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 Me 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.

We appreciate the Reviewer's valuable feedback and suggestions. Our responses to each comment are provided below.

Line 84: the final Mei for each event is computed as the median over the Meij of each station j. Why is it as computed as the median and not as some kind of average?

The median is more robust to outliers than the average. However, other options such as trimmed mean (as used in the SeiscomP application) are also possible. As the catalog reports also single station values, the end-user can apply other statistics to compute Me.

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.

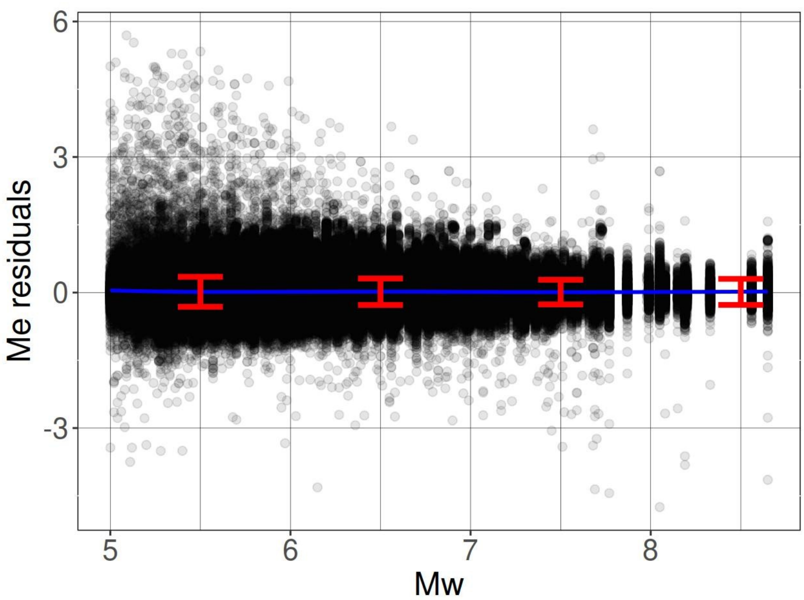
An anomaly score is computed to further refine the data set by flagging anomalous amplitudes using the software sdaas (Zaccarelli et al. 2022). The software, developed from the work of Zaccarelli et al (2021) is based on a machine learning algorithm specifically designed for outlier detection (Isolation forest) which computes an anomaly score in [0, 1], representing the degree of belief of a waveform to be an outlier. The score can be used to assign robustness weights, or to define thresholds above which data can be discarded. We added this sentence to the manuscript around line 96.

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

Following the reviewer's comment, we removed the term 'extended' as it could be misleading. It was used to indicate that D3 is the largest disseminated catalog to which further selections, such as limiting magnitude to values larger than 5.8, were applied. We decided to publish D3 as it is the largest data set after quality checks. We then applied further selections to D3 to produce D6 and flagged the entries corresponding to D6 in D3. As the end-user may wish to apply a different filter, we prefer to disseminate D3 while also specifying our selections for D6.

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.

As stated in the caption of Figure 3, the vertical error bars represent the residuals averaged over 1 m. u. and 20° (*Figure 3. Energy magnitude residuals versus distance (a) and moment magnitude (b) for data set D3. The 90% confidence interval [-0.43,0.50] of the residual distribution is bounded by the horizontal red lines, while the error bars indicate the mean ± 1 standard deviation of the residuals computed over different distance (20° wide) and magnitude (1 m. u. wide) intervals*). The reviewer has requested computing averages over denser intervals. However, increasing the density of the grid will not alter the message, as there are no discernible trends. This is demonstrated in the figure below, where the blue trend line was computed using a localized regression (loess method applied through the function geom_smooth of ggplot2, in R).



Line 114: The mixed-effect regression of eq. (2) is underdetermined because the number of unknown coefficients to be determined (i+j+i x 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).

In mixed effects regressions, the parameters to be determined are the fixed effects (i.e. the model parameters) and the covariance matrix of the random effects, including the variance of the left-over residuals (in the case of equation 2, 5 quantities: c1, c2, tau, phi_S, phi_0). We performed the mixed-effects regression using the standard *lmer* function of R (Bates et al., 2015). For a detailed discussion of the mixed-effects regression and its application to the ground motion variability, please see Stafford (2014) [this reference has been added to the manuscript] and the references therein.

Line 116: “intercept c1 and slope c2 parameters define the median model”. What does it mean? C1 and c2 are not parameters obtained from the inversion of a matrix? How the parameter errors are calculated?

The parameters c1 (intercept) and c2 (slope) are the regression parameters determined through the mixed-effects regressions (the so-called fixed effects); when used for predictive purposes, they define the median model. Errors on the regression parameters are estimated from the asymptotic variances extracted from the covariance matrix of the fit.

Line 127: σ = sqrt(ϕ^2_0 + ϕ^2_S) (ϕ_S is square, check the text), and σ=sqrt(τ^2+ϕ^2_0 + ϕ^2_S). I don’t understand why to divide in two terms this calculation if it is the same as σ=sqrt(τ^2+ϕ^2_0 + ϕ^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 τ=0.246 but in this case different from the previous one (τ=0.27)?

We thank the Reviewer for bringing the missing square to our attention. We corrected the value of tau in line 126.

Eq. 2 allows to calculate Me from Mw, what is the error on Me?

The standard deviation of 0.246 for the between-event residuals (random effects) can be used to quantify the uncertainty of Me from equation 2. It is important to note that due to the simplicity of the linear model and the large population of data used for the regression (~750000 data points), the uncertainty of the median model defined by c1 and c2 is very low. When evaluating the uncertainty of the median model using:

$var[\bar{Me}]_{Mw} = J_o^T [varCov] J_o$ (eq_a)

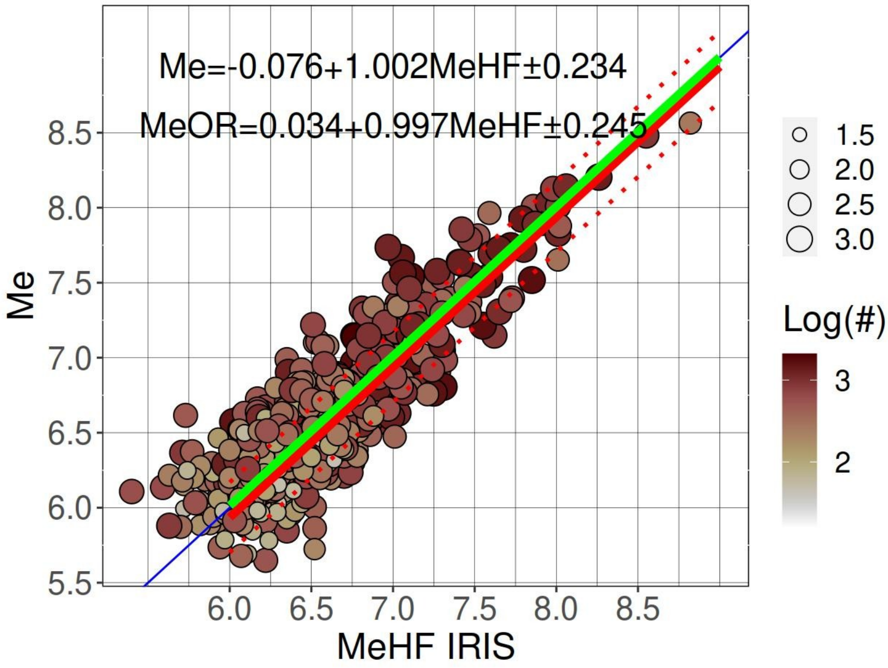
which includes the Jacobian matrix (J_o) and the variance-covariance matrix (varCov), the standard deviation of the variance of Me regression in (eq_a) for Mw=6 and 9 is 0.007 and 0.039, respectively.

Line 174: “varying from 0.17 to -0.04 m.u. for Me vs Me(HF)”: Me(HF) is used in place of Me(BB).

Thanks, corrected.

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 (▬) equal to 1 is more appropriate. What kind of regression was applied?

A robust least squares regression was performed using the rlm function of R. In general, we agree with the Reviewer that total least squares regression (orthogonal regression) could be advisable when accounting for uncertainties on both axes. Considering the high density of points in Figure 9 and assuming equal errors on both axes, the best-fit model obtained by performing an orthogonal regression (using the *odregress* function from the R library *pracma*, represented by the green line) is very similar to the ordinary least squares model (represented by the red line), as shown in the figure below for MeHF. We have chosen to maintain the results obtained with the robust regression.



What do the values 0.234 and 0.175 in the regression formulas correspond to?

They are the standard deviations of the residual distributions.

And also, what are the parameter errors?

We added the errors to our statement of the regression relations.

The scaling of the obtained Me against SPUD Me(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 Me(HF) and also for M e(BB).

For MeHF, a Student's t-test shows that the null-hypothesis that the slope is 1 cannot be rejected at 95% confidence (slope=1.0019, SE=0.0331, DF=363); for MeBB, the null hypothesis can be rejected (slope=0.8958, SE=0.0271, DF=363).

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

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

A mixed-effects regression using maximum likelihood is preferred, as it allows for the introduction of the SOF grouping factor to partition the residuals. Please, also refer to the previous answer for the MeHF regression.

Statistical analysis of the difference between the two types of Me 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 Me value dataset compared to the one currently on the GEOFON site.

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

Thanks, corrected.