Revision on: “Enriching the GEOFON seismic catalogue with automatic energy magnitude estimations” by Dino Bindi, Riccardo Zaccarelli, Angelo Strollo, Domenico Di Giacomo, Andres Heinloo, Peter Evans, Fabrice Cotton, and Frederik Tilmann

The paper aims to implement and extend back to 2011 the M_e dataset furnished in real-time by GEOFON from December 2021. The importance of the energy magnitude in relation to the damage is known and having an extended database could be very useful for hazard studies.
In general, the goal is clear but some explanations about the used methodology are necessary to allow the publication.

Line 84: the final M_{ei} for each event is computed as the median over the M_{eij} of each station j. Why is it as computed as the median and not as some kind of average?

Line 96: The anomaly score is here introduced but some explanation of what it is, what the reported values mean, and how it is used to refine the dataset is needed.

Line 99: Why the preferred data set is also the extended one? What does extended mean in this case?

Line 100: From Fig. 3 a) and b) is hard to deduce that the residual analysis is unbiased and a trend is not present. The residual must be averaged over intervals of magnitude and distance (i.e., 0.1 m.u. and 1°) and plotted with the relative s.d. to show the lack of bias.

Line 114: The mixed-effect regression of eq. (2) is underdetermined because the number of unknown coefficients to be determined ($i+j+i \times j+2$) is larger tha the number of equations. I don’t understand how it is possible to obtain all the parameters. The same underdetermination also holds for eq. (3).

Line 116: “*intercept c_1 and slope c_2 parameters define the median model”*. What does it mean? c_1 and c_2 are not parameters obtained from the inversion of a matrix? How the parameter errors are calculated?

Line 127: $\phi = \sqrt{(\phi^2_0 + \phi^2 s)}$ (ϕs is square, check the text), and $\sigma=\sqrt{(\tau^2+\phi^2_0 + \phi^2 s)}$. I don’t understand why to divide in two terms this calculation if it is the same as $\sigma=\sqrt{(\tau^2+\phi^2_0 + \phi^2 s)}=0.407$. Why the term 0.407 is not used anymore and in the relation reported in Fig. 8 the variability is only $\tau=0.246$ but in this case different from the previous one ($\tau=0.27$)?
Eq. 2 allows to calculate Me from Mw, what is the error on Me?

Line 174: “*varying from 0.17 to -0.04 m.u. for M_e vs $M_e(HF)$*”: M_e(HF) is used in place of M_e(BB). As both regressed variables are affected by errors of the same error a general orthogonal regression (GOR; Fuller, 2007; Castellaro et al., 2006), a squared error ratio (η) equal to 1 is more appropriate. What kind of regression was applied? What do the values 0.234 and 0.175 in the regression formulas correspond to? And also, what are the parameter errors?
The scaling of the obtained M_e against SPUD M_e(HF) seems to be close to 1:1. A simple statistical test (Student's t-test) could be useful to show if there is a significative difference from 1 of the slope for M_e(HF) and also for M_e(BB).

Line 200: Like the previous ones, the regressions of the equation (4) between M_e for different faulting styles should be GOR (Fig. 11).

Line 233: Also in this case, a GOR is more appropriate.

Statistical analysis of the difference between the two types of M_e could be useful to conclude that they are the same and the method proposed here could be implemented in real-time in the future providing an extended M_e value dataset compared to the one currently on the GEOFON site.

The caption of Fig. 7: check equation 2 2.

Castellaro, S., F. Mulargia, and Y. Y. Kagan (2006). Regression problems for magnitudes, Geophys. J. Int. 165, 913–930, doi: 10.1111/j.1365- 246X.2006.02955.x.

Fuller, W. A. (1987). Measurement Error Models, John Wiley & Sons Inc., New York, 440 pp.