Pure and Applied Geophysics

CrossMark

Interval Estimation of Seismic Hazard Parameters

BEATA ORLECKA-SIKORA[1] and STANISLAW LASOCKI[1]

Abstract—The paper considers Poisson temporal occurrence of earthquakes and presents a way to integrate uncertainties of the estimates of mean activity rate and magnitude cumulative distribution function in the interval estimation of the most widely used seismic hazard functions, such as the exceedance probability and the mean return period. The proposed algorithm can be used either when the Gutenberg–Richter model of magnitude distribution is accepted or when the nonparametric estimation is in use. When the Gutenberg–Richter model of magnitude distribution is used the interval estimation of its parameters is based on the asymptotic normality of the maximum likelihood estimator. When the nonparametric kernel estimation of magnitude distribution is used, we propose the iterated bias corrected and accelerated method for interval estimation based on the smoothed bootstrap and second-order bootstrap samples. The changes resulted from the integrated approach in the interval estimation of the seismic hazard functions with respect to the approach, which neglects the uncertainty of the mean activity rate estimates have been studied using Monte Carlo simulations and two real dataset examples. The results indicate that the uncertainty of mean activity rate affects significantly the interval estimates of hazard functions only when the product of activity rate and the time period, for which the hazard is estimated, is no more than 5.0. When this product becomes greater than 5.0, the impact of the uncertainty of cumulative distribution function of magnitude dominates the impact of the uncertainty of mean activity rate in the aggregated uncertainty of the hazard functions. Following, the interval estimates with and without inclusion of the uncertainty of mean activity rate converge. The presented algorithm is generic and can be applied also to capture the propagation of uncertainty of estimates, which are parameters of a multiparameter function, onto this function.

Key words: Aggregated uncertainty in the activity rate and magnitude, magnitude cumulative distribution function, interval estimation of seismic hazard functions, resampling methods, bootstrap, jackknife.

1. Introduction

The probabilistic seismic hazard is a potential possibility of the occurrence of ground motion caused by seismicity, expressed in the form of likelihoods.

This possibility results from probabilistic properties of the seismic source, propagation of seismic waves from the source to a receiver and receiving site. The probabilistic seismic hazard analysis (PSHA) problem can be presented as

$$\Pr[\text{amp}(x_0, y_0) \geq a(x_0, y_0) \text{ in } D \text{ time units}] = p, \tag{1}$$

where D and p are given values and one looks for the value of amplitude parameter of ground motion, $a(x_0, y_0)$, at the given point (x_0, y_0) whose exceedance probability in D time units is p.

The classic formulation of PSHA assumes that the earthquake occurrence process is Poissonian (e.g. Cornell 1968; Cornell and Toro 1970; Reiter 1991). There are numerous papers indicating the non-Poissonian character of tectonic (e.g. Shlien and Toksöz 1970; Vere-Jones 1970; Kiremidjian and Anagnos 1984; Cornell and Winterstein 1988; Parvez and Ram 1997; Lana et al. 2005; Xu and Burton 2006; Chang et al. 2006; Jimenez 2011; Martin-Montoya et al. 2015) as well as anthropogenic seismic processes (e.g. Lasocki 1992; Weglarczyk and Lasocki 2009; Marcak 2013). However, the classic formulation with Poisson model for earthquake occurrence is still often used (e.g. Petersen et al. 2014, 2015) as it may be appropriate for the cases involving broad average of mixtures of seismic processes. Nonetheless, its practical application should be preceded by a rigorous check of the applicability of Poisson model. When the Poisson model for earthquake occurrences is accepted, the exceedance probability, that is the probability that the amplitude parameter of ground motion will exceed a at (x_0, y_0) in any time intervthe seismic source and pointal of length D time units is:

[1] Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland. E-mail: orlecka@igf.edu.pl; lasocki@igf.edu.pl

Birkhäuser

$$\Pr[\text{amp}(x_0, y_0) \geq a(x_0, y_0)]$$

$$= \int\limits_{0}^{\infty} \int\limits_{M} \Pr[\text{amp}(x_0, y_0) \geq a(x_0, y_0)|r, M] \quad (2)$$

$$f(r)f(M|N(D) \neq 0)\mathrm{d}M\mathrm{d}r,$$

where $N(D)$ is the number of seismic events in the time interval of length D time units, M is the event magnitude, and $f(M|N(D) \neq 0)$ is the probability density function of M, conditional upon the occurrence of seismic events in D, r is the distance between the seismic source and point (x_0, y_0), $f(r)$ is the probability density function of r, and $\Pr[\text{amp}(x_0, y_0) \geq a(x_0, y_0)|r, M]$ is the probability of occurrence at (x_0, y_0) the ground motion amplitude greater than or equal to $a(x_0, y_0)$, when the event of magnitude M is located at the distance r from the point (x_0, y_0). It is also assumed in (2) that the event magnitude and location are independent. The conditional magnitude density reads:

$$f(M|N(D) \neq 0) = -\frac{\mathrm{d}}{\mathrm{d}M}\left[\frac{R(M, D)}{1 - \exp(-\lambda D)}\right], \quad (3)$$

where,

$$R(M, D) = 1 - \exp[-\lambda D(1 - F(M))], \quad (4)$$

$R(M, D)$, referred to as exceedance probability, is the total probability that in D time units there will be events equal to or greater than M, where $F(M)$ is the cumulative distribution function (CDF) of magnitude, λ is the mean activity rate, that is the parameter of Poisson's distribution of earthquake occurrence. $R(M, D)$ represents a potential of the seismic source.

The other function often used to express probabilistic properties of seismic sources when the Poisson model for occurrence is applied, is the reciprocal of the rate of occurrence of earthquakes of magnitude M or greater referred to as the mean return period (e.g. Lomnitz 1974; Baker 2008):

$$T(M) = \{\lambda[1 - F(M)]\}^{-1}. \quad (5)$$

The mean return period is the average recurrence time of events of magnitude M or greater. Both these hazard functions, $R(M, D)$ and $T(M)$, depend on the mean event rate of the Poisson temporal occurrence of earthquakes and the distribution of magnitude. The interval estimation of the CDF of magnitude, $F(M)$ and subsequently the interval estimation of $R(M,$ $D)$ and $T(M)$ where the Poisson occurrence model is accepted but only $F(M)$ uncertainty is taken into considerations have been presented in Orlecka-Sikora (2004, 2008). Extending these works, here we propose a method for the interval estimation of $R(M, D)$ and $T(M)$ functions that accounts for aggregated uncertainty resulting from the uncertainty of the Poisson mean event rate, λ, estimate and the uncertainty of CDF of magnitude, $F(M)$, estimate. On synthetic and actual seismicity cases we analyze improvements introduced by such an integrated approach.

2. Interval Estimation of Seismic Hazard Parameters

When taking into account the aggregated uncertainty in the activity rate and magnitude CDF estimates the confidence intervals (CI) of hazard functions are evaluated on the plug-in curve in the following way:

1. First, the percentiles of the mean activity rate distribution, $\lambda^{(\alpha)}$, are estimated, where α is the percentile order;
2. Next, at each value of M the percentiles of CDF, $F(M)^{(\alpha)}$, are calculated for each α;
3. The values of $R(M, D)$, $T(M)$ (or other hazard functions), are products of all combinations of the percentiles $\hat{\lambda}^{(\alpha)}$ with the percentiles $\hat{F}(M)^{(\alpha)}$;
4. The confidence intervals on the plug-in $\hat{R}(M, D)$, $\hat{T}(M)$ are determined directly from the sorted values of $\hat{R}^k(M, D)$, $\hat{T}^k(M)$ obtained for the particular value of M, where $k = 1, 2, \ldots, l^2$ denotes the combinations of l evenly spaced percentiles λ with l the same percentiles for magnitude CDF. The interval of intended coverage $1 - 2\alpha$ is given by the $\lfloor \alpha \cdot l^2 \rfloor$-th and $\lceil (1 - \alpha) \cdot l^2 \rceil$-th values of the series of $\hat{R}^k(M, D)$, $\hat{T}^k(M)$, where $\lfloor a \rfloor / \lceil a \rceil$ denotes the largest/smallest integer less/greater than or equal to a, respectively.

For the assumed here Poisson earthquake occurrence the mean activity rate estimate is $\lambda = N(D)/D$. In this case, the standard method of confidence interval construction for the Poisson mean is based on inverting an equal tailed test for the null hypothesis $H_0 : \lambda = \lambda_0$ using the exact distribution, e.g. normal.

However, Patil and Kulkarni (2012) present that this approach provides conservative and too wide confidence intervals. After review of the existing methods for obtaining the Poisson confidence intervals they recommend to choose method adjusted to the value of mean activity rate. In the case where the value of mean activity rate is lower than 2 they propose to use one of the following methods:

(a) Modified Wald (Barker 2002):

$$
\text{CI}: \begin{cases} \left[0; -\log\left(\frac{\alpha}{2}\right)\right] & \text{for } x = 0 \\ \left[x + z_{\frac{\alpha}{2}} \cdot \sqrt{x}; x + z_{1-\frac{\alpha}{2}} \cdot \sqrt{x}\right] & \text{for } x > 0 \end{cases} \quad (6)
$$

(b) Wald continuity correction (Schwertman and Martinez 1994):

$$
\text{CI}: \left[(x - 0.5) + z_{\frac{\alpha}{2}} \cdot \sqrt{x - 0.5}; (x + 0.5) + z_{1-\frac{\alpha}{2}} \cdot \sqrt{x + 0.5}\right], \quad (7)
$$

where x is the number of observation in the considered time period, $z_{\frac{\alpha}{2}}$ and $z_{1-\frac{\alpha}{2}}$ are the quantiles of the standard Gaussian distribution of the order $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$, respectively. In the case where the mean activity rate is larger authors suggest to use one of the four methods:

(a) Garwood (1936):

$$
\text{CI}: \left[\frac{\chi^2_{\left(2x, \frac{\alpha}{2}\right)}}{2}; \frac{\chi^2_{\left(2(x+1), 1-\frac{\alpha}{2}\right)}}{2}\right] \quad (8)
$$

(b) Wilson and Hilferty (1931):

$$
\text{CI}: \left[x\left(1 - \frac{1}{9}x + \frac{1}{3}z_{\frac{\alpha}{2}}\sqrt{x}\right); (x + 1) \\ \times \left(1 - \frac{1}{9}(x+1) + \frac{1}{3}z_{1-\frac{\alpha}{2}}\sqrt{x+1}\right)\right] \quad (9)
$$

(c) Molenaar (1970):

$$
\text{CI}: \left[(x - 0.5) + \frac{\left(2z_{\frac{\alpha}{2}}^2 + 1\right)}{6} + z_{\frac{\alpha}{2}}\sqrt{(x - 0.5) + \frac{\left(z_{\frac{\alpha}{2}}^2 + 2\right)}{18}}; (x + 0.5) \\ + \frac{\left(2z_{1-\frac{\alpha}{2}}^2 + 1\right)}{6} + z_{1-\frac{\alpha}{2}}\sqrt{(x + 0.5) + \frac{\left(z_{1-\frac{\alpha}{2}}^2 + 2\right)}{18}}\right] \quad (10)
$$

(d) Begaud et al. (2005):

$$
\text{CI}: \left[\left(\sqrt{x + 0.02} + \frac{z_{\frac{\alpha}{2}}}{2}\right)^2; \left(\sqrt{x + 0.96} + \frac{z_{1-\frac{\alpha}{2}}}{2}\right)^2\right], \quad (11)
$$

where $\chi^2_{(n,\alpha)}$ are the quantiles of the α order of the χ^2 distribution with n degrees of freedom.

We consider two approaches to model the magnitude distribution. First is the most popular exponential magnitude distribution model, which results from the Gutenberg–Richter relation and reads:

$$
f(M) = \beta e^{-\beta(M - M_{\min})}; \ F(M) = 1 - e^{-\beta(M - M_{\min})} \\ M \geq M_{\min}, \quad (12)
$$

$f(M) = F(M) = 0$ for $M < M_{\min}$, $\beta = b \ln 10$, where b is the Gutenberg–Richters' constant and $M_{\min}$ known as magnitude completeness is the lower limit of magnitude of events, which statistically all are present in the analyzed sample of earthquakes. For this model, the interval estimation of its parameter is usually based on the asymptotic normality of the maximum likelihood estimator.

The second approach is applicable to deal with multicomponental seismic processes in which the magnitude distribution does not follow the Gutenberg–Richter relation but is more complex, often multimodal. It is then proposed to use the nonparametric kernel estimation of magnitude distribution (e.g. Lasocki et al. 2000; Kijko et al. 2001; Orlecka-Sikora and Lasocki 2005; Lasocki and Papadimitriou 2006; Lasocki 2008; Quintela-del-Rio 2010; Francisco-Fernandez et al. 2011; Francisco-Fernandez and Quintela-del-Rio 2011). The adaptive kernel estimate of magnitude probability density function (PDF), $f(M)$, is constructed by summing up the Gaussian kernel functions:

$$
\hat{f}^a(M) = \begin{cases} 0 & \text{for } M < M_{\min} \\ \dfrac{\frac{1}{\sqrt{2\pi}}\sum_{i=1}^{n}\frac{1}{\omega_i h}\exp\left[-0.5\left(\frac{M - M_i}{\omega_i h}\right)^2\right]}{n - \sum_{i=1}^{n}\Phi\left(\frac{M_{\min} - M_i}{\omega_i h}\right)} & \text{for } M \geq M_{\min} \end{cases}, \quad (13)
$$

where n is the number of events greater than or equal to $M_{\min}$, M_i are the sizes of these events, $\Phi(\bullet)$

denotes the standard Gaussian cumulative distribution, h is the smoothing factor automatically selected from the data using the least squares cross-validation technique (Silverman 1986). For the Gaussian kernel function and this h selection method it is the root of the equation (Kijko et al. 2001):

$$\sum_{i,j}\left\{2^{-0.5}\left[\frac{(M_i-M_j)^2}{2h^2}-1\right]\exp\left[-\frac{(M_i-M_j)^2}{4h^2}\right]\right.$$
$$\left.-2\left[\frac{(M_i-M_j)^2}{h^2}-1\right]\exp\left[-\frac{(M_i-M_j)^2}{2h^2}\right]\right\}-2n=0.$$
$$(14)$$

The local bandwidth factors $\omega_i, i=1,\ldots,n$ cause the smoothing factor to adapt to uneven data density along the magnitude range. They are estimated as follows

$$\omega_i=\left[\frac{\tilde{f}(M_i)}{g}\right]^{-0.5}, \qquad (15)$$

where $\tilde{f}(\bullet)$ is the pilot, constant kernel estimator

$$\tilde{f}(M)=\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{nh}\sum_{i=1}^{n}\exp\left[-0.5\left(\frac{M-M_i}{h}\right)^2\right], \quad (16)$$

and $g=\left[\prod_{i=1}^{n}\tilde{f}(M_i)\right]^{\frac{1}{n}}$ is the geometric mean of all constant kernel estimates. Such adaptive approach improves effectiveness of the nonparametric estimator in high magnitude intervals where the data are sparse. The corresponding magnitude CDF estimator is:

$$\hat{F}^a(M)=\begin{cases}0 & \text{for } M<M_{\min} \\ \dfrac{\sum_{i=1}^{n}\left[\Phi\left(\frac{M-M_i}{\omega_i h}\right)-\Phi\left(\frac{M_{\min}-M_i}{\omega_i h}\right)\right]}{n-\sum_{i=1}^{n}\Phi\left(\frac{M_{\min}-M_i}{\omega_i h}\right)} & \text{for } M\geq M_{\min}\end{cases}.$$
$$(17)$$

Further details on the nonparametric estimator and its adoption for magnitude distribution estimation are provided in Lasocki et al. (2000), Kijko et al. (2001), and Orlecka-Sikora and Lasocki (2005) and the references therein.

For the nonparametric modeling of magnitude distribution we propose the iterated bias corrected and accelerated method (IBCa method) for interval estimation (Orlecka-Sikora 2004, 2008). This procedure is based on the smoothed bootstrap and second-order

bootstrap samples. The algorithm begins from the so-called bias corrected and accelerated method (BCa method, Efron 1987). The BCa intervals are second-order accurate and transformation respecting (Efron 1987; Efron and Tibshirani 1998). To improve the accuracy of results of the magnitude CDF confidence interval estimation we use of the iterated bootstrap for estimating the bias-correction parameter. According to the iterated BCa method, for any magnitude value the interval of intended coverage $1-2\alpha$ of the nonparametric magnitude CDF is given by:

$$\left(\hat{F}_{\alpha 1}^{a*},\hat{F}_{\alpha 2}^{a*}\right), \qquad (18)$$

where $\hat{F}_{\alpha 1}^{a*}$ and $\hat{F}_{\alpha 2}^{a*}$ are bootstrap estimated percentiles of the distribution of nonparametric magnitude CDF estimator, $\hat{F}^a$. The orders of percentiles, $\alpha 1$ and $\alpha 2$, are calculated from the equations:

$$\alpha 1=\Phi\left(\hat{z}_0+\frac{\hat{z}_0+z_\alpha}{1-\hat{a}(\hat{z}_0+z_\alpha)}\right), \qquad (19)$$

$$\alpha 2=\Phi\left(\hat{z}_0+\frac{\hat{z}_0+z_{1-\alpha}}{1-\hat{a}(\hat{z}_0+z_{1-\alpha})}\right), \qquad (20)$$

where z_α and $z_{1-\alpha}$ are percentiles of the standard Gaussian distribution, $\hat{z}_0$ is the estimate of bias-correction, and $\hat{a}$ is the estimate of the acceleration constant. The bias-correction, z_0, measures the discrepancy between the median of $\hat{F}_i^{a*}$ and $\hat{F}_i^a$, in normal units. According to IBCa method $\hat{z}_0$ is estimated as a mean value of the bootstrap estimates of $\hat{z}_0$, $\hat{z}_0^*$. Each value of $\hat{z}_0^*$ is obtained from the proportion of the second-order bootstrap CDF estimates, $\hat{F}_i^{a**}$, less than the magnitude CDF estimated from the b-th bootstrap data sample, $\hat{F}^a(m)_b^*$, where $b=1, 2, \ldots, B$ and B is the number of the first order bootstrap samples (Orlecka-Sikora 2008):

$$\hat{z}_0^*=\Phi^{-1}\left(\frac{\text{number of II order bootstrap estimates }\hat{F}^a(M)_i^{**}<\hat{F}^a(M)_b^*}{j}\right),$$
$$(21)$$

where $\Phi^{-1}(\bullet)$ indicates the inverse function of the standard Gaussian CDF, j is the number of second-order bootstrap samples drawn from every bootstrap sample and used to estimate the magnitude CDF, $\left\{\hat{F}^a(M)_i^{**}, i=1, 2, \ldots, j\right\}$.

The acceleration constant refers to the rate of change of the standard error of $\hat{F}_i^a$ with respect to the

actual value of magnitude CDF. The acceleration constant can be evaluated in various ways, for instance from the equation (Efron and Tibshirani 1998):

$$\hat{a} = \frac{\sum_{i=1}^{n} \left(\hat{F}^a_{(\bullet)} - \hat{F}^a_{(\text{ijack})} \right)^3}{6 \left\{ \sum_{i=1}^{n} \left(\hat{F}^a_{(\bullet)} - \hat{F}^a_{(\text{ijack})} \right)^2 \right\}^{3/2}}, \qquad (22)$$

where $\hat{F}^a_{(\text{ijack})}$ denotes the i-th jackknife nonparametric estimate of magnitude CDF, and $\hat{F}^a_{(\bullet)}$ is the arithmetic mean of all jackknife estimates.

The bootstrap samples are generated by sampling n-times with replacement from the original data set. Given a data sample $M = \{M_i\}$, $i = 1, 2, \ldots, n$, the bootstrap sample is obtained from the formula:

$$y_i = M'_i + h \cdot \omega_i \cdot \varepsilon, \qquad (23)$$

where M'_i represents the results of resampling with replacement from the original data points, the smoothing factor h is estimated on the basis of the original data sample, the local bandwidth factors ω_i are calculated on the basis of the original data sample for M'_i values, and ε is the standard normal random variable, (Silverman 1986). The i-th jackknife sample is defined as the original sample with the i-th data point removed (Efron and Tibshirani 1998).

To achieve a desired level of accuracy of the quantile level of CI of magnitude CDF the number of bootstrap samples can be calculated using three-step method (Andrews and Buchinsky 2002; Orlecka-Sikora 2008). Further details on the IBCa interval estimation and justification of its use for magnitude CDF estimation when nonparametric approach is applied can be found in the cited works and the references therein.

3. Performance of the Algorithm

The performance of the proposed approach is studied on Monte Carlo generated seismic catalogues linked to three models of magnitude distribution. The functional form of the first two models is the one-side truncated exponential distribution of magnitude, Eq. 12. The parameters for the simulations are: $b = 1.7$ ($\beta = 3.8$), $M_{\min} = 1.1$ for the first model and $b = 0.6$ ($\beta = 1.4$), $M_{\min} = 1.0$ for the second one. An actual example of the first model-like magnitude distribution is the seismic sequence that occurred in connection with a geothermal well in Basel in Switzerland (e.g. Haege et al. 2012; Urban et al. 2015). The second model corresponds for instance to the seismicity triggered by a surface reservoir impoundment of the hydropower plant Song Tranh 2 in Central Vietnam (e.g. Wiszniowski et al. 2015; Urban et al. 2015). The third model is a mixture of two one-side truncated exponential distributions, and reads:

$$f(x) = \begin{cases} \lambda \cdot \beta_1 \cdot e^{-\beta_1 x} & \text{dla} \quad 0 \leq x \leq x_c, \\ \mu \cdot \beta_2 \cdot e^{-\beta_2 x} & \text{dla} \quad x \geq x_c \end{cases}, \qquad (24)$$

where $x = M - M_{\min}$, $x_c = M_c - M_{\min}$, M_c is the magnitude for which the break of linear scaling is observed, $\lambda = \left\{ 1 - \left(1 - \frac{\beta_1}{\beta_2} \right) \cdot e^{-\beta_1 x_c} \right\}^{-1}$, $\mu = \lambda \cdot \frac{\beta_1}{\beta_2} \cdot \frac{e^{\beta_2 x_c}}{e^{\beta_1 x_c}}$. This function models complex magnitude generation processes. The parameters for the simulation are $b_1 = 1.05$ ($\beta_1 = 2.42$), $b_2 = 1.55$ ($\beta_1 = 3.57$), $M_{\min} = 3.5$, $M_c = 5.0$.

From each of these model distributions we draw 50 samples of 50 elements each and 50 samples of 100 elements each. Every sample is used to estimate the cumulative distribution, $F(M)$, and the seismic hazard functions, $R(M, D)$, $T(M)$. The estimation is done by fitting the parametric exponential model, Eq. 12, to data drawn from model 1 and model 2 and using the adaptive nonparametric kernel estimator, Eqs. 13–17, for data drawn from model 3. We use mean activity rate values from the range 0.1–10 events/time unit. In this way, we obtain an opportunity to track scenarios stemming from combinations of: (a) seismic sequence with low activity rate, (b) seismic sequence with high activity rate, (c) seismic sequence with low value of magnitude CDF for the specified M, and (d) seismic sequence with high value of magnitude CDF for the specified M.

In Figs. 1, 2, 3 and 4 the exact values of the exceedance probability, $R(M, D)$, and mean return period function, $T(M)$, are compared with the estimates of 95% CI calculated with and without inclusion of the activity rate uncertainty. The results come from one of the above mentioned 100 and 50 event sample drawn from model 1 and model 2

distributions, respectively, and from 100 event sample drawn from model 3 distribution. The estimation of $R(M, D)$ has been performed for magnitude $M_p = 2.0$ in models 1 and 2 and for $M_p = 4.5$ in model 3.

Figures 5, 6, 7 and 8 show the relative disparity of mean upper/lower bound of 95% CI of the exceedance probability when assuming an aggregated uncertainty of the activity rate and magnitude CDF and when accounting only for CDF uncertainty. The disparity is evaluated by:

$$\delta_{U/L}(M_p) = \frac{\bar{R}_{U/L}(M_p, D) - \bar{R}^B_{U/L}(M_p, D)}{\bar{R}^B_{U/L}(M_p, D)}, \quad (25)$$

where $\bar{R}^B_{U/L}(M_p, D)$ is the mean of 50 estimates of the upper/lower bound of 95% CI of exceedance probability when assuming the aggregated uncertainty, and $\bar{R}_{U/L}(M_p, D)$ is this mean when the mean activity rate estimate is assumed to be error free.

The analysis shows that the uncertainty of mean activity rate affects significantly the interval estimates of hazard functions only when the product λD is small, the activity rate is small and the inference does not concern very long time period D. With increasing λ, the impact of the uncertainty of magnitude CDF dominates the impact of λ uncertainty in the aggregated uncertainty of the hazard functions and the interval estimates with and without inclusion

of the λ uncertainty converge. This agrees with and stems from the functional forms of the hazard functions. For larger M, $1 - F(M)$ tends to zero and for moderate λ and D it dominates λD in the exponent in $R(M, D)$ (see: Eq. 4).

When the λ uncertainty effect is significant, its neglecting results in underestimation of the upper bound of CI of $R(M_p, D)$ and overestimation of its lower bound. Increasing sample size reduces the level of this misestimation. For the same sample size we observe that the effect of λ uncertainty becomes greater for smaller magnitudes. This is due to smaller magnitude CFD uncertainty for smaller magnitudes and hence a reduction of its effect in total uncertainty due to both factors: λ and $F(M)$ (Figs. 5a, 6a, b).

4. Practical Examples

The two considered approaches to CI estimation of hazard functions have been applied to two actual sets of earthquakes related to anthropogenic seismicity accompanying, respectively, (1) underground exploitation of copper ore in the Legnica-Głogów Copper District (LGCD) in Poland (Orlecka-Sikora et al. 2012) and (2) Song Tranh 2 in Vietnam reservoir impoundment (Wiszniowski et al. 2015; Urban et al. 2015).
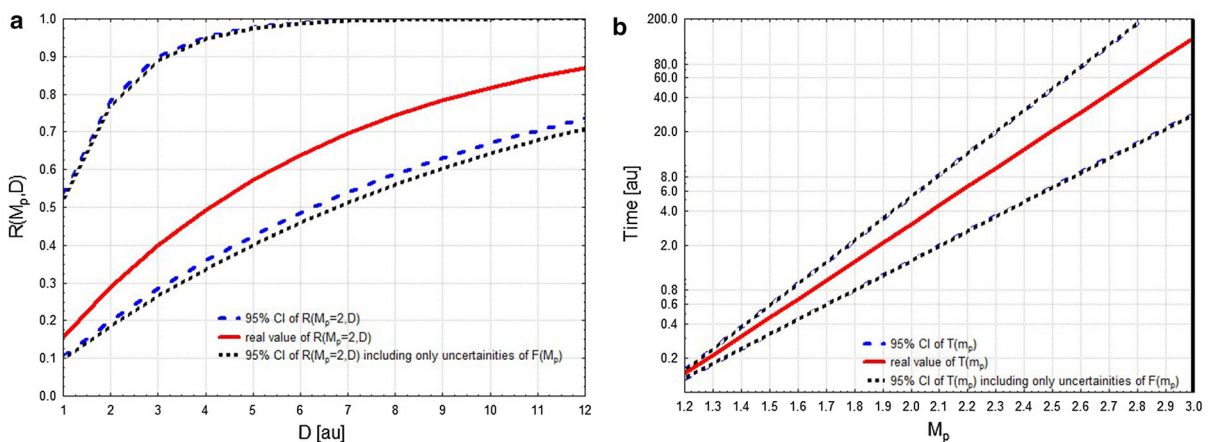

Figure 1
Exact values (*solid red*) and 95% CI-s estimated with (*dashed blue*) and without (*dotted black*) the inclusion of the mean activity rate uncertainty of **a** the exceedance probability of events of magnitude $M_p = 2.0$ and **b** the mean return period function. The results have been obtained for 100 event sample drawn from the model 1 distribution. The mean activity rate has been assumed as 10 events/arbitrary unit
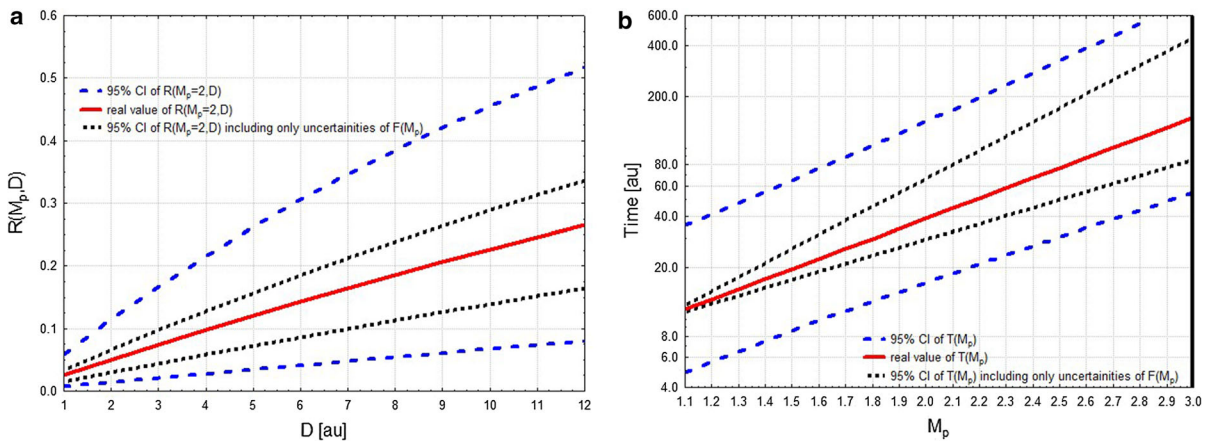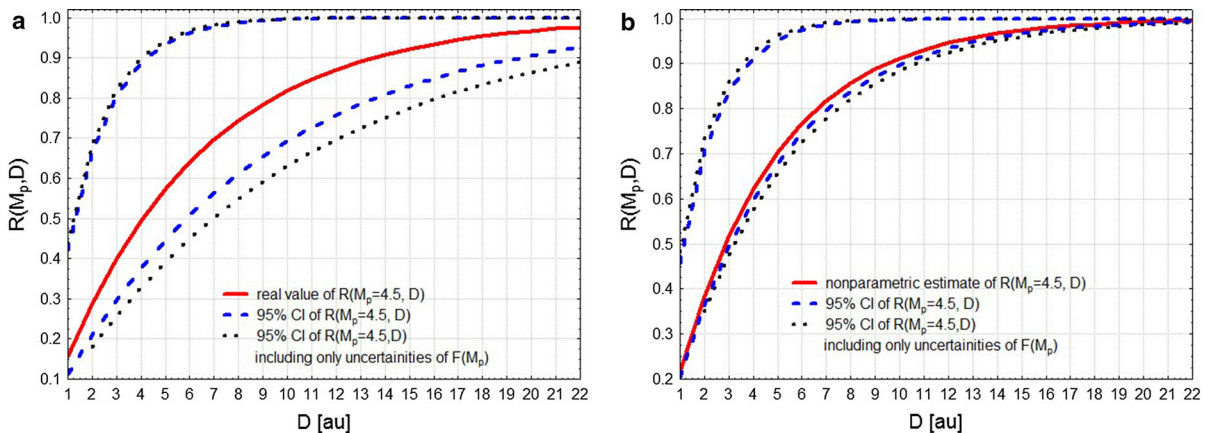
Figure 2
Exact values (*solid red*) and 95% CI-s estimated with (*dashed blue*) and without (*dotted black*) the inclusion of the mean activity rate uncertainty of **a** the exceedance probability of events of magnitude $M_p = 2.0$ and **b** the mean return period function. The results have been obtained for 50 event sample drawn from the model 2 distribution. The mean activity rate has been assumed as 0.1 events/arbitrary unit

Figure 3
Exact values (*solid red*) and 95% CI-s estimated with (*dashed blue*) and without (*dotted black*) the inclusion of the mean activity rate uncertainty of the exceedance probability of events of magnitude $M_p = 4.5$. The results have been obtained for **a** 50 and **b** 100 event sample drawn from the model 3 distribution. The mean activity rate has been assumed as 2.1 and 3 events/arbitrary unit for the **a** and **b**, respectively

The first dataset from LGCD is associated with mining exploitation in section G-11/8 of Rudna mine. An in-mine seismic monitoring system records all events from there of magnitudes 1.2 and more. Mining works in the section G-11/8 began in 2002 and have been continued until present. In this study, 242 seismic events that occurred in the period from 2.01.2004 to 30.12.2005 are analyzed. The strongest tremors of local magnitudes 3.7 and 3.5 took place on 7.01 and 20.01.2005, respectively. The b-value for this dataset is very low, 0.32, and the mean activity rate is 0.3 event/day. Detailed analyses of the

empirical frequency–magnitude relations of the seismicity from the LGCD area revealed that the magnitude distribution did not follow the Gutenberg–Richter relation but had a complex structure (e.g., Orlecka-Sikora and Lasocki 2005; Lasocki and Orlecka-Sikora 2008; Orlecka-Sikora 2008). In such cases, the nonparametric kernel estimator is used to estimate the magnitude CDF. We calculate point and interval estimates of $R(3.0, 30)$ and $T(M)$ in the moving time window of 100 events advancing by one event. For each time window 10,000 bootstrap replicas of the data in the window are used to
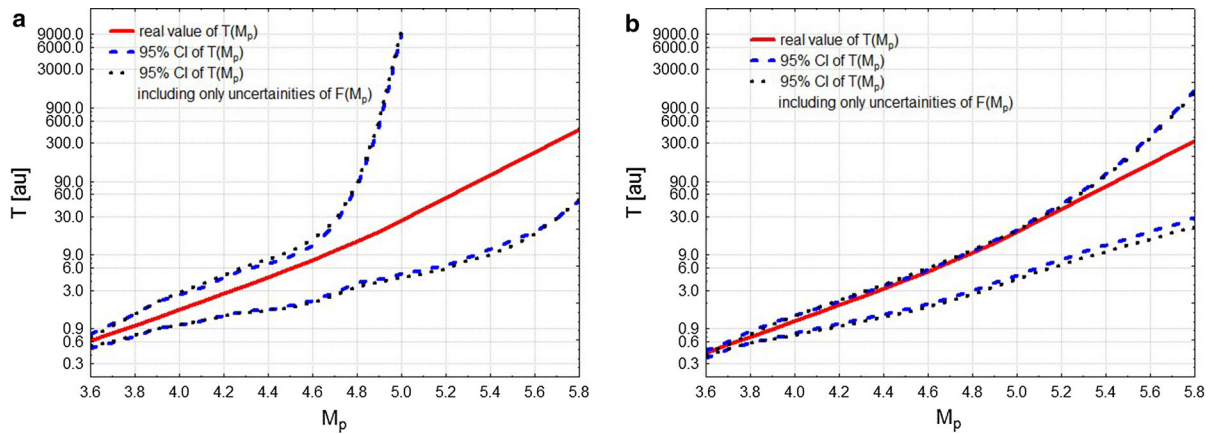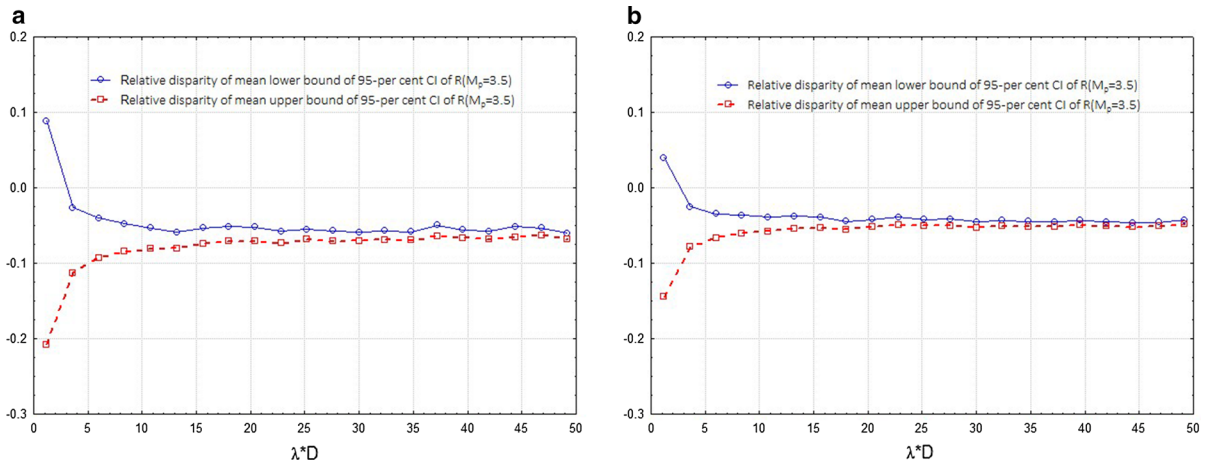
Figure 4
Exact values (*solid red*) and 95% CI-s estimated with (*dashed blue*) and without (*dotted black*) the inclusion of the mean activity rate uncertainty of the mean return period function. The results have been obtained for **a** 50 and **b** 100 event sample drawn from the model 3 distribution. The mean activity rate has been assumed as 2.1 and 3 events/arbitrary unit for the **a** and **b**, respectively

Figure 5
The relative disparity between the mean 95% CI-s of exceedance probability estimated with and without inclusion of the mean activity rate uncertainty. *Red lines* correspond to the upper bound and *blue lines* to the lower bound of CI-s. The calculations have been done for $M_p = 3.5$, $D = 12$ arbitrary units and for λ ranging from 0.1 to 10. The **a** 50 and **b** 100 element magnitude samples have been drawn from model 1 of magnitude distribution with parameters: $\beta = 3.8$, $M_{\min} = 1.1$

evaluate 95% CI of the mentioned hazard functions. The final results of the analysis are shown in Fig. 8. The presented $T(M)$ estimates have been obtained for the window No 20.

Song Tranh 2 dam, the base for the second practical example, locates on the River Song Tranh in Quang Nam province in central Vietnam. The dam was built as a part of hydropower plant. Filling of the reservoir started in November 2010. Up to the beginning of 2011, the seismic activity in this area

increased significantly. Two strongest earthquakes, of magnitudes 4.6 and 4.7, took place on 22nd October and on 15th November 2012, respectively. The seismic activity continues until the present. We analyze a set of 822 earthquakes recorded from 1.09.2012 to 10.11.2014. The range of magnitudes is [1.0; 4.7] and the set is complete. The b-value from the whole set is 0.82, however, Urban et al. (2015) ascertained statistically a highly significant deviation of the observed magnitude distribution from the Gutenberg–
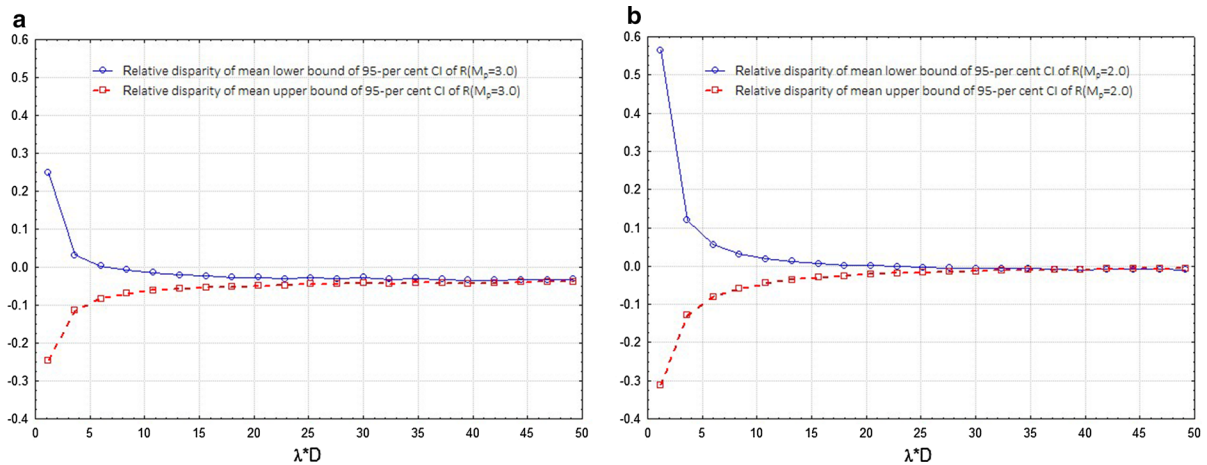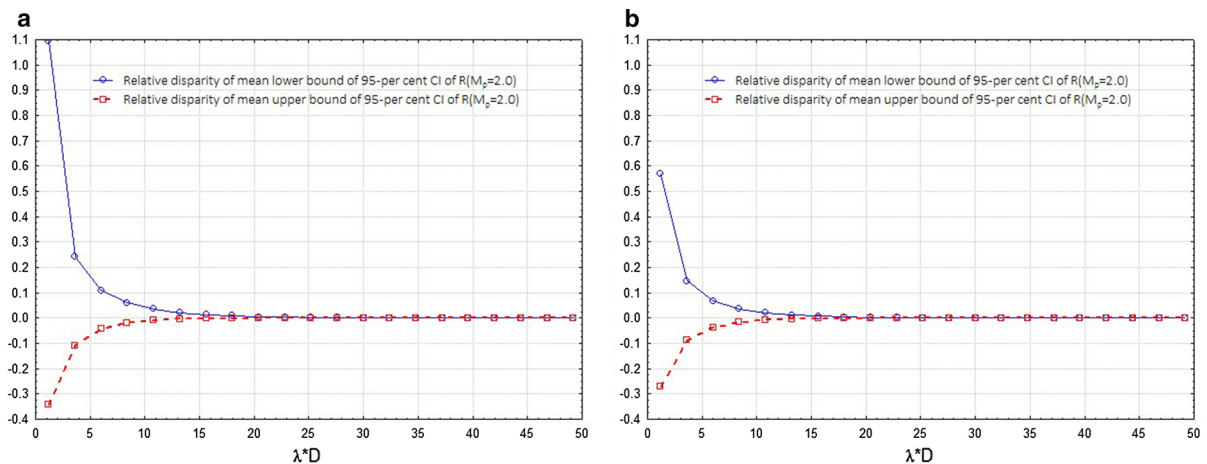
Figure 6

The relative disparity between the mean 95% CI-s of exceedance probability estimated with and without inclusion of the mean activity rate uncertainty. *Red lines* correspond to the upper bound and *blue lines* to the lower bound of CI-s. The calculations have been done for $M_p = 3.0$ (**a**) and for $M_p = 2.0$ (**b**), for $D = 12$ arbitrary units and λ ranging from 0.1 to 10. The 50 element magnitude samples have been drawn from model 1 of magnitude distribution with parameters: $\beta = 3.8$, $M_{\min} = 1.1$

Figure 7

The relative disparity between the mean 95% CI-s of exceedance probability estimated with and without inclusion of the mean activity rate uncertainty. *Red lines* correspond to the upper bound and *blue lines* to the lower bound of CI-s. The calculations have been done for $M_p = 2$, $D = 12$ arbitrary units and λ ranging from 0.1 to 10. The **a** 50 and **b** 100 element magnitude samples have been drawn from model 2 of magnitude distribution with parameters: $\beta = 1.4$, $M_{\min} = 1$

Richter related exponential model, Eq. 12. Therefore, we also use the nonparametric approach to estimate the seismic hazard functions and their uncertainties. We calculate the estimates in the moving time window comprising 200 events and advancing by 10 events. The mean activity rate varies between the windows within the range of 0.6–3.5 event/day. We calculate point and interval estimates of $R(3, 7 \text{ days})$

and $T(M)$. For each time window 10,000 bootstrap replicas of the data in the window are used to evaluate 95% CI of the hazard functions. The results are shown in Fig. 9. The presented mean return period estimates have been obtained for the window No 5.

The first observation drawn from Figs. 8 and 9 is that in both considered cases the exceedance probability considerably varies in time. In the example
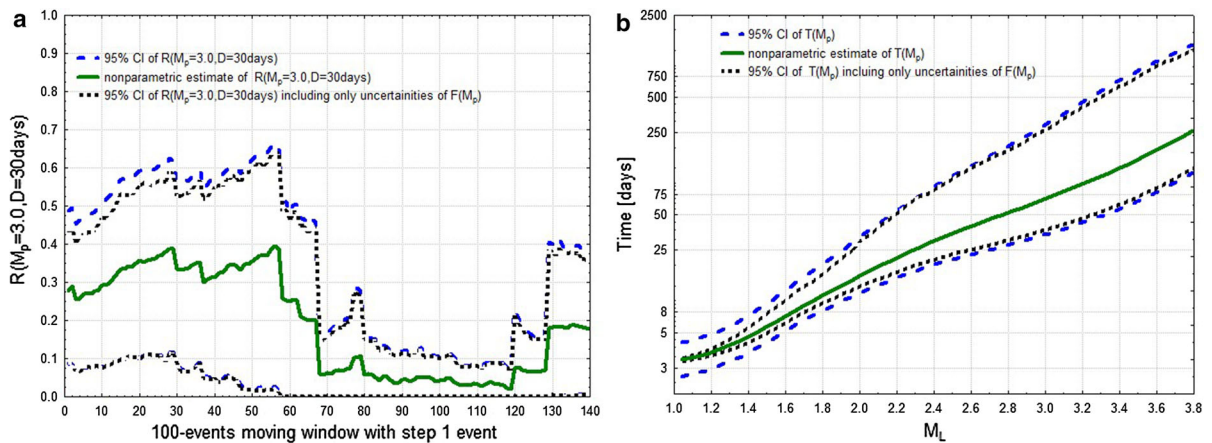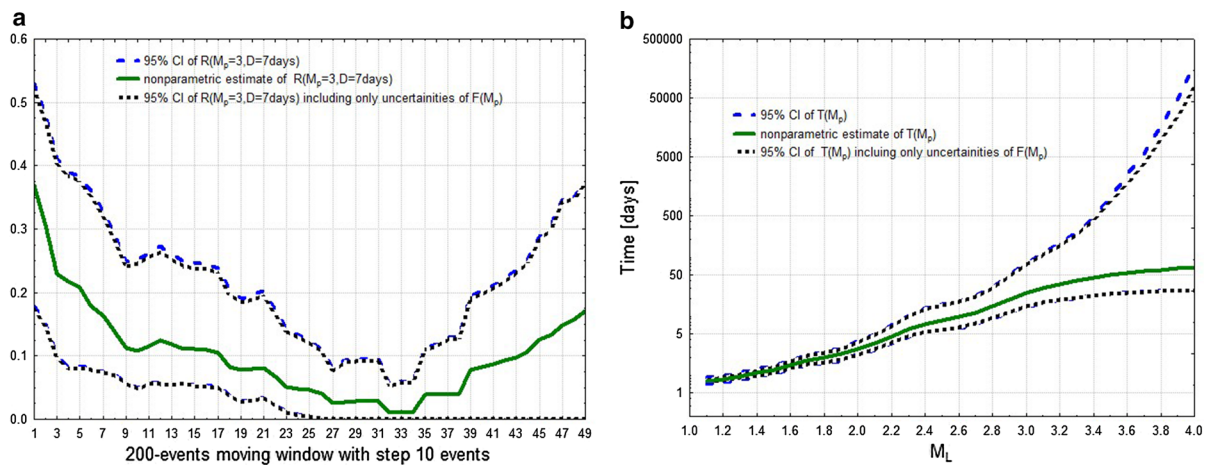
Figure 8

The results for the induced seismicity episode from G11/8 panel in Rudna Mine. **a** Time changes of the estimated exceedance probability, $R(M_p, D)$ for $M_p = 3.0$ and $D = 30$ days calculated in moving time window of 100 events advancing by 1 event. **b** The mean return period estimates for the time window No 20. The mean activity rate for this window is 0.36 events/day. The *solid green lines* represent the point estimates, the *blue dashed lines* represent the 95% CI estimates when the mean activity rate uncertainty has been taken into account and the *black dotted lines* represent the 95% CI estimates when the mean activity rate uncertainty has been neglected

Figure 9

The results for the induced seismicity episode from Song Tranh 2 reservoir. **a** Time changes of the estimated exceedance probability, $R(M_p, D)$ for $M_p = 3.0$ and $D = 30$ days calculated in moving time window of 200 events advancing by 10 events. **b** The mean return period estimates for the time window No 5. The mean activity rate for this window is 1.36 events/day. The *solid green lines* represent the point estimates, the *blue dashed lines* represent the 95% CI estimates when the mean activity rate uncertainty has been taken into account and the *black dotted lines* represent the 95% CI estimates when the mean activity rate uncertainty has been neglected

from Rudna mine during the time of the first 60 windows this probability, that is the seismic hazard, was much stronger than the hazard during the time of the windows from 70 to 120. In the Song Tranh 2 case the hazard was initially quite high and then steadily decreased until the 33-rd time window to increase again within the time period of the last 16 windows.

Second, the confidence intervals are generally wide, that is the uncertainty of hazard functions estimation is considerable. For instance in the Rudna mine case, the point estimate of the exceedance probability of M3 events in a month is 0.4 for the window No 58 and the 95% CI is [0.01, 0.65]. For the Song Tranh 2 case in the window No 1 we have 0.37 for the point estimate and [0.18, 0.52] for 95% CI of

the exceedance probability of M3 events in a month. These results underline the need for interval estimation of hazard functions illustrating how much one can be misled regarding hazard when only point estimates are in hand.

Third, there are no significant differences between the 95% CI estimates including and not including the uncertainty of the mean activity rate, λ. The one visible difference is tiny. The biggest differences between these estimates reach 12-per cent of $R(M_p = 3, D = 30)$ for the Rudna mine case when the time periods with the mean activity rate is the lowest, equal to 0.22–0.28 event per day.

5. Conclusions

We have presented a way to integrate the uncertainty of mean activity rate and magnitude CDF estimates in the interval estimation of the most widely used seismic hazard functions, namely the exceedance probability, $R(M, D)$ and the mean return period, $T(M)$. The proposed algorithm can be used in both situations, either when the parametric model of magnitude distribution is accepted or when the nonparametric estimation is in use. The performance of this algorithm and the changes resulted from this integrated approach with respect to the approach, which neglects the uncertainty of the mean activity rate estimate have been studied on synthetic and actual datasets. The following conclusions can be drawn:

1. Assuming that earthquake occurrences are governed by the Poisson distribution, the algorithm deals with the uncertainty of seismic hazard functions, which depend on the magnitude distribution and the Poisson mean activity rate, both elements being uncertain. However, it is generic, hence can be applied also to capture the propagation of uncertainty of estimates, which are parameters of a multiparameter function, onto this function.

2. Taking into account also the uncertainty of the mean activity rate in the interval estimation of hazard functions makes differences only when the product λD is small, at about 5.0 or less. In such cases, CI of the considered seismic hazard functions should be estimated capturing uncertainty of both their random components: the mean activity rate and magnitude CDF.

3. When λD is bigger, the impact of the uncertainty of magnitude CDF dominates the values of confidence intervals of hazard functions. This results from the particular forms of the hazard functions hence is specific for these functions. In such cases, the uncertainty of λ can be safely neglected.

4. In any case the variance of hazard functions estimates, resulting from the variance of estimates of their components, is significant. Further developments of PSHA should aim at including this source of uncertainty into seismic hazard assessments.

References

Andrews, D. W. K., & Buchinsky, M. (2002). On the number of bootstrap repetitions for BCa confidence intervals. *Econometric Theory, 18*, 962–984.

Barker, L. (2002). A comparison of nine confidence intervals for a Poisson parameter when the expected number of events is ≤ 5. *The American Statistician*, 56(2), 86–89.

Baker, J.W. (2008). An introduction to probabilistic seismic hazard analysis (PSHA). Version 1.3. pp. 72. http://www.stanford.edu/~bakerjw/Publications/Baker_(2008)_Intro_to_PSHA_v1_3.pdf. Accessed 01 Oct 2008.

Begaud, B., Karin, M., Abdelilah, A., Pascale, T., Nicholas, M., Yola, M. (2005). An easy to use method to approximate Poisson confidence limits. *European Journal of Epidemiology, 20*(3), 213–216.

Chang, Y. F., Chen, C. C., & Huang, H. C. (2006). Rescaled range analysis of microtremors in the Yun-Chia area, Taiwan. *Terrestrial Atmospheric and Oceanic Sciences, 17*, 129–138.

Cornell, C. A. (1968). Engineering seismic risk analysis. *BSSA, 58*, 1583–1606.

Cornell, C.A., & Toro, G., (1970). Seismic Hazard Assessment, In R.L. Hunter & C.J. Mann (Eds.), *International association for mathematical geology studies in mathematical geology*, No. 4, techniques for determining probabilities of geologic events and processes (pp. 147–166). Oxford: Oxford University Press.

Cornell, C. A., & Winterstein, S. R. (1988). Temporal and magnitude dependence in earthquake recurrence models. *Bulletin of the Seismological Society of America, 78*, 1522–1537.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of American Statistical Association, 82*(397), 171–200.

Efron, B., & Tibshirani, R. J. (1998). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Francisco-Fernandez, M., & Quintela-del-Rio, A. (2011). Nonparametric seismic hazard estimation: a spatio-temporal application to the northwest of the Iberian Peninsula. *Tectonophysics, 505*(1–4), 35–43.

Francisco-Fernandez, M., Quintela-del-Rio, A., & Casal, R. F. (2011). A nonparametric analysis of the spatial distribution of earthquake magnitudes. *Bulletin of the Seismological Society of America, 101*(4), 1660–1673.

Garwood, F., (1936). Fiducial limits for the Poisson distribution. *Biometrika*, 28(3–4), 437–442. doi:10.1093/biomet/28.3-4.437.

Haege, M., Blascheck, P., & Joswig, M. (2012). EGS hydraulic stimulation monitoring by surface arrays—location accuracy and completeness magnitude: the Basel Deep Heat Mining Project case study. *Journal of seismology, 17*, 51–61. doi:10.1007/s10950-012-9312-9.

Jimenez, A. (2011). Comparison of the Hurst and DEA exponents between the catalogue and its clusters: the California case. *Physica A-Statistical Mechanics and its Applications, 390*, 2146–2154. doi:10.1016/j.physa.2011.01.023.

Kijko, A., Lasocki, S., & Graham, G. (2001). Nonparametric seismic hazard analysis in mines. *Pure and Applied Geophysics, 158*, 1655–1676.

Kiremidjian, A. S., & Anagnos, T. (1984). Stochastic slip-predictable model for earthquake occurrences. *Bulletin of the Seismological Society of America, 74*, 739–755.

Lana, X., Martinez, M. D., Posadas, A. M., & Canas, J. A. (2005). Fractal behavior of the seismicity in the Southern Iberian Peninsula. *Nonlinear Processes in Geophysics, 12*, 353–361.

Lasocki, S. (1992). Non-Poissonian structure of mining induced seismicity. *Acta Montana, 84*, 51–58.

Lasocki, S. (2008). Some unique statistical properties of the seismic process in mines. In Y. Potvin, J. Carter, A. Dyskin, R. Jeffrey (Eds.), Proceedings of the 1st Southern Hemisphere International Rock Mechanics Symp., Vol. 1 (pp. 667–678). Perth: Mining and Civil. **(Australian Centre for Geomechanics, Nedlands, Western Australia).**

Lasocki, S., Kijko, A., & Graham, G. (2000) Model-free seismic hazard estimation, In H. Gokcekus (Ed.), Proc. Int. Conf. Earthquake Hazard and Risk in the Mediterranean Region, EHRMR'99 (Educational Foundation of Near East University, Lefkosa, T. R. N. Cyprus) (pp. 503–508).

Lasocki, S., & Orlecka-Sikora, B. (2008). Seismic hazard assessment under complex source size distribution of mining-induced seismicity. *Tectonophysics, 456*, 28–37. doi:10.1016/j.tecto.2006.08.013.

Lasocki, S., & Papadimitriou, E. E. (2006). Magnitude distribution complexity revealed in seismicity from Greece. *Journal Geophysical Research, 111*, B11309. doi:10.1029/2005JB003794.

Lomnitz, C. (1974). *Global tectonics and earthquake risk*. Amsterdam: Elsevier Sc. Publ. Co.

Marcak, H. (2013). Cycles in mining seismicity. *Journal of Seismology, 17*, 961–974.

Martin-Montoya, L. A., Aranda-Camacho, N. M., & Quimbay, C. J. (2015). Long-range correlations and trends in Colombian seismic time series. *Physica A-Statistical Mechanics and its Applications, 421*, 124–133. doi:10.1016/j.physa.2014.10.073.

Molenaar, W. (1970). Approximations to the Poisson, binomial and hypergeometric distribution functions, *Mathematical Center Tracts* 31, Mathematisch Centrum, Amsterdam.

Orlecka-Sikora, B. (2004). Bootstrap and jackknife resampling for improving in the nonparametric seismic hazard estimation, In Y.T. Chen, G.F. Panza, Z.L. Wu. (Eds.), The IUGG 2003 Proceedings Volume "Earthquake. Hazard, Risk, and Strong Ground Motion" (pp. 81–92). Seismological Press.

Orlecka-Sikora, B. (2008). Resampling methods for evaluating the uncertainty of the nonparametric magnitude distribution estimation in the Probabilistic Seismic Hazard Analysis. *Tectonophys, 456*(1–2), 38–51.

Orlecka-Sikora, B., Lasocki, S. (2005) Nonparametric characterization of mining induced seismic sources, In Y. Potvin, & M. Hudyma (Eds.), The Sixth International Symposium on Rockbursts and Seismicity in Mines "Controlling Seismic Risk" Proceedings (pp. 555–560). Perth: ACG.

Orlecka-Sikora, B., Lasocki, S., Lizurek, G., Rudziński, Ł. (2012). Response of seismic activity in mines to the stress changes due to mining induced strong seismic events. *International Journal of Rock Mechanics and Mining Sciences, 53*, 151–158, 10.1016/j.ijrmms.2012.05.010.

Parvez, I. A., & Ram, A. (1997). Probabilistic assessment of earthquake hazards in the north-east Indian peninsula and Hindukush regions. *Pure and Applied Geophysics, 149*, 731–746.

Patil, V.V., Kulkarni, H.V. (2012). Comparison of confidence intervals for the Poisson mean: some new aspects. *REVSTAT – Statistical Journal, 10*(2), 211–227.

Petersen, M.D., Moschetti, M.P., Powers, P.M., Mueller, C.S., Haller, K.M., Frankel, A.D., Zeng, Y., Rezaeian, S., Harmsen, S.C., Boyd, O.S., Field, N., Chen, R., Rukstales, K.S., Luco, N., Wheeler, R.L., Williams, R.A., & Olsen, A.H. (2014). Documentation for the 2014 update of the United States national seismic hazard maps: U.S. Geological Survey Open-File Report 2014–1091, pp. 243. doi:10.333/ofr20141091.

Petersen, M.D., Mueller, C.S., Moschetti, M.P., Hoover, S.M., Rubinstein, J.L., Llenos, A.L., Michael, A.J., Ellsworth, W.L., McGarr, A.F., Holland, A.A., & Anderson, J.G. (2015).

Incorporating induced seismicity in the 2014 United States National Seismic Hazard Model—Results of 2014 workshop and sensitivity studies: U.S. Geological Survey Open-File Report 2015–1070, pp. 69. doi:10.3133/ofr20151070.

Quintela-del-Rio, A. (2010). On non-parametric techniques for area-characteristic seismic hazard parameters. *Geophysical Journal International, 180*(1), 339–346.

Reiter, L. (1991). *Earthquake hazard analysis*. New York: Columbia University Press.

Schwertman, N.C., Martinez, R. (1994). Approximate Poisson confidence limits. *Communication in Statistics — Theory and Methods*, *23*(5), 1507–1529.

Shlien, S., Toksöz, M.N. (1970). A clustering model for earthquake occurrences. *BSSA*, *60*, 1765–1787.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.

Urban, P., Lasocki, S., Blascheck, P., do Nascimento, A. F., Van Giang, N., & Kwiatek, G. (2015). Violations of Gutenberg–Richter relation in anthropogenic seismicity. *Pure and Applied Geophysics,*. doi:10.1007/s00024-015-1188-5.

Vere-Jones, D. (1970). Stochastic models for earthquake occurrence. *Journal of the Royal Statistical Society: Series B, 32*, 1–62.

Węglarczyk, S., & Lasocki, S. (2009). Studies of short and long memory in mining-induced seismic processes. *Acta Geophysica, 57*, 696–715. doi:10.2478/s11600-009-0021-x.

Wilson, E.B., Hilferty, M.M. (1931). The distribution of chi-square. *Proceedings of the National Academy of Sciences*, *17*, 684–688.

Wiszniowski, J., Van Giang, N., Plesiewicz, B., Lizurek, G., Van Quoc, D., Quang, Khoi L., et al. (2015). Preliminary results of anthropogenic seismicity monitoring in the region of Song Tranh 2 reservoir, Central Vietnam. *Acta Geophysica, 63*(3), 843–862. doi:10.1515/acgeo-2015-0021.

Xu, Y., & Burton, P. W. (2006). Time varying seismicity in Greece: hurst's analysis and Monte Carlo simulation applied to a new earthquake catalogue for Greece. *Tectonophysics, 423*, 125–136. doi:10.1016/j.tecto.2006.03.006.