a region, the uncertainties on probabilistic hazard due to the uniform hypothesis in space can be bounded accordingly. A correlation between the impacts on hazard and the *D*-values of the source zones is derived, showing that a uniform distribution of seismicity inside an areal source zone leads on average to conservative hazard estimates. Applying the approach in the Alps and in the Lower Rhine area yields an overestimation of the probabilistic hazard of up to 70% for the 85% percentile (475 yr). The effect is even stronger for longer return periods. More realistic and reliable hazard estimates would be obtained if active faults could be identified, which would require longer instrumental catalogs and more multidisciplinary studies (e.g., geophysical, paleoseismological, historical) in low-seismicity countries such as France or Germany. In the meantime, our results are allowing an adequate correction of systematic errors as well as the incorporation of realistic uncertainty bounds for hazard assessment.

Acknowledgments

We are grateful to the Laboratoire de Détection et de Géophysique (LDG/CEA, Bruyère-le-Châtel) and to the Bundesanstalt für Geowissenschaften und Rohstoffe (BGR, Hannover) for providing us with the seismic catalogs. We thank Klaus-Günter Hinzen (University of Cologne) for numerous stimulating discussions on seismic hazard in the Lower Rhine Embayment. The manuscript has greatly benefited from comments by Ivan Wong and an anonymous reviewer. This work was supported by the Deutsche Forschungsgemeinschaft (SCHE280/14-1).

References

Autran, A., J. L. Blès, P. Combes, M. Cushing, P. Dominique, C. Durouchoux, J. C. Gariel, X. Goula, B. Mohammadioun, and M. Terrier (1998). Probabilistic seismic hazard assessment in France. Part One: Seismotectonic zonation, *Proceedings of the 11th ECEE, 6–11 September, Paris, France*.

Bazzurro, P., and C. A. Cornell (2004). Nonlinear soil-site effects in probabilistic seismic-hazard analysis, *Bull. Seism. Soc. Am.* **96,** 2110–2123.

Beauval, C., S. Hainzl, and F. Scherbaum (2006a). Probabilistic seismic hazard estimation in low-seismicity regions considering non-Poissonian seismic occurrence, *Geophys. J. Int.* **164,** 543–550.

Beauval, C., O. Scotti, and L. F. Bonilla (2006b). The role of seismicity models in probabilistic seismic hazard estimation: comparison of a zoning and a smoothing approach, *Geophys. J. Int.* **165,** 584–595.

Berge-Thierry, C., F. Cotton, O. Scotti, D. A. Griot-Pommera, and Y. Fukushima (2003). New empirical response spectral attenuation laws for moderate European earthquakes, *J. Earthquake Eng.* **7,** 193–222.

Cao, T., M. D. Petersen, and M. S. Reichle (1996). Seismic hazard estimate from background seismicity in Southern California, *Bull. Seism. Soc. Am.* **86,** no. 5, 1372–1381.

Cornell, C. A. (1968). Engineering seismic risk analysis, *Bull. Seism. Soc. Am.* **58,** 1583–1606.

Cramer, C. (2003). Site-specific seismic hazard analysis that is completely probabilistic, *Bull. Seism. Soc. Am.* **93,** 1841–1846.

Cramer, C. H. (2001). The New Madrid seismic zone: capturing variability in seismic hazard analyses, *Seism. Res. Lett.* **72,** no. 6, 664–670.

Cramer, C. H., M. D. Petersen, T. Cao, T. R. Toppozada, and M. Reichle (2002). A time-dependent probabilistic seismic-hazard model for California, *Bull. Seism. Soc. Am.* **90,** 1–21.

Frankel, A. (1995). Mapping seismic hazard in the Central and Eastern United States, *Seism. Res. Lett.* **66,** 8–21.

Goltz, C. (1998). Fractal and chaotic properties of earthquakes, in *Lecture Notes in Earth Sciences*, Springer, New York, 175 pp.

Grassberger, P., and I. Procaccia (1983). Measuring the strangeness of strange attractors, *Physica* **9D,** 189–208.

Lapajne, J., B. S. Motnikar, and P. Zupancic (2003). Probabilistic seismic hazard assessment methodology for distributed seismicity, *Bull. Seism. Soc. Am.* **93,** no. 6, 2502–2515.

Leydecker, G. (2004). Earthquake catalogue for the Federal Republic of Germany and adjacent areas, www.bgr.bund.de (last accessed October 2006).

Leydecker, G., and H. Aichele (1998). The seismogeographical regionalisation for Germany: the prime example of third-level regionalisation, *Geol. Jahr.* **E55,** 85–98.

McGuire, R. K. (1976). Fortran computer program for seismic risk analysis, *U.S. Geol. Surv. Open-File Rept. 76-67*.

Nicolas, M., N. Bethoux, and B. Madeddu (1998). Instrumental seismicity of the Western Alps: a revised catalogue, *Pageoph* **152,** 707–731.

Sue, C., J. R. Grasso, F. Lahaie, and D. Amitrano (2002). Mechanical behavior of western alpine structures inferred from statistical analysis of seismicity, *Geophys. Res. Lett.* **29,** no. 8, 1224, doi 10.1029/2001GL014050.

Turcotte, D. L. (1997). *Fractals and Chaos in Geology and Geophysics*, Cambridge University Press, New York.

Institute of Earth Sciences
University of Potsdam
Postfach 60 15 53
14415 Potsdam, Germany
(C.B., S.H., F.S.)

Géosciences Azur - IRD
250 av. Albert Einstein
Sophia Antipolis
06560 Valbonne, France
beauval@geoazur.unice.fr
(C.B.)

Manuscript received 29 March 2006.