Bulletin of the Seismological Society of America, Vol. 107, No. 4, pp. 1979–1982, August 2017, doi: 10.1785/0120160303

Reply to "Comment on 'Estimation of Earthquake Hazard Parameters from Incomplete Data Files. Part III. Incorporation of Uncertainty of Earthquake-Occurrence Model' by Andrzej Kijko, Ansie Smit, and Markvard A. Sellevoll" by Gert Zöller

by Andrzej Kijko, Ansie Smit, and Markvard A. Sellevoll

Abstract This reply focuses on comments made by Zöller (2017). We sincerely appreciate the comment by Zöller. This comment and our response creates a perfect opportunity to clarify the controversial issue of whether area-characteristic maximum possible earthquake magnitude m_{max} can be estimated using only the seismic-event catalog. In his comment, Zöller is attempting to convince the reader that based on the seismic-event catalog, the area-characteristic m_{max} cannot be estimated using equation (25) of Kijko *et al.* (2016). In this reply, we will argue the opposite: that with the help of statistical theory, it is possible to assess the m_{max}.

Reality is defined by fact. The fact is that in many instances, an estimation of the maximum possible earthquake magnitude m_{max} is essential. This is especially true for seismic-hazard assessments of critical infrastructures such as bridges, dams, mines, airports, and nuclear facilities. In all these cases, engineers need to have an estimate of m_{max} to allow them to design structures that are both safe and economically viable. A second fact is that the quality and length of earthquake catalogs are not always what we wish them to be, nor do they adhere strictly to the stochastic models we apply to the data. These catalogs are only a small sample of the true population of earthquake events. Aleatory and epistemic uncertainties are therefore an inherent part of any earthquake model or parameter estimation. The goal of parameter estimation is to obtain the best possible estimator based on available sample information, but it should be remembered that it remains only an estimate—a best calculated guess of sorts—of the true value based on available but imperfect data.

One of the tools used to verify fact is numerical simulation (e.g., Monte Carlo simulation). Although Zöller expresses his concern regarding the use of this technique to test the theory, it is a proven and well-used statistical tool used to determine the effectiveness of statistical estimation and tests. If done properly, the simulation process does not lie. Figure 1 supports this by showing how closely it can simulate reality.

Figure 1 shows results of a Monte Carlo simulation of the seismicity that can occur in many seismic active areas. It is assumed that the sought m_{max} is 6.8, the b-value of Gutenberg–Richter is 1.0, and the available seismic-event catalog is complete from magnitude $m_0 = 5.0$. A total of 1000 catalogs were simulated for each of the events indicated in Figure 1. By default, each simulated catalog contains its own m_{max}^{obs}. The first moment (the mean value) of the unknown m_{max} is estimated with equation (1) (equation 25 in Kijko *et al.*, 2016; hereafter, KS estimator) over each set of the 1000 simulated catalogs. As Figure 1 shows, the estimated mean of m_{max} has a negligible error of significantly less than 0.1 magnitude unit. At the point of 100 events in the catalog, the mean square error of estimated m_{max} already does not exceed 0.2 units of magnitude. The simulation results are in full disagreement of Zöller's opinion, that the point estimator of m_{max} cannot be estimated from seismic-event catalog.

Let us address some points and opinions expressed by Zöller in more detail. There are countless estimation procedures that attempt to assess the upper limit of a distribution of stochastic processes. One of these procedures is the estimator used in Kijko *et al.* (2016; their equation 25). This estimator was defined in Kijko (2004) as

$$m_{max} = m_{max}^{obs} + \int_{m_{min}}^{m_{max}} [F_M(m)]^n dm, \quad (1)$$

in which m_{max}^{obs} is the maximum observed magnitude in the catalog and $F_M(m)$ denotes the cumulative distribution function (CDF) of earthquake magnitude. For the doubly truncated frequency–magnitude Gutenberg–Richter relation, the CDF (equation 1) takes the form

$$F_M(m) = \begin{cases} 0 & \text{for } m < m_{min} \\ \frac{1-\exp[-\beta(m-m_{min})]}{1-\exp[-\beta(m_{max}-m_{min})]} & \text{for } m_{min} \le m \le m_{max} \\ 1 & \text{for } m < m_{max} \end{cases} \quad (2)$$

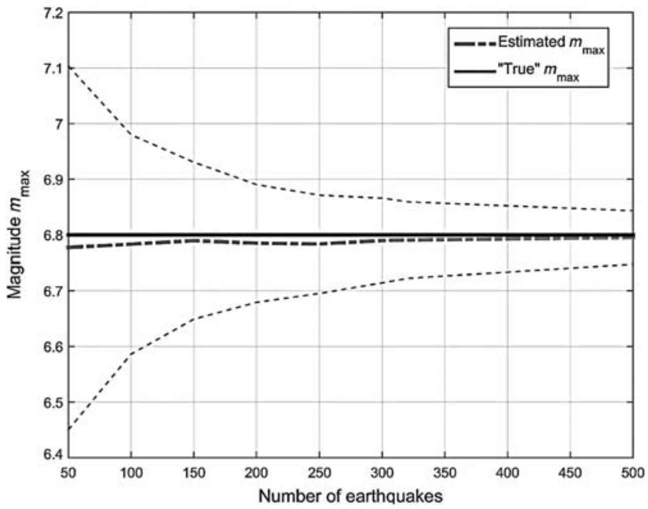

Figure 1. Monte Carlo simulation of the area-characteristic maximum magnitude based on equation (25) of Kijko *et al.* (2016), as well as the respective confidence intervals.

(e.g., Page, 1968), in which $\beta = b\ln(10)$, b is the parameter of the frequency–magnitude Gutenberg–Richter relation, $m_{\min}$ is the level of completeness, and n denotes the number of seismic events equal to or exceeding $m_{\min}$. If the generic estimator (equation 1) is applied to the Gutenberg–Richter CDF of earthquake magnitudes (equation 2), it states that the maximum regional earthquake magnitude $m_{\max}$ is equal to the largest observed magnitude $m_{\max}^{\text{obs}}$, increased by a correction factor $\Delta = \int_{m_{\min}}^{m_{\max}} [F_M(m)]^n dm$. This correction factor depends on the seismic parameters, supporting the expectation that it is always positive and that its value decreases as the time span of the catalog (more precisely, the number of seismic events) increases.

The derivation of equation (1) is based on a generic estimator of the upper bound of a random variable as provided by Cooke (1979). The derivation of the estimator includes analysis of its asymptotic confidence intervals and tests of hypotheses for its bounds. In his subsequent work, Cooke (1980) discusses the problem of the assessment of an upper limit for a random variable when the database is highly incomplete and only a few of the largest observations are known.

An essential part of the derivation of the Cooke's estimator (equation 1) is the replacement of the maximum observed random value (in our case maximum observed magnitude $m_{\max}^{\text{obs}}$) by the expected value $E\left[m_{\max}^{\text{obs}}\right]$ of the sample (equation 4 in Zöller, 2017). The legitimacy of this replacement is questioned by Zöller. Conceptually, such a replacement has its roots in the classic method of moments (MM) estimation of parameters, developed by Pearson (1894). The MM involves equating sample moments with their theoretical counterparts. Because the sample moments are consistent estimates of population moments, the parameter estimates by MM are generally consistent. Because the distribution function of $m_{\max}^{\text{obs}}$ is known and equal to $[F_M(m)]^n$, introducing the condition $m_{\max}^{\text{obs}} = E\left[m_{\max}^{\text{obs}}\right]$ makes it possible to assess all quantiles of the unknown $m_{\max}$. It can be done despite the fact that $F_M(m)$ depends on an unknown $m_{\max}$, because $m_{\max}^{\text{obs}}$ is a sufficient statistic for a sample and any function of $m_{\max}^{\text{obs}}$ is an unbiased estimator of its expectation with the lowest possible variance. Theoretical justification of such an approach can be found in the original paper by Cooke (1979) and in any textbook of theoretical statistics that discusses the Rao–Blackwell theorem. Also, the application of the condition $m_{\max}^{\text{obs}} = E\left[m_{\max}^{\text{obs}}\right]$ for the assessment of the $m_{\max}$ magnitude, including a comprehensive justification of the applied approach, can be found in Pisarenko *et al.* (1996).

It is important to note that any given seismic-event catalog can be only a sample of the true population of earthquake events in history and that the sufficient statistic $E[m_{\max}^{\text{obs}}]$ is only valid for a specific sample of the data at a given point in time. If the catalog is updated, it will constitute a new sample for which a new estimator is required. It can and does happen that a new magnitude value that exceeds the previous $m_{\max}^{\text{obs}}$ is added to the sample, thus updating both $E[m_{\max}^{\text{obs}}]$ and $\hat{m}_{\max}$. The previous estimation of $\hat{m}_{\max}$ and therefore the procedure should not be seen as wrong, but rather that it was the best possible estimation based on available data.

The $\hat{m}_{\max}$ estimator in equation (1), or equation (25) in Kijko *et al.* (2016), cannot be solved without an iteration process, because $m_{\max}$ also appears on the right side of the equation. The integral in equation (1) can be calculated analytically or approximately. In both cases, the estimation of $m_{\max}$ requires iteration. In all of our computational codes, the solution of equation (1) is obtained by utilizing the so-called fixed point iteration (e.g., Hoffman, 2001), functional iterations, or simply the iterative method. The first approximation of $\hat{m}_{\max}$ is obtained by replacing $m_{\max}$ with $m_{\max}^{\text{obs}}$ in two places on the right side of equation (1), in the CDF $F_M(m)$ and in the upper limit of integration. The next approximation of $\hat{m}_{\max}$ is obtained by replacing the $m_{\max}^{\text{obs}}$ with its previous solution. This procedure is repeated until the last correction of $\hat{m}_{\max}$ is small (in our computer code its absolute value ≤ 0.01), or if the number of iterations exceeds 10. In most cases, the procedure converges very fast and does not exceed 10 iterations. An extensive analysis and formal conditions of convergence of the above iterative procedure are discussed, for example, by Legras (1971).

Regarding the statement that the authors are providing only the approximation of the correction factor $\Delta = \int_{m_{\min}}^{m_{\max}} [F_M(m)]^n dm$: following equation (1) and the assumption that earthquake magnitudes follow the doubly truncated Gutenberg–Richter distribution (equation 2), the KS estimator of $m_{\max}$ requires calculation of the integral

$$\Delta = \int_{m_{\min}}^{m_{\max}} \left[\frac{1 - \exp[-\beta(m - m_{\min})]}{1 - \exp[-\beta(m_{\max} - m_{\min})]}\right]^n dm. \qquad (3)$$

The previously mentioned analytical solution of the integral has the closed-form solution as defined by Dwight (1961), but it is neither simple nor easily tractable. This closed-form

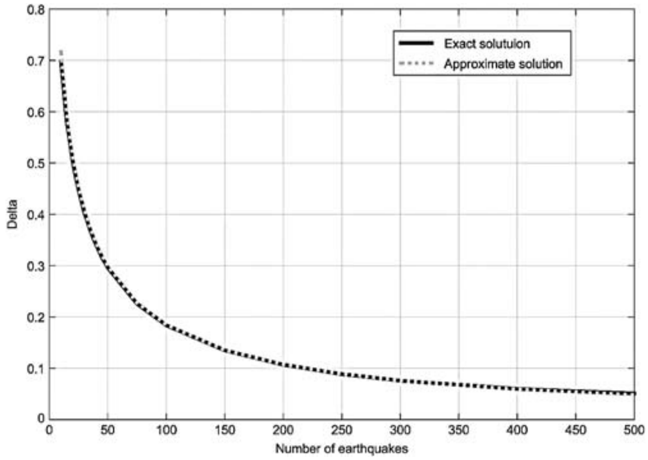

Figure 2. Comparison of the estimation of the correction factor Δ using the exact and the approximated solutions.

solution is provided by equations (15) and (16) in Kijko and Singh (2011; Section 2.1.3: Kijko-Sellevoll Procedure). An alternative solution for the integral (equation 3) can be obtained by applying Cramer's approximation, in which $[F_M(m)]^n \cong \exp[-n(1 - F_M(m))]$. It provides a solution that is no less accurate than the exact one but is compact and easily tractable. Therefore, as a personal preference, we apply the approximation for equation (3) to most of our applications as

$$\Delta = \frac{E_1(n_2) - E_1(n_1)}{\beta \exp(-n_2)} + m_{\min} \exp(-n), \tag{4}$$

in which $n_1 = n/\{1 - \exp(-\beta(m_{\max} - m_{\min}))\}$, $n_2 = n_1 \exp[-\beta(m_{\max} - m_{\min})]$, and $E_1(\cdot)$ denotes an exponential integral function (Abramowitz and Stegun, 1970). The comparison of the two solutions, the exact with the approximate, is provided in Figure 2.

Figure 2 shows that the two solutions for the integral (equation 3) are indistinguishable, even for a very small number of observations ($n \sim 10$). It therefore does not matter which solution, analytical or approximate, you use—the estimates you obtain will be indistinguishable.

It is interesting to note that in 2002, when we started the project to design a toolbox for the assessment of the maximum possible earthquake magnitude (Project Number 691/2002, Council for Geoscience, Pretoria), we had the same doubts that Zöller expressed in his comment: if the application of the Cramer's approximation provided accurate enough values for the integral (equation 3). This question was investigated thoroughly by creating the MATLAB computer code *mmax.m*, which compares 12 different methods of $m_{\max}$ estimation (Kijko and Singh, 2011). Among others, this program and the subsequent publication provides the estimators of $m_{\max}$ per the KS procedure, the first of which based on the exact solution of the integral (equation 3), and the second based on Cramer's approximation. Numerous applications of the KS procedure showed that both versions of procedures provide almost identical results. The *mmax.m* code is freely available from the authors of this note.

Zöller raises two questions regarding the application of the formalism by Cooke (1979) in the KS estimator, namely the range of integration in equation (3) and the application of an analytical form of earthquake magnitude distribution $F(m)$. The exact application of Cooke's formalism requires integration of $[F_M(m)]^n$ in the range $\langle m_{\min}, m_{\max} \rangle$. As mentioned above, this leads to $m_{\max}$ being present in both sides of equation (1). Because $m_{\max}$ is unknown, its assessment can be done only by iteration. To avoid the iteration process, Cooke (1979) replaced $m_{\max}$ by $m_{\max}^{\text{obs}}$. Such replacement simplifies the search procedure but can lead to a slight underestimation of $m_{\max}$. In early applications of the KS procedure (e.g., Kijko and Graham, 1998), Cooke's simplification was used. However, in the more recent works of Kijko (2004), Kijko and Singh (2011), and Kijko *et al.* (2016), the integration of $[F_M(m)]^n$ is performed in the range $\langle m_{\min}, m_{\max} \rangle$ with an iterative search of $\hat{m}_{\max}$. The computer code *mmax.m* provides the user with a choice between the fast and iteration-free approach by Cooke with integration range $\langle m_{\min}, m_{\max}^{\text{obs}} \rangle$, as well as the time-consuming iterative procedure with an integration range $\langle m_{\min}, m_{\max} \rangle$.

Regarding Zöller's comment on the use of the empirical distribution $\hat{F}_n(m)$ compared to the assumed analytical magnitude distribution $F_M(m)$: by its nature, the formalism by Cooke (1979) is generic and applicable to any form of distribution $F(m)$. As long as the assumed functional form of the magnitude distribution is correct, its application can only improve the performance of the estimator, not aggravate it.

As stated by Zöller, the confidence interval of the estimator $\hat{m}_{\max}$ can indeed be infinite, particularly for cases that lack sufficient data for the reliable assessment of $\hat{m}_{\max}$ (Pisarenko, 1991). Infinite confidence intervals are not an uncommon phenomenon and can been seen in the cases of the Pareto distribution (with an infinite variance), the Cauchy distribution (with no finite moments), and even the Gaussian distribution with its infinite tails.

In summary, there is no doubt that it is not easy and not always possible to accurately assess the area-characteristic maximum possible earthquake magnitude $m_{\max}$. Nobody is questioning this. But life teaches us that any extreme viewpoint, in this case the view that we are not able to assess $m_{\max}$, could be dangerous. As evidenced by the above discussion, we believe that if we have enough observations, we can do it. No mathematical or statistical tricks can replace the data, and we have sound mathematical tools that provide the best possible estimates, as well as techniques to assess how reliable our estimates are.

As stated earlier, one of the tools used to determine the reliability of estimates is numerical simulation (e.g., Monte Carlo simulation). Simulation, if done properly, does not lie. Simulations can also reveal if available catalogs contain enough information about $m_{\max}$. Pisarenko (1991) presented it elegantly by introducing special indicator (parameter α), which quantifies the likelihood that, based on the available catalog, one has a chance (or not) to assess m_{max}.

In our opinion, the critical question that remains is what to do when not enough information is available but we still have to provide an estimate of m_{max} for the area. Real-world situations like these are abundant in the design and construction of critical structures. One obvious course of action is to constrain the m_{max} estimator by providing additional and independent information such as regional or local geology, tectonics, paleoearthquakes, capable tectonic faults, seismic history of similar regions, and/or an m_{max} for similar seismogenic regions. This information in conjunction with the available seismic-event catalog will be able to constrain the estimated m_{max}. The combination of different sources of information can easily be done with the application of Bayesian statistics.

Addendum

A comprehensive discussion on virtually of all the concerns raised by Zöller regarding the assessment of the m_{max} estimator, as proposed by Kijko *et al.* (2016), can be found in the two papers by Haarala and Orosco (2016a,b). These papers were made known to the authors after the review of this article.

Data and Resources

The data used in this article was synthetically derived using Monte Carlo simulation of seismicity. It was assumed that $m_{min} = 5.0$, $b = 1$, and $m_{max} = 6.8$. The figures were generated using MATLAB from Mathworks (https://www.mathworks.com/products/matlab.html, last accessed May 2017).

References

Abramowitz, M., and I. A. Stegun (1970). *Handbook of Mathematical Functions*, Ninth Ed., Dover, Mineola, New York.

Cooke, P. (1979). Statistical inference for bounds of random variables, *Biometrika* **66,** 367–374.

Cooke, P. (1980). Optimal linear estimation of bounds of random variables, *Biometrika* **67,** 257–258.

Dwight, H. B. (1961). *Tables of Integrals and Other Mathematical Data*, Third Ed., The Macmillan Co., New York, New York.

Haarala, M., and L. Orosco (2016a). Analysis of Gutenberg–Richter *b*-value and m_{max}. Part I: Exact solution of Kijko–Sellevoll estimator of m_{max}, in *Cuadernos de Ingeniería*, Nueva Serie, Vol. 9, Publicaciones Académicas Fac. Ingeniería, Universidad Católica de Salta, 51–77, available at http://www.ucasal.edu.ar/eucasa/documentos/174-cuaderno-ingenieria-9.pdf (last accessed May 2017).

Haarala, M., and L. Orosco (2016b). Analysis of Gutenberg–Richter *b*-value and m_{max}. Part II: Estimators for *b*-value, in *Cuadernos de Ingeniería*, Nueva Serie, Vol. 9, Publicaciones Académicas Fac. Ingeniería, Universidad Católica de Salta, 79–106, available at http://www.ucasal.edu.ar/eucasa/documentos/174-cuaderno-ingenieria-9.pdf (last accessed May 2017).

Hoffman, J. D. (2001). *Numerical Methods for Engineers and Scientists*, Second Ed., Marcel Dekker, Inc., Basel, Switzerland.

Kijko, A. (2004). Estimation of the maximum earthquake magnitude, m_{max}, *Pure Appl. Geophys.* **161,** 1655–1681.

Kijko, A., and G. Graham (1998). Parametric-historic procedure for probabilistic seismic hazard analysis. Part I: Estimation of maximum regional magnitude m_{max}, *Pure Appl. Geophys.* **152,** no. 3, 413–442, doi: 10.1007/s000240050161.

Kijko, A., and M. Singh (2011). Statistical tools for maximum possible earthquake magnitude estimation, *Acta Geophys.* **59,** 674–700.

Kijko, A., A. Smit, and M. A. Sellevoll (2016). Estimation of earthquake hazard parameters from incomplete data files. Part III. Incorporation of uncertainty of earthquake occurrence model, *Bull. Seismol. Soc. Am.* **106,** 1210–1222.

Legras, J. (1971). *Methodes et Techniques de L'analyse Numerique*, Dunod, Paris, France (in French).

Page, R. (1968). Aftershocks and microaftershocks, *Bull. Seismol. Soc. Am.* **58,** 1131–1168.

Pearson, K. (1894). Contribution to the mathematical theory of evolution, *Phil. Trans. Roy. Soc. Lond. A* **185,** 71–127.

Pisarenko, V. F. (1991). Statistical evaluation of maximum possible earthquakes, *Phys. Solid Earth* **27,** no. 9, 757–763.

Pisarenko, V. F., A. A. Lyubushin, V. B. Lysenko, and T. V. Golubieva (1996). Statistical estimation of seismic hazard parameters: Maximum possible magnitude and related parameters, *Bull. Seismol. Soc. Am.* **86,** 691–700.

Zöller, G. (2017). Comment on "Estimation of Earthquake Hazard Parameters from Incomplete Data Files. Part III. Incorporation of Uncertainty of Earthquake-Occurrence Model" by Andrzej Kijko, Ansie Smit, and Markvard A. Sellevoll, *Bull. Seismol. Soc. Am.* **107,** doi: 10.1785/0120160193.

Department of Geology
University of Pretoria Natural Hazard Centre
University of Pretoria
Private Bag X20, Hatfield
Pretoria 0028
South Africa
andrzej.kijko@up.ac.za
ansie.smit@up.ac.za
(A.K., A.S.)

Department of Earth Science
University of Bergen
Allegaten 41
5007 Bergen
Norway
markvard.sellevoll@broadpark.no
(M.A.S.)

Manuscript received 30 September 2016;
Published Online 13 June 2017