Mak, S., Clements, R., Schorlemmer, D. (2016): Reply to “Comment on ‘Validating Intensity Prediction Equations for Italy by Observations’ by Sum Mak, Robert Alan Clements, and Danijel Schorlemmer” by Mathias Raschke. - *Bulletin of the Seismological Society of America*, *106*, 5, pp. 2414—2415.

DOI: http://doi.org/10.1785/0120160200

THE SEISMOLOGICAL SOCIETY OF AMERICA
400 Evelyn Ave., Suite 201
Albany, CA 94706-1375
(510) 525-5474; FAX (510) 525-7204
www.seismosoc.org

Bulletin of the Seismological Society of America, Vol. 106, No. 5, pp. 2414–2415, October 2016, doi: 10.1785/0120160200

Reply to "Comment on 'Validating Intensity Prediction Equations for Italy by Observations' by Sum Mak, Robert Alan Clements, and Danijel Schorlemmer" by Mathias Raschke

by Sum Mak, Robert Alan Clements, and Danijel Schorlemmer

Abstract Raschke (2016) commented on Mak *et al.* (2015) regarding the use of the performance metric, calculation of the epicentral intensity, and treatment of the focal depths for historical earthquakes when the focal depths have not been estimated in the databases we used. In reply, we present our opinions on these issues. We hope this may help the reader to decide how our study should be interpreted and how future studies of similar kinds should be conducted.

Performance Metric

Raschke (2016) suggested use of the Bayesian information criterion (BIC), a metric that penalizes excessive degrees of freedom of the regression model, to evaluate intensity prediction equations (IPEs). Recently, Roselli *et al.* (2016) also made a similar proposal of using the BIC on ground-motion prediction equation (GMPE) evaluation, although they frankly admitted that GMPE evaluations were seldom conducted on the same set of data used to calibrate the models, and so the strict mathematical meaning for likelihood-based selection criteria (including Akaike information criterion, BIC, and others) to penalize excessive degrees of freedom was not achieved. This situation also applies to our IPE evaluation, even for the retrospective datasets we used ($DBMI11_{RM}$ and $DBMI11_{RL}$; see table 4 of Mak *et al.*, 2015): The retrospective datasets were (at least partially) included in the data used by the modelers to develop their IPEs, but this did not mean that the IPEs were developed using exactly the retrospective datasets. In fact, by only the information published by the modelers, it is usually impossible to reconstruct exactly the dataset the modelers used to calibrate their models.

The use of the likelihood-based selection criteria (including the probabilistic score log likelihood [LLH] presented by Scherbaum *et al.*, 2009) requires models that provide probabilistic predictions. A major reason for us to use the nonparametric score of weighted mean absolute error (wMAE), as explained in the Appropriateness of the Score section of Mak *et al.* (2015), and regarded by Raschke as "not a plausible reason," was that not all IPEs we evaluated have provided the necessary parameters to implement a probabilistic score. Models 1987GPT, 1996P, and 2000P (see table 2 of Mak *et al.*, 2015) did not provide the information on the uncertainty σ of their predictions. When we implemented the LLH (fig. S13 of Mak *et al.*, 2015), we had to assign a σ to the three models. We therefore mentioned the implementation was tentative. Our use of the term "point prediction" simply meant the prediction was a single numeric value, instead of an interval or a distribution. We regret if this has led to any of Raschke's misunderstandings.

Raschke advocated for some conditional advantages of the use of mean square error (MSE) over MAE. Our choice of MAE was not rooted in any theories that are based on specific assumptions. Our choice was purely practical: The results of using MAE were consistent with those of other common metrics (including MSE and LLH), and the results were consistent over multiple datasets. Raschke showed interest in our results of using MSE. We provide here the results of root mean square error (rmse) (Fig. 1), which we produced in our analysis but did not present in Mak *et al.* (2015). It was calculated in the same way as figure 2 of Mak *et al.* (2015), except that the scores for each event were not computed (i.e., not event based). Instead, all data before and including the event indicated by the abscissa were used to calculate the scores. Data were weighted only for the datasets "Hai sentito il terremoto?" (HSIT) and "Did You Feel It?" (DYFI), as described in the Measure-Oriented Evaluation section of Mak *et al.* (2015). Figure 1 does not provide any additional information to what we have already provided in figure 2 of Mak *et al.* (2015), in that the models performing well under the score of wMAE also perform well here; we therefore did not present it in the first place.

We are not surprised by the results presented in the appendix of Raschke (2016). As Raschke has mentioned, there is a close relation between the normal distribution and MSE. The use of MSE on synthetics based on the normal distribution, as Raschke presented in his appendix, is reasonable. For real-world data, mistakes are not uncommon in the data generation process. It is often difficult to completely rule out such mistakes, especially when the data are not created by the user. It is common to use MAE to suppress the effect of such mistakes. We do not disagree with using MSE, especially when the data are carefully prepared such that deterministic outliers are excluded. We acknowledge the theoretical topics that Raschke mentioned. We do not see them as very relevant to our study.

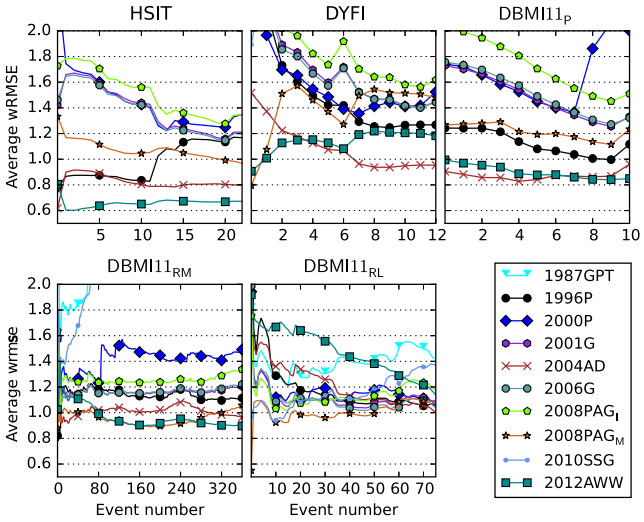

Figure 1. Root mean square error (rmse) of intensity prediction equations. The color version of this figure is available only in the electronic edition.

In summary, Raschke focused on the technical details of parameter estimation of models and the suitability of a metric for in-sample model selection. In our empirical evaluation of IPEs, we were more concerned with the effect of hidden mistakes in the observations and a metric for out-of-sample model evaluation. We also showed consistent results of using different metrics and among multiple datasets. We reiterate here "we focused on the IPE performance on prospective data but also show their performance under retrospective data as reference" (Mak *et al.*, 2015, p. 2946).

Epicentral Intensity and Focal Depth

We here correct a typographic error in our article. We wrote "For the remaining one-third, about 90% have $0 < I_0 - I_{max} \le 1$" (Mak *et al.*, 2015, p. 2945). The correct statement is "For the remaining one-third, about 90% have $0 < I_{max} - I_0 \le 1$," meaning that most of the epicentral intensities in the DBMI04 dataset are within one degree of the maximum intensity. We regret if this typo has led to any of Raschke's misunderstandings.

Raschke commented on the accuracy of the estimation of epicentral intensity I_0, although he has not provided a better alternative of the estimation. We welcome a universally approved accurate estimation of epicentral intensity. A second-best choice is an estimation endorsed by an authority. We did not find any of them when we conducted our study. We opted for another alternative: an estimation that was fair to most (if not all) IPEs under evaluation, regardless of its accuracy. A fair choice ideally means the same choice used by the modelers. We saw that many IPEs under evaluation, which used I_0 as an input parameter, were based on the macroseismic intensity database DBMI04 or older databases that it superseded. The I_0 given in DBMI04 is based on the earthquake catalog CPTI04 (see Data and Resources). In CPTI04, the I_0 of most earthquakes is assigned to be identical to the maximum observed intensity I_{max}, if such I_{max} has been observed at multiple sites. $I_0 \neq I_{max}$ is assigned when I_{max} has been observed at few sites or when other information is available to justify an I_0 different (usually smaller, occasionally larger) than I_{max}.

In CPTI04, whether the uncommon case of $I_0 \neq I_{max}$ should happen involves expert judgment. Therefore, it was difficult for us to follow exactly the I_0 estimation method used in CPTI04. The estimation method we adopted (equation 1 of Mak *et al.*, 2015, equation 5 of Raschke, 2016, or appendix 2 of Gasperini *et al.*, 1999) is conceptually consistent with the explanation given in CPTI04 and does not require expert judgment. Furthermore, the fact that the method is proposed by Paolo Gasperini, the lead editor of CPTI04, is a psychological support for its consistency with the I_0 estimation used in CPTI04.

Raschke advocated for the treatment of epicentral intensity used in Stromeyer and Grünthal (2009), which is entirely about modeling, but not out-of-sample model evaluation. This again shows Raschke's focus on parameter estimation, instead of the empirical evaluation of the predictive power of models, which was the focus of our study.

Raschke questioned our assignment of 10 km to the focal depth of historical earthquakes when the focal depth is not reported, again, without providing a better alternative. Raschke built his argument based on his figure 2, which was based on a fictitious IPE he constructed. We do not see how such a hypothetical model can help with interpreting any parts of our results.

Data and Resources

The description of CPTI04 is given in emidius.mi.ingv.it/CPTI04 (last accessed June 2016).

References

Gasperini, P., F. Bernardini, G. Valensise, and E. Boschi (1999). Defining seismogenic sources from historical earthquake felt reports, *Bull. Seismol. Soc. Am.* **89,** no. 1, 94–110.

Mak, S., R. A. Clements, and D. Schorlemmer (2015). Validating intensity prediction equations for Italy by observations, *Bull. Seismol. Soc. Am.* **105,** no. 6, 2942–2954, doi: 10.1785/0120150070.

Raschke, M. (2016). Comment on "Validating intensity prediction equations for Italy by observations" by Sum Mak, Robert Alan Clements and Danijel Schorlemmer, *Bull. Seismol. Soc. Am.* **106,** no. 5, doi: 10.1785/0120160007.

Roselli, P., W. Marzocchi, and L. Faenza (2016). Toward a new probabilistic framework to score and merge ground-motion prediction equations: The case of the Italian region, *Bull. Seismol. Soc. Am.* **106,** no. 2, 720–733, doi: 10.1785/0120150057.

Scherbaum, F., E. Delavaud, and C. Riggelsen (2009). Model selection in seismic hazard analysis: An information-theoretic perspective, *Bull. Seismol. Soc. Am.* **99,** no. 6, 3234–3247, doi: 10.1785/0120080347.

Stromeyer, D., and G. Grünthal (2009). Attenuation relationship of macroseismic intensities in central Europe, *Bull. Seismol. Soc. Am.* **99,** no. 2A, 554–565, doi: 10.1785/0120080011.

Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ
Helmholtzstraße 6
14467 Potsdam
Germany
smak@gfz-potsdam.de

Manuscript received 29 June 2016;
Published Online 30 August 2016