In this article, the authors present a methodology to automatically compute Energy Magnitudes (Me). They apply the methodology to the GEOFON catalogue (2011-2023). The authors present several quality checks and statistical analyses of the dataset analyzed. They further compare their Me estimates with those made available by IRIS. The codes used both for off-line and real-time computations are made openly available.
The article if clear and well-written, and represents an important contribution by adding an additional magnitude estimation to the reference Geofon catalog.
We appreciate the Reviewer's valuable feedback and suggestions. Our responses to each comment are provided below.

I suggest only minor revisions below:
Abstract: Maybe add a brief sentence explaining the added value of Me estimates?
We added in the introduction (line 27) the sentence “Me estimates have been shown to play an important role when used in conjunction with Mw to better characterise the tsunami and shaking potential of an earthquake (Newman and Okal, 1998; Di Giacomo et al., 2010).”

Line 24: “… the low frequency end…”, in practice we often measure Mw from the low frequency end of spectra, but really it represents the static (f = 0 Hz) component. Maybe just add that?
We substituted ‘characterized’ with ‘extrapolated’.

L26: Correct to “… fraction of the total energy being radiated…”; “energy” is currently missing.
Thanks, we added ‘energy’.

L 29: “parameter”; singular, not plural.
Thanks, corrected.

L 36: I’m not familiar with the methodologies to compute Me in detail, so I was a bit surprised to read that you compute Me from P waves, in opposition to S waves, which carry most of the energy. I guess it’s related to the SNR. For the more unaware readers, maybe add a brief explanation on why you compute Me from P waves?
Papers illustrating methodologies to compute Me using teleseimic recordings go back to the 1980s (E.g., papers by Boatwrigth&Choy) and we feel that we do not need to repeat all the background but focus on our Me catalogue. However, we added the following after “vertical-component P-waveforms” at line 37 in the preprint: “(teleseismic P-waves are commonly used to compute Me for global earthquakes as their energy loss during propagation can be more reliably modeled compared to S-waves)”

L 44: Distance range: 20˚ to 98˚? I guess 98˚ is related to the P-wave shadow zone. Why disregard near source recordings? Add a brief explanation, again for the sake of the more unaware readers.
Similarly to the previous point we did not want to repeat the reasons for our setup because it largely follows what it is well established in the literature. However, we added the following after “98◦ ” at line 44 in the preprint: “(standard teleseismic range

usually starts at 30◦, but we use 20 to allow closer stations to be used for rapid response purposes. The shortest distances, however, are difficult to include for global earthquakes as regional effects are not be well accounted for with a global 1-D model)"

L 47: Suggestion: add "each" before "single station"
Added

Eq 1: Please double check this equation – is it dimensionally correct? Maybe I'm missing something…
We have double checked the equation, and also compared with other papers, such as Vassiliou and Kanamori, 1982, The energy release in earthquakes, BSSA. The equation appears to be correct.

L 53: Clarify what is "a wide range of plausible focal mechanisms"
We replaced ", which are computed across a wide range of plausible focal mechanism solutions and the median value is extracted" with "computed from multiple combinations of focal mechanisms, varying strike, dip and rake over regular grid (Di Giacomo et al. 2008)."

L 55: How much is "just before" the P wave arrival?
The configuration file of Me-compute allows for setting the number of seconds by which the starting time of the extracted window is shifted with respect to the theoretical P-wave arrival time. For our application, we used a 10-second shift [information added to the manuscript].

L 57: Correct to "a single event-station pair"
Done

L 66: Can you provide a rationale for starting in 2011?
The reason for this is that the Mw Geofon catalog starts from 2011.

L 86: Wouldn't you want to take into account static station corrections, once you've analyzed a large enough dataset? It seems like that would provide more robust Me estimates. It's a common correction when computing ML.
We acknowledge that station corrections can reduce the variance of magnitude computation. Therefore, we have provided station-specific residuals that can be used as station adjustments for future computations. The rationale behind the catalog compilation was to provide users with station magnitude values and all necessary information for computing and refining the event magnitude assessment.

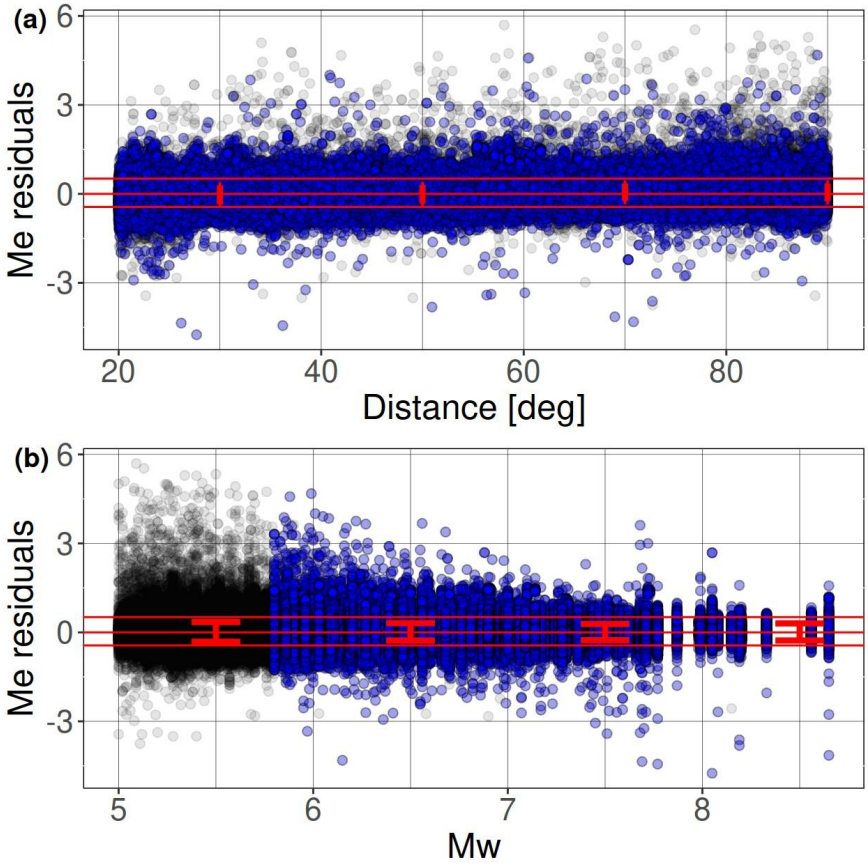
L 87: 246 networks: Do you have a smart way to cite the DOIs of all those networks in your work??
The citations and DOIs provided as Supplement as written in the acknowledgments (https://zenodo.org/records/10200493) has been created writing a simple bash script running the IRIS service for citation (https://www.fdsn.org/networks/citation/). This information is provided in the supplement.

L 90: It was surprising to me that you find entire networks outside the 5-95 percentiles. Wouldn't it be enough to exclude stations outside the 5-95 percentiles? It's not very clear to me why you need to exclude entire networks.
The decision to remove certain networks was based on the fact that most of their stations were providing outlier values, likely due to incorrect or misused information in their station inventory files (e.g. units of generation constant).

L 96: "Anomaly score": maybe give a brief explanation on the grounds on which this method flags anomalous amplitudes?
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.

Datasets D0, D1, …: it's not clear to me if you apply cumulatively or independently the quality criteria D1 -> D3. Please clarify.
The quality checks and selections indicated in Table 1 are applied sequentially in the order indicated [information added in the heading of Table 1, last column].

Figure 3: Can you overlay the dataset D6, in front of the black dots and behind the lines? In case the figure doesn't become illegible, it would be nice to see how much we lose from D3 to D6.
In terms of residual distribution with respect to Mw, only points above 5.8 were selected for D6, as shown in Table 1 (see the numbers). As for the distribution with respect to distance, it is difficult to distinguish due to the large number of overlapping points, with D3 having over 1 million points and D6 having about 750,000. It is worth noting that some of the large residuals are from Mw<6, as seen in the distribution with respect to magnitude, and therefore these values are not carried over to D6. We changed Figure 3.



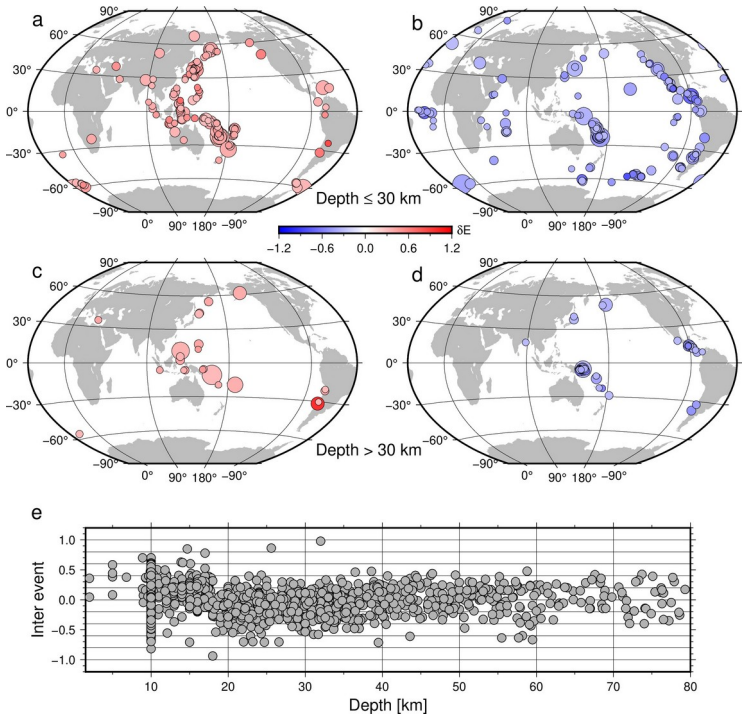
L 127: Change to “with the intra-event, equal to…”
Agreed

L 127: phi = sqrt(…) , I believe a square (^2) is missing in the last parameter, phi_S.
Thanks, corrected.

L 137: East African Rift: active, but cratonic…
Thanks

L 149: ”Similar to our approach” instead of “Like us”
Agreed

Fig 6e: Do you really need to use a log x scale? A big part of the plot is empty…
Done



L 166: “on the analysis”
Thanks

L 174: I believe it should be BB instead of HF, right before “, i.e.,”
Thanks

Figure 8: Very nice! It shows a lot of inter-station variability…
Thanks

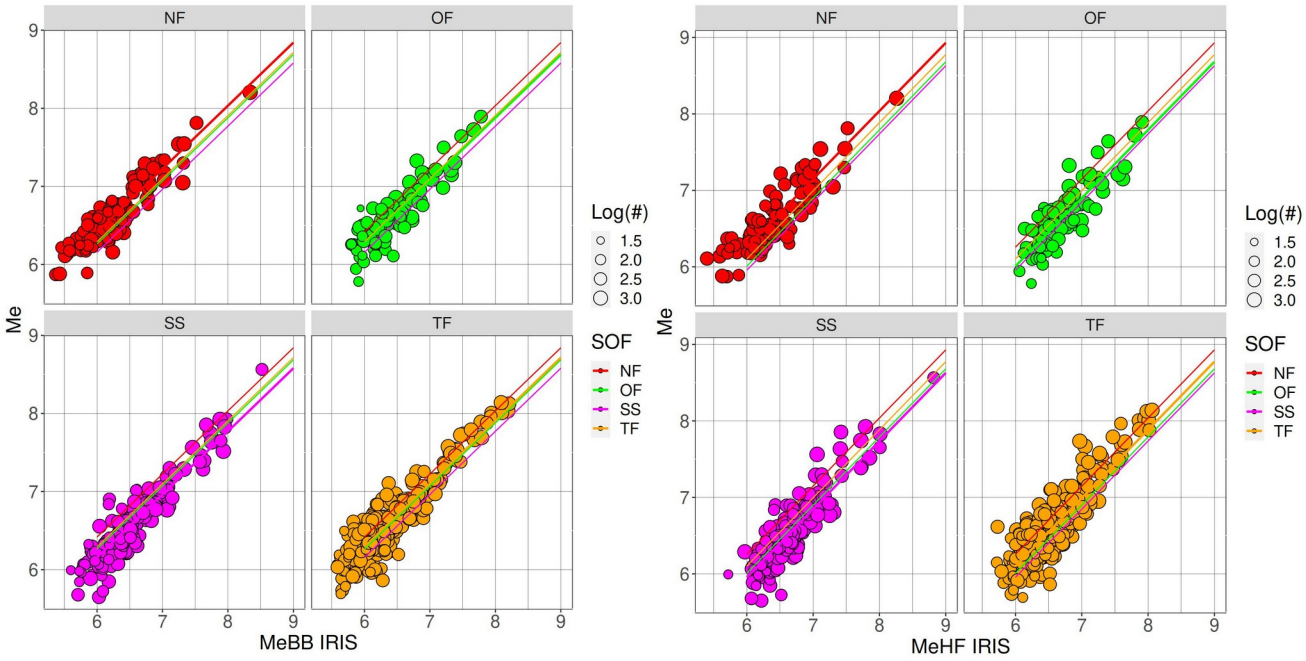
L 180: Is it really 50 deg? All other styles of faulting have 60 deg.
Yes, see Frohlich & Apperson (1992)

L 180: Just write out that OF means “other faulting styles”
Done

L 196: “, where different faulting…”
Done

L 197: It was a surprise to read here that the method does not consider radiation pattern. When in line 52 you write that G(f) is computed for a range of plausible focal mechanisms, I though you took into account the focal mechanism, therefore the radiation pattern. Please clarify in the text. It seems like you should take into account the focal mechanism/radiation pattern, to get better Me estimates…
That’s why we use median values from the Green’s functions out of several computations from different focal mechanisms. The procedure is designed to be used without the knowledge of the focal mechanism, as, for example, already done by Newman&Okal, JGR 1998.

Fig 11: Maybe easier to read in a 2 rows x 4 columns plot? Top-BB, Bottom-HF.
We prefer to keep the grouping per magnitude type but we arranged the panels horizontally.



Finally, it would be really nice to have a “Conclusions” section, where you summarize the main take-away lessons from your new Me catalog. Why is this catalog useful? What new things are we learning from it?
We added a section ‘Conclusive remarks’ where we stated:
We computed the energy magnitude Me for 6349 events in the moment magnitude catalog disseminated by Geofon. When combined with Mw, Me allows for a better characterization of the tsunami and shaking potential of an earthquake. The procedure used to compile the data set, which includes 1031396 Me values for each recording station, is described in detail. Residuals are evaluated using a mixed-effects regression, which partitions the overall residuals into event-specific and station-specific contributions. These random effects are included in the distributed catalog, enabling the computation of Me for future events using inter-station residuals as station corrections to reduce the uncertainty on Me. They also enable the assessment of energy magnitude adjustments for specific regions or faulting mechanisms by using inter-event residuals, and locating propagation anomalies with respect to the global model used to compute Green's functions using the left-over residuals. The methodology employed for computing Me (Di Giacomo et al, 2008) is suitable for the rapid assessment of Me (Di Giacomo et al, 2010). Therefore, it has been implemented as a module for SeiscomP, allowing for the automatic computation of Me in real-time and keeping the Me catalog up-to-date.

Great work! Thank you for this contribution.
Thank you and thanks for your comments.