Review Reports

Reviewer Comments (Reviewer A)

Dear authors,

This is a fine submission and I have recommended minor/moderate revision. In this document I have included all my comments and requests for clarifications. I hope you find them useful, and that all hyperlinks work.

Domenico Di Giacomo

Comment 1. Since the title, the authors put lots of emphasis on the use of homogenous and similar lingo (e.g., "**Homogenization** of instrumental earthquake catalogs", "**homogenized** earthquake catalog"). Whilst we all have used this terminology in studies regarding magnitude conversions, I believe that the community should abandon the use of "homogenization" when talking of magnitude conversion. I recently advocated to change our lingo in this sense during the ECEES conference in Bucharest, September 2022. The reason is summarized in the following figures, taken from slides presented in Bucharest.

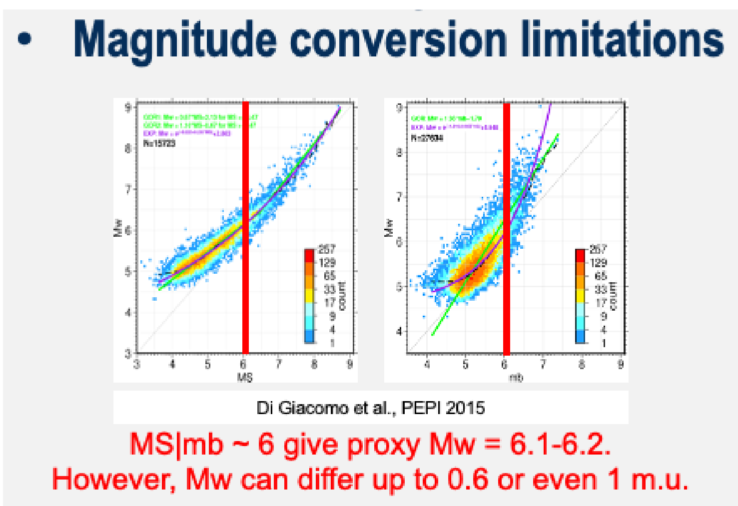

Figure C1.

Figure C1 reports the figure published by Di Giacomo et al. (2015) with the addition of the solid tick vertical red line just to point out the fact that when you use either MS or mb as basis to get a proxy Mw, the latter will always be 6.1-6.2 starting from MS|mb ~ 6 (in that given model, but the same can be said for any other model). However, in fact, Mw can differ significantly from what that (or any other) model would give. Below I provide an example to reinforce this argument. Figure C2 compares the magnitude information for two great Kuril Islands earthquakes as extracted from the ISC Bulletin. These interesting pair of earthquakes (occurred close in space and time) has been the subject of several studies, and the work by Ammon, Kanamori and Lay (Nature 2008) gives us the source spectra for both earthquakes. We can use these source spectra to make sense of the significant difference in mb, MS (energy magnitude ME as well) for these two earthquakes. Indeed, the thrust event (Mw 8.3-8.4) has systematically smaller mb, MS and ME than the normal fault event despite a smaller Mw of 8.1 for the normal fault event.

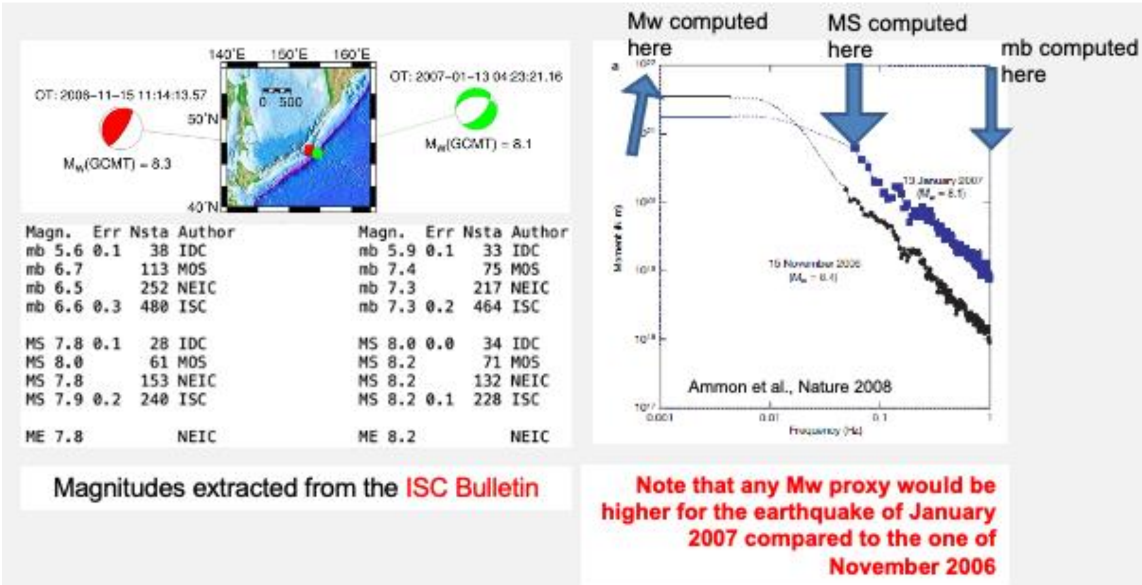

Figure C2

This is because different magnitudes mean different things as they are related to different parts of the source spectrum. Any conversion starting from any magnitude type would provide a higher Mw for the normal fault event, but authoritative direct Mw solutions tell the opposite.

Hence, I believe that the use of “homogenization” is generally misleading. From various English dictionary definitions, in our context we use “homogenize” with the meaning “to make homogeneous; to make uniform or similar, as in composition or function”. When talking of magnitude conversion, I suggest to replace (throughout the text) “homogenize” with “standardize”, with the latter meaning (again from various English dictionaries) “to choose or establish a standard for”. To me it seems more appropriate because, in fact, what we do is to establish a standard (Mw) and try to convert other magnitudes to that standard.

In addition, I have to point out that the catalog presented here is focused on magnitude conversion and locations are not reassessed. As such, locations in the catalog are the result of multiple location techniques, and this makes the case for “**homogenization** of earthquake catalogs” somewhat weaker. I suggest to add magnitude in the title, e.g.: “Standardizing magnitudes in instrumental earthquake catalogs...”.

Comment 2. Both in the abstract and on the first line of Section 6 the authors state that the catalog here presented is “more than 5500 events”. The supplementary file “DSTFZ_v8d_SI.csv” lists 25000 events. I realize that the events with magnitude field Mw >= 3 are 5637. It would therefore be desirable to a have a README file accompanying the csv file to explain its content (why 25000 events when the reader expects ~5,500, the meaning of the acronyms inside the file, e.g., what ISC-RAW means, etc.). I guess the csv file will go on Zenodo with its own DOI, if not I suggest to register the dataset with the **ISC dataset repository**. I did not check but it would be good to know if the statistics shown in Tables S1 and S2 refer to the whole 25,000 events or only to the~5,500 events with Mw >= 3. Nevertheless, I tried to extract from the csv file the events missing in the ISC Bulletin (it is written to be 8% of the catalogue, I found 561 entries with ISC not mentioned as source in any field) and dumped the events in the study area that occur within 15 seconds of the origin time of those entries. I did not look too much in detail but I provide the file for you to peruse and seek to verify why some events in the ISC Bulletin were excluded or not used in your general listing.

Comment 3. The authors have a deep knowledge and understanding of the literature, including the work by Di Giacomo et al. (PEPI 2015). In there, however, the original dataset to obtain conversion relationships to Mw have been done by splitting it in training and validating datasets. It would be recommendable that the authors do a similar split, at least for the datasets that have enough points over a satisfactory magnitude range, to show the performance of the derived models on the validating dataset (maybe adding the results in the supplement).

Comment 4. In the text the authors refer several times to the "ISC bulletin". As it is the name of a data product, it would be appreciated if referred to as "ISC Bulletin" consistently throughout the text.

Other points/comments/suggestions:

- Page 3, Section 3, lines 2-3: unless you want to refer to the ISC data collection status back in 2018, please replace "about 130 networks worldwide (ISC, 2018)" with "about 150 networks worldwide (ISC, 2022)";
- Page 3, Section 3, after the description of the content of the ISC-GEM Catalogue (from page 3, line 8 to page 4, line 5): I think it would be appropriate to add that in later versions (V10 recently released) the ISC-GEM Catalogue has dropped the global cut-off magnitude to 5.5 pre-1964 and to 5.0 from 1976;
- Table 1: the description of the sources is all fine, but in reality the parametric data stored in IRIS comes from other parties (ISC, NEIC, GCMT etc.) and it is there to support IRIS dataset and services (i.e., waveforms). As such, it should not be used for earthquake catalog research. I suggest to delete any mention of it in the text;
- some thoughts on the paragraph soon after Table 1 (end of page 4 and beginning of page 5): there are two main situations for the ISC not reporting a local event: a) before 1999 small events from many contributors were not included in the final ISC Bulletin; b) after

1999 if a reported event, no matter how small it is, has been removed from the ISC Bulletin, there is a reason for that. This means that, at least from 1999, say if a M3.5 is missing from the ISC Bulletin is not necessarily a bad thing from our side, and reintegrating an event from the local report may not be the best way to proceed. In such cases I would suggest the authors to verify the existence of associated station data with the local reports, and if yes, check if the data fits in the event. This is just to say that no catalogue is immune to errors or spurious entries (e.g., for several decades the main global catalogs listed the mythical M8.2 Peru earthquake of 12 December 1908, see Di Giacomo and Dewey, 2020), but adding events without validating them can lead to spurious entries (see also Di Giacomo and Storchak, Compt. Rendus Geosci. 2023). Trickier is the case where ISC lists the solution of a local agency but that solution is missing from the local report. There are multiple situations for that to happen, but again, in such cases the best we can do is to verify the event with stations data, if any. All of this is just to say that yes I understand your reasoning but care should be used when integrating events from multiple sources and not using station data to validate them. If the authors do that, it should be mentioned. Perhaps, a short synthesis of this argument should be added in this paragraph. On a smaller note, please replace "detected events" with "reported events", as the ISC does not do detection of events.

- Page, 5, at the first mention $m_{B:}$ please refer to Gutenberg (1945a,b);
- Page, 5, at the first mention $M_{S:}$ please refer to Gutenberg (1945c);
- page 5, line 9 of paragraph #3: replace "M_S is not appropriate" with "M_S strictly measured around 20 s is not appropriate". You may add that broadband M_S suffers less of saturation and deviates from Mw only for tsunami earthquakes (Kanamori, 1972) and ~M9 earthquakes (Di Giacomo and Storchak, 2022);
- page 7, first mention of Figure S3: a bit odd that in the text this is the first reference to a figure in the supplement and it is not Figure S1; please check this is not the case for all other figures;
- page 7, end of section 4.2.2: the statement "with the exception of MED-RCMT (Pondrelli et al. 2011), which showed significantly larger scatter when compared to the other moment magnitudes" comes somewhat as a surprise. It would be desirable to see this large scatter (does it affect M<5 only?), maybe in a supplementary figure, since most readers would consider the MED_RCMT as the reference Mw source for this area after GCMT (literature has plenty of such examples). Hence the question hangs why the authors decided to exclude the most long-term and well-established direct source of Mw for the EuroMed area. On a similar note, there are a handful of ZUR_RMT solutions the authors could use, any reason to exclude them?
- page 13, second line after beginning of Section 5: the use of "duplicates" here is misleading, as a duplicate event is such when is part of the same dateset more than once (in reality there is one physical event but two or more entries for it in a catalogue). When merging multiple catalogues, you are facing a grouping problem, hence I suggest to replace "I.e duplicates" with "matching events" or "events to be grouped" or any other way the authors may prefer. Update accordingly the use of "duplicates" in the remaining of the text;
- General comment on Section 5. I understand the author's strategy to group events from different inputs, and theirs is the best I have read from papers dealing with this subject. However, the shortcomings that I already mentioned remain, namely that earthquake parameters alone may not be enough to properly group events. I think this aspect should be mentioned here;
- page 13, section 5.1, line 8: replace "he/she" with "one";

- page 14, section 5.1, line 4: it would be better to add “when one relies on earthquake solutions alone (i.e., without stations data)” to “impossible to distinguish a fore/aftershock from a duplicate”. The ISC analysts are indeed in a difficult and time-consuming situation when there are close-by events (in space and/or time), and the only way to address such cases (difficult but not impossible) is to make sense of the associated data. I believe this is important to mention because the wording used in the text sounds too general but it is actually worse if one relies on earthquake parameters only;
- end of page 14, on the direct Mw hierarchy: you should use GCMT, ISC-GEM if Mw from bibliography, MED_RCMT (unless evidence is shown why not, see earlier comment), NEIC, EMSC, GII, EMEC, EMME. Not sure if here ISC means from ISC Bulletin or Mw from ISC as an author (ISC does not compute Mw, except from PPSM solutions for recent earthquakes). You should really discard IRIS from this list. It may help looking at the EuroMed area figures in Di Giacomo et al. (2021);
- at the end of page 17 it seems that the catalogue has event “most likely related to potash mines…”. Does that mean that induced or anthropogenic events may be included in the catalogue? A clarification to this regard would be desirable for any catalogue user.

Reviewer Comments (Reviewer B)

For author and editor

The manuscript “Homogenizing instrumental earthquake catalogs - a case study around the Dead Sea Transform Fault Zone” presents a framework to merge multiple instrumental earthquake catalogs and homogenise their magnitudes for seismic hazard applications. The authors apply the framework to the Dead Sea Transform Fault Zone (DSTFZ) and provide the derived catalog as a supplement to the paper. The motivation for the work is good since earthquake catalogs are of utmost importance for the study of seismicity, in particular seismic hazard assessments.

There are a number of ways in which the authors could streamline and improve the presentation of the paper, as outlined further below. Also, the authors could be more explicit about what sets their work apart from previously published methods and tools (e.g. Weatherill et al., 2016).

There is also some detail missing about the data selection, or maybe I have missed it. Are the earthquakes in the bulletins that are merged and for which the magnitudes are homogenized in the region shown in Figure 1 and correspond to the numbers in Table 1? Or do they cover a wider area?

A major component of the work is deriving regression relations for magnitudes. As introduced by Castellaro and Bormann (2007) for linear regression and widely accepted in the literature, the authors use particular orthogonal regression (POR) but extend from linear to other functional forms, in particular exponential and power law. POR requires two key assumptions to be met. First, orthogonal regression assumes that there is only measurement error between the variables. However, since different magnitude types, in particular moment magnitude and magnitudes determined from amplitude readings such as local magnitude, body-wave and surface wave magnitudes, capture different physical aspects of the earthquake, there is also a so-called equation error present (Carroll and Ruppert, 1996). Second, the assumption of POR is that the ratio of measurement errors is between 0.7 and 1.8 (Castellaro and Bormann, 2007). The authors do not provide details of the measurement errors in this manuscript although at least the ISC catalogue includes uncertainty estimates for different magnitude types.

The authors derive a magnitude fitting procedure that includes grouping magnitudes of similar type, deriving conversion equations for different magnitude types to Mw, selecting the best functional form and the applicable range. The results are presented in a number of figures and a table with the ten most common magnitude types.

The section on magnitude homogenization is followed by the section on merging multiple catalogs, which seems the wrong order when deriving regressions for earthquake magnitudes from different catalogs. The method to merge the catalogs seems logical and sufficiently described. However, it is again not so clear what is new compared to previous work. Results are again presented by various figures and a table.

The penultimate section is the presentation of the new catalog. Results are shown by two figures, a density of earthquake magnitudes and a map with earthquakes of Mw>=3.0. It is not clear for which time period, if at all, completeness was achieved within the target magnitude range.

The conclusions rephrase the problem that was addressed and summarise the work undertaken and its results.

Further comments and questions:

Abstract: Please define upfront that “homogenizing” involves merging different catalogs and that you aim for Mw>=3.0 equivalent magnitude estimate for seismic hazard analysis. Also please provide more details on the form of the regression, i.e. linear, exponential and power law.

Section 1 Introduction: For ease of reading and clarity, please consider stating your aim upfront and then providing details on the importance of the work.

Section 2 Seismotectonic session: Figure 1 covers a very long period of time; please consider also providing a stem plot of magnitudes versus time to give the reader, who is not familiar with the area, a temporal understanding of the earthquake occurrence.

Section 3 Input datasets:

Provide details on the catalogue sources in the same order as they are listed in Table 1 - that will hugely help the reader to take in the information.

Make sure all acronyms like EMC, CSEM, EMEC, CENEC etc are explained at first mention. Please consider adding a glossary of terms since the paper features many acronyms.

Table 1: Please add the full name for each acronym

Also please clarify what area in each catalog was searched as addressed above.

Section 4 Magnitude homogenisation

How come that this section precedes the section on Merging multiple catalogs? How can you combine the analyse the magnitudes if you have not gathered them from various sources before? Or are you initially using only magnitudes that are reported in the ISC catalogue?

Section 4.1 I enjoyed the section on magnitude scales and their limitations; however, it is not new material and could be shortened.

Section 4.2.1 Carroll and Ruppert, 1996, is not cited correctly. Their main point is “that orthogonal regression is often misused in error-in variables linear regression because of a failure to account for equation errors". Equation error is inherent scatter in the data, other than uncertainty in the measurements. Equation error is present for different magnitude types, e.g. moment magnitude and local magnitude since they capture different physical aspects of the earthquake. Carroll and Ruppert recommend using the method of moments to derive the regression.

The authors cite Gasperini et al. (2015) "it is widely accepted that the error in Mw is generally smaller than those of traditional magnitudes". I am not so sure what the basis for the statement is. Since Mw is calculated from the simultaneous inversion of many station recordings, it is not possible to estimate uncertainty by variation in measurements between stations. For New Zealand, the estimate of uncertainty of regional moment magnitude is around 0.2 (Section 3.8 in Christophersen et al., 2022) and 0.018 for ML in the sample set to derive regressions of Mw versus ML (Rhoades and Christophersen, 2017). It would be good to include the uncertainties that are available in the paper (for the ISC data as stated in Section 4.2.6).

Section 5.2 What was your rationale for your preferred hierarchy of Mw?

Figure 6 was a bit excited and seems to have jumped into the manuscript three times. Figure 7 might be easier to interpret if you selected percentiles similar to Figure 8. Maybe combine them by a vertical line to help the eye.

Figure 8: The symbols and colour for the axis on the right do not stand out enough – maybe select an open larger circle instead of a little start? Also, it might help to connect the symbols for each station with a vertical line.

Section 6: Catalog overview

Have you investigated the magnitude of completeness? Is your target of Mw>=3.0 achieved at completeness level, at least for the later years?

Data availability section

Why not make the missing conversion relations available in the supplement?

I have tried all the dataset links; https://earthquake.co.il/en/earthquake/searchEQS.php does not work for me. It would be helpful to include the search criteria for each catalog and when the dataset was last accessed.

Additional question:

Have you thought about making the code for your procedures available?

Thank you for the opportunity to review this work. My recommendation: Revision required.

Kind regards,

Annemarie Christophersen, GNS Science, New Zealand

References:

Carroll, R.J., and Ruppert, D. (1996). The Use and Misuse of Orthogonal Regression in Linear Errors-in-Variables Models. *The American Statistician* 50(1), 1-6. doi: 10.1080/00031305.1996.10473533.

Castellaro, S., and Bormann, P. (2007). Performance of Different Regression Procedures on the Magnitude Conversion Problem. *Bulletin of the Seismological Society of America* 97(4), 1167.

Christophersen, A., Bourguignon, S., Rhoades, D., TI, A., Salichon, J., Ristau, J., et al. (2022). "Consistent magnitudes over time for the revision of the New Zealand National Seismic Hazard Model", in: *GNS Science report 2021/56.* (Lower Hutt, New Zealand: GNS Science).

Gasperini, P., Lolli, B., and Castellaro, S. (2015). Comparative Analysis of Regression Methods Used for Seismic Magnitude Conversions. *Bulletin of the Seismological Society of America* 105(3), 1787.

Rhoades, D.A., and Christophersen, A. (2017). Magnitude Conversion of Earthquake Rate Forecasts. *Bulletin of the Seismological Society of America* 107(6), 3037-3043. doi: 10.1785/0120170225.

Weatherill, G.A., Pagani, M., and Garcia, J. (2016). Exploring earthquake databases for the creation of magnitude-homogeneous catalogues: tools for application on a regional and global scale. *Geophysical Journal International* 206(3), 1652-1676. doi: 10.1093/gji/ggw232.

Response to Reviewers

Reviewer A (Domenico Di Giacomo):

We thank Dr. Di Giacomo for reviewing our study. We address his comments one-by-one below (in purple).

Comment 1. Since the title, the authors put lots of emphasis on the use of homogenous and similar lingo (e.g., "**Homogenization** of instrumental earthquake catalogs", "**homogenized** earthquake catalog"). Whilst we all have used this terminology in studies regarding magnitude conversions, I believe that the community should abandon the use of "homogenization" when talking of magnitude conversion. I recently advocated to change our lingo in this sense during the ECEES conference in Bucharest, September 2022. The reason is summarized in the following figures, taken from slides presented in Bucharest.

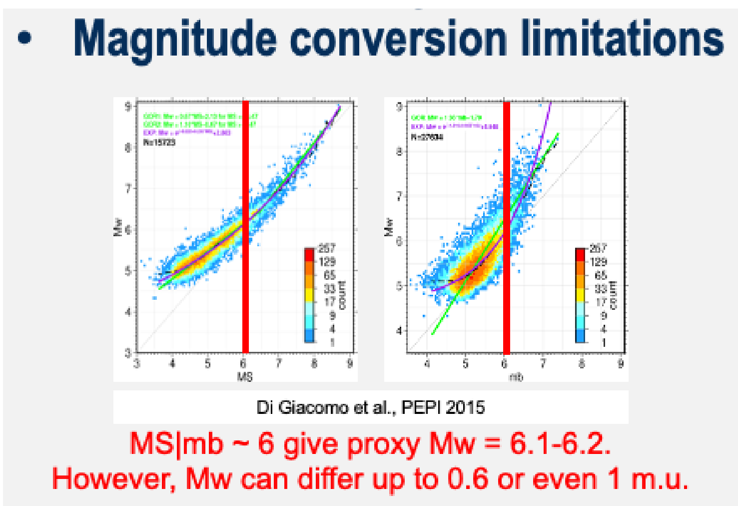

Figure C1.

Figure C1 reports the figure published by Di Giacomo et al. (2015) with the addition of the solid tick vertical red line just to point out the fact that when you use either MS or mb as basis to get a proxy Mw, the latter will always be 6.1-6.2 starting from MS|mb ~ 6 (in that given model, but the same can be said for any other model). However, in fact, Mw can differ significantly from what that (or any other) model would give. Below I provide an example to reinforce this argument. Figure C2 compares the magnitude information for two great Kuril Islands earthquakes as extracted from the ISC Bulletin. These interesting pair of earthquakes (occurred close in space and time) has been the subject of several studies, and the work by Ammon, Kanamori and Lay (Nature 2008) gives us the source spectra for both earthquakes. We can use these source spectra to make sense of the significant difference in mb, MS (energy magnitude ME as well) for these two earthquakes. Indeed, the thrust event (Mw 8.3-8.4) has systematically smaller mb, MS and ME than the normal fault event despite a smaller Mw of 8.1 for the normal fault event.

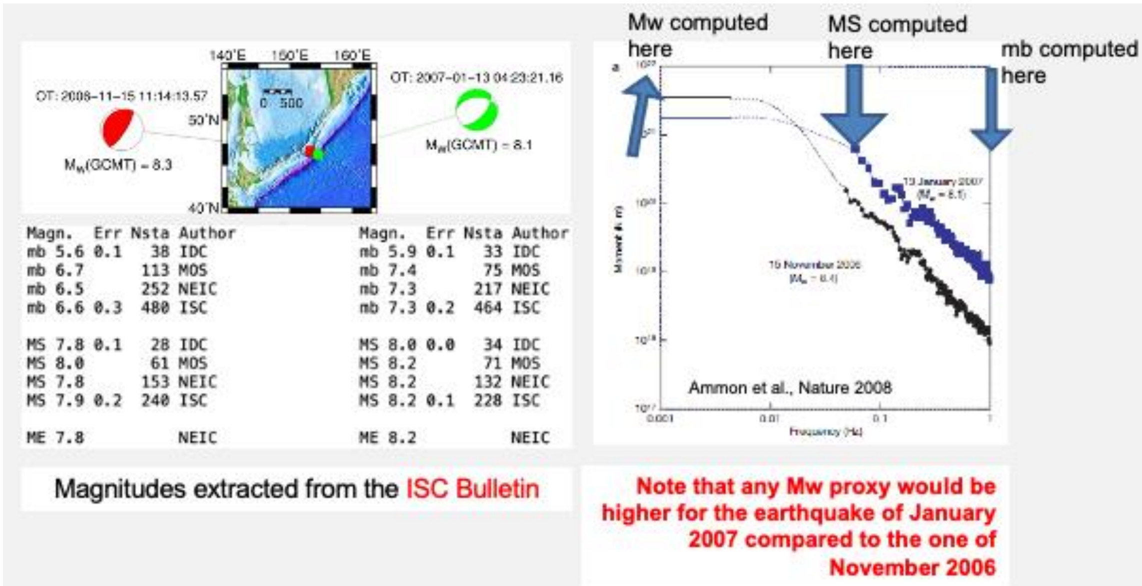

Figure C2

This is because different magnitudes mean different things as they are related to different parts of the source spectrum. Any conversion starting from any magnitude type would provide a higher Mw for the normal fault event, but authoritative direct Mw solutions tell the opposite.

Hence, I believe that the use of “homogenization” is generally misleading. From various English dictionary definitions, in our context we use “homogenize” with the meaning “to make homogeneous; to make uniform or similar, as in composition or function”. When talking of magnitude conversion, I suggest to replace (throughout the text) “homogenize” with “standardize”, with the latter meaning (again from various English dictionaries) “to choose or establish a standard for”. To me it seems more appropriate because, in fact, what we do is to establish a standard (Mw) and try to convert other magnitudes to that standard.

In addition, I have to point out that the catalog presented here is focused on magnitude conversion and locations are not reassessed. As such, locations in the catalog are the result of multiple location techniques, and this makes the case for “**homogenization** of earthquake catalogs” somewhat weaker. I suggest to add magnitude in the title, e.g.: “Standardizing magnitudes in instrumental earthquake catalogs...”.

We conceptually understand the reviewer’s perspective; that is why we use the term “proxy Mw” in the text. Furthermore, both in the *abstract* and in the *introduction*, we clearly define how we interpret the word “homogenization”: “*common events within the bulletins have to be identified and assigned with the most suitable origin-time and location solution, while all the events have to be harmonized into a single magnitude scale” & “a unified instrumental earthquake catalog containing only unique events with standardized parametric information“.* Crucially, we feel uncomfortable being the first ones to adopt this new terminology. The term “catalog homogenization” has been well established by now and anyone who tries to do what we are presenting will search these words on *google scholar*. We would like that person to find our paper and use our methods. Finally, we did more than just standardize the magnitudes; we also removed duplicated entries after merging all the sources. We wanted to use a single word that captures both procedures.

Comment 2. Both in the abstract and on the first line of Section 6 the authors state that the catalog here presented is “more than 5500 events”. The supplementary file “DSTFZ_v8d_SI.csv” lists 25000 events. I realize that the events with magnitude field Mw >= 3 are 5637. It would therefore be desirable to a have a README file accompanying the csv file to explain its content (why 25000 events when the reader expects ~5,500, the meaning of the acronyms inside the file, e.g., what ISC-RAW means, etc.).

There was indeed some lack of clarity. We modified the first sentence in section 6 to clarify that the attached catalog contains 25000 evets, but only around 5500 of them are above Mw 3.

I guess the csv file will go on Zenodo with its own DOI, if not I suggest to register the dataset with the ISC dataset repository.

The journal will handle the electronic supplement. We are keen on adding the catalog to the *ISC dataset repository*, regardless.

I did not check but it would be good to know if the statistics shown in Tables S1 and S2 refer to the whole 25,000 events or only to the~5,500 events with Mw >= 3.

Table S1 refers to all events above M 2.

Table S2 refers only to events above Mw 3, as stated in its caption.

Nevertheless, I tried to extract from the csv file the events missing in the ISC Bulletin (it is written to be 8% of the catalogue, I found 561 entries with ISC not mentioned as source in any field) and dumped the events in the study area that occur within 15 seconds of the origin time of those entries. I did not look too much in detail but I provide the file for you to peruse and seek to verify why some events in the ISC Bulletin were excluded or not used in your general listing.

Unfortunately, we did not receive a csv file within the journal’s email. The statistics were performed only for Mw >= 3 events, while the attached catalog contains more events. We also used data last retrieved in Feb 2018. Furthermore, in our hierarchy ISC-GEM was above ISC, so not having ISC as its source does not always mean it was not available within ISC. It is reasonable to assume that some events from GII were not listed within ISC, either because GII did not provide them to ISC or because ISC filtered them out as low-quality.

Comment 3. The authors have a deep knowledge and understanding of the literature, including the work by Di Giacomo et al. (PEPI 2015). In there, however, the original dataset to obtain conversion relationships to Mw have been done by splitting it in training and validating datasets. It would be recommendable that the authors do a similar split, at least for the datasets that have enough points over a satisfactory magnitude range, to show the performance of the derived models on the validating dataset (maybe adding the results in the supplement).

We thank the reviewer for his kind words. We do agree that showcasing the performance out of sample is useful, and it is almost never done in the literature. Even if we were to do that for academic purposes, we would still eventually provide the public with the regressions derived from all the data, as this would be the best available fit for future use. Furthermore, in most cases, we do not have rich enough datasets to do reliable testing. In any case, our scripts were not designed with that in mind and it would be quite tricky to modify them now; they are already quite complex to begin with, because they have to use $RMSOE_{adj}$ to optimize for magnitude range, functional form, etc.

Comment 4. In the text the authors refer several times to the “ISC bulletin”. As it is the name of a data product, it would be appreciated if referred to as “ISC Bulletin” consistently throughout the text.

We have now replaced “ISC” with “ISC bulletin”, whenever necessary.

Other points/comments/suggestions:

- Page 3, Section 3, lines 2-3: unless you want to refer to the ISC data collection status back in 2018, please replace “about 130 networks worldwide (ISC, 2018)” with “about 150 networks worldwide (ISC, 2022)”;

Yes, we are referring to the 2018 status. Last data-update was Feb 2018 (now stated in the *Data availability* section).

- Page 3, Section 3, after the description of the content of the ISC-GEM Catalogue (from page 3, line 8 to page 4, line 5): I think it would be appropriate to add that in later versions (V10 recently released) the ISC-GEM Catalogue has dropped the global cut-off magnitude to 5.5 pre-1964 and to 5.0 from 1976;

We now mention in the text that newer versions have lower thresholds.

- Table 1: the description of the sources is all fine, but in reality the parametric data stored in IRIS comes from other parties (ISC, NEIC, GCMT etc.) and it is there to support IRIS dataset and services (i.e., waveforms). As such, it should not be used for earthquake catalog research. I suggest to delete any mention of it in the text;

We agree with the reviewer that the IRIS dataset seems redundant. We mention in the text that it is just a collection of events from other input sources, without further review. We argue why we included it after Table 1. In fact, there are about 35 events in the catalog that use IRIS as a source;

some of them have magnitude above 4. Removing IRIS from the text would require us to redo the entire process related to the duplicate-search, a process that includes manual inspection of the results, in case the margins missed edge-cases. We would really like to avoid redoing all this just to refine the source of 35 events.

- some thoughts on the paragraph soon after Table 1 (end of page 4 and beginning of page 5): there are two main situations for the ISC not reporting a local event: a) before 1999 small events from many contributors were not included in the final ISC Bulletin; b) after 1999 if a reported event, no matter how small it is, has been removed from the ISC Bulletin, there is a reason for that. This means that, at least from 1999, say if a M3.5 is missing from the ISC Bulletin is not necessarily a bad thing from our side, and reintegrating an event from the local report may not be the best way to proceed. In such cases I would suggest the authors to verify the existence of associated station data with the local reports, and if yes, check if the data fits in the event. This is just to say that no catalogue is immune to errors or spurious entries (e.g., for several decades the main global catalogs listed the mythical M8.2 Peru earthquake of 12 December 1908, see Di Giacomo and Dewey, 2020), but adding events without validating them can lead to spurious entries (see also Di Giacomo and Storchak, Compt. Rendus Geosci. 2023). Trickier is the case where ISC lists the solution of a local agency but that solution is missing from the local report. There are multiple situations for that to happen, but again, in such cases the best we can do is to verify the event with stations data, if any. All of this is just to say that yes I understand your reasoning but care should be used when integrating events from multiple sources and not using station data to validate them. If the authors do that, it should be mentioned. Perhaps, a short synthesis of this argument should be added in this paragraph. On a smaller note, please replace “detected events” with “reported events”, as the ISC does not do detection of events.

We thank the reviewer for all this valuable information. We did replace the word “detected”. To be clear, we outlined these facts in order to justify using extra catalogs, apart from the ISC bulletin, given that other people have criticized us for not solely relying on the ISC bulletin. We did not imply any mishandling of the events by the ISC bulletin. We trust the reviewing processes followed by the ISC bulletin. We certainly did not look at station-data when conducting our analysis, because this would be outside the scope of our framework. We always assume that each catalog-source has removed duplicated entries and is only reporting real earthquakes. This implies that if one source is missing an event reported by another source, we will end up including this event regardless of the hierarchy of sources. We believe that such edge cases are so rare that they do not influence the final activity rates. We are not sure how exactly we should communicate our message better. We are open to specific suggestions.

- Page, 5, at the first mention $m_{B:}$ please refer to Gutenberg (1945a,b);

Reference added.

- Page, 5, at the first mention $M_{S:}$ please refer to Gutenberg (1945c);

Reference added.

- page 5, line 9 of paragraph #3: replace “M_S is not appropriate” with “M_S strictly measured around 20 s is not appropriate”. You may add that broadband M_S suffers less of saturation and deviates from Mw only for tsunami earthquakes (Kanamori, 1972) and ~M9 earthquakes (Di Giacomo and Storchak, 2022);

Reference added. We also applied the suggested phrasing.

- page 7, first mention of Figure S3: a bit odd that in the text this is the first reference to a figure in the supplement and it is not Figure S1; please check this is not the case for all other figures;

We thank the reviewer for this correction. We fixed this issue now.

- page 7, end of section 4.2.2: the statement “with the exception of MED-RCMT (Pondrelli et al. 2011), which showed significantly larger scatter when compared to the other moment magnitudes” comes somewhat as a surprise. It would be desirable to see this large scatter (does it affect M<5 only?), maybe in a supplementary figure, since most readers would consider the MED_RCMT as the reference Mw source for this area after GCMT (literature has plenty of such examples). Hence the question hangs why the authors decided to exclude the most long-term and well-established direct source of Mw for the EuroMed area. On a similar note, there are a handful of ZUR_RMT solutions the authors could use, any reason to exclude them?

We added the relevant requested plots in the supplement (Figure S3), to provide clarity. As you can see, MED_RCMT has by far the largest error in relation to the diagonal (50% larger than any other agency) and a clear trend to overestimate magnitudes below 5.

- page 13, second line after beginning of Section 5: the use of “duplicates” here is misleading, as a duplicate event is such when is part of the same dateset more than once (in reality there is one physical event but two or more entries for it in a catalogue). When merging multiple catalogues, you are facing a grouping problem, hence I suggest to replace “I.e duplicates” with “matching events” or “events to be grouped” or any other way the authors may prefer. Update accordingly the use of “duplicates” in the remaining of the text;

From our perspective, once you merge the input-catalogs together, then the unified dataset has duplicated entries for each events, i.e. duplicates. The term “duplicates” is almost always used in that way in the global literature, for example:

- https://www.nature.com/articles/s41597-019-0234-z
- https://www.mdpi.com/2076-3417/12/10/5010
- https://www.frontiersin.org/articles/10.3389/feart.2022.820277/full
- https://link.springer.com/article/10.1007/s10518-015-9833-z
- http://www.seismo.ethz.ch/static/ecos-09/ECOS-2009_Report_final_WEB.pdf [SED; ECOS]
- https://doi.org/10.1785/0220170108 [USGS]
- https://link.springer.com/article/10.1007/s10950-012-9302-y [EMEC]

Adopting a single word to describe what we mean is advantageous; people can easily search for it and find our methods. It also makes the text a lot more easy to follow, from a phrasing perspective.

- General comment on Section 5. I understand the author's strategy to group events from different inputs, and theirs is the best I have read from papers dealing with this subject. However, the shortcomings that I already mentioned remain, namely that earthquake parameters alone may not be enough to properly group events. I think this aspect should be mentioned here;

We are not sure what the implied "aforementioned shortcomings" are. We just followed the standard procedure, refining the way one defines the margins.

- page 13, section 5.1, line 8: replace "he/she" with "one";

We thank the reviewer for this suggestion, which we followed.

- page 14, section 5.1, line 4: it would be better to add "when one relies on earthquake solutions alone (i.e., without stations data)" to "impossible to distinguish a fore/aftershock from a duplicate". The ISC analysts are indeed in a difficult and time-consuming situation when there are close-by events (in space and/or time), and the only way to address such cases (difficult but not impossible) is to make sense of the associated data. I believe this is important to mention because the wording used in the text sounds too general but it is actually worse if one relies on earthquake parameters only;

  We followed this suggestion. What do you exactly mean by *stations data*? Do you mean *waveform data* or metadata from all stations?

- end of page 14, on the direct Mw hierarchy: you should use GCMT, ISC-GEM if Mw from bibliography, MED_RCMT (unless evidence is shown why not, see earlier comment), NEIC, EMSC, GII, EMEC, EMME. Not sure if here ISC means from ISC Bulletin or Mw from ISC as an author (ISC does not compute Mw, except from PPSM solutions for recent earthquakes). You should really discard IRIS from this list. It may help looking at the EuroMed area figures in Di Giacomo et al. (2021);

  On page 14 we list the hierarchy for the proxy Mw, not for the original Mw. We added the word "proxy" in the text now to be more clear. That is why ISC is there.

- at the end of page 17 it seems that the catalogue has event "most likely related to potash mines…". Does that mean that induced or anthropogenic events may be included in the catalogue? A clarification to this regard would be desirable for any catalogue user.

  The input sources list real earthquakes, without filtering for *induced* seismicity. We just inherit the attributes of our input sources. Therefore, our compilation may indeed include events triggered by human activities. We identified the only area where we have evidence that this is the case.

Reviewer B (Annemarie Christophersen):

We thank Dr. Christophersen for reviewing our study. We address her comments one-by-one below (in purple).

Recommendation: Revisions Required

The manuscript “Homogenizing instrumental earthquake catalogs - a case study around the Dead Sea Transform Fault Zone” presents a framework to merge multiple instrumental earthquake catalogs and homogenise their magnitudes for seismic hazard applications. The authors apply the framework to the Dead Sea Transform Fault Zone (DSTFZ) and provide the derived catalog as a supplement to the paper. The motivation for the work is good since earthquake catalogs are of utmost importance for the study of seismicity, in particular seismic hazard assessments.

There are a number of ways in which the authors could streamline and improve the presentation of the paper, as outlined further below. Also, the authors could be more explicit about what sets their work apart from previously published methods and tools (e.g. Weatherill et al., 2016).

Weatherill et al. (2016) provided a toolkit that performs POR to derive magnitude conversion relations. However, the user must decide, by expert judgement, the functional form, the valid magnitude range and whether to group magnitude types. We adopted a data-driven and automated approach that makes all these decisions for the user. When it comes to merging different catalogs, Weatherill et al. (2016) again ask the user to define the margins for the duplicate-search. We define them in a data-driven way, utilizing the ISC data.

There is also some detail missing about the data selection, or maybe I have missed it. Are the earthquakes in the bulletins that are merged and for which the magnitudes are homogenized in the region shown in Figure 1 and correspond to the numbers in Table 1? Or do they cover a wider area?

Correct, the region is shown in Figs 1, 10, and the boundaries are mentioned in the abstract and in L71 of the original manuscript.

A major component of the work is deriving regression relations for magnitudes. As introduced by Castellaro and Bormann (2007) for linear regression and widely accepted in the literature, the authors use particular orthogonal regression (POR) but extend from linear to other functional forms, in particular exponential and power law. POR requires two key assumptions to be met. First, orthogonal regression assumes that there is only measurement error between the variables. However, since different magnitude types, in particular moment magnitude and magnitudes determined from amplitude readings such as local magnitude, body-wave and surface wave magnitudes, capture different physical aspects of the earthquake, there is also a so-called equation error present (Carroll and Ruppert, 1996). Second, the assumption of POR is that the ratio of measurement errors is between 0.7 and 1.8 (Castellaro and Bormann, 2007). The authors do not provide details of the measurement errors in this manuscript although at least the ISC catalogue includes uncertainty estimates for different magnitude types.

Indeed, the ISC provides some magnitude uncertainty estimates, for a subset of the earthquakes, depending on the availability of the raw data. We would not be able to test every magnitude pair we analyzed though; the data is not rich enough. The references we list that promote the use of POR, have tested the aforementioned assumptions for different magnitude scales and have concluded that in most cases POR is the better choice. We just followed the current state-of-the-art.

The authors derive a magnitude fitting procedure that includes grouping magnitudes of similar type, deriving conversion equations for different magnitude types to Mw, selecting the best functional form and the applicable range. The results are presented in a number of figures and a table with the ten most common magnitude types.

The section on magnitude homogenization is followed by the section on merging multiple catalogs, which seems the wrong order when deriving regressions for earthquake magnitudes from different catalogs. The method to merge the catalogs seems logical and sufficiently described. However, it is again not so clear what is new compared to previous work. Results are again presented by various figures and a table.

We agree with the reviewer that others might choose the opposite order. We justified in detail why we ordered the chapters (and our workflow) in the way we did, in section 5.1:

The order in which the compiler will do the merging and the magnitude homogenization is not always fixed. If the compiler chooses to discard the magnitude as a criterion for the identification of duplicates and has great confidence in the selected margins for origin time and location, then he/she could first merge all catalogs into one, storing all magnitude solutions for each unique event, and then perform the magnitude homogenization. This way, the compiler ends up with more magnitude solutions per event, hence more potential magnitude pairs to derive the conversion relations. We however preferred to do the opposite; first homogenize all sources in Mw (using the magnitude solutions of ISC and GII) and then merge all datasets utilizing the event's size to constrain the duplicate finding algorithm. One reason to do that is that it is often impossible to distinguish a fore/aftershock from a duplicate, even if the margins are ideally selected. Our main concern was not to contaminate the derivation of the magnitude conversion relations with any artifacts that the merging process might introduce. For example, events of the same earthquake sequence can be misidentified as duplicates, mixing up incompatible solutions as magnitude pairs for the regression.

Our main innovation compared to previous works is in the data-driven way we defined the margins for the duplicate-search. The last paragraph before section 5.1 reads:

Most modern unified catalogs either do not clearly state their criteria for identifying the duplicate events or choose arbitrary values based more on expert opinion than data-driven analysis. The chosen margins for instrumental catalogs vary significantly, in the order of 10-120 seconds and of 30-100km (e.g. Wang et al. 2009; Faeh et al. 2011; Beauval et al. 2013; Poggi et al. 2017), while the magnitude is rarely used as a deciding factor.

The penultimate section is the presentation of the new catalog. Results are shown by two figures, a density of earthquake magnitudes and a map with earthquakes of Mw>=3.0. It is not clear for which time period, if at all, completeness was achieved within the target magnitude range.

The catalog contains all available earthquakes above the homogenized magnitude of 3. This does not mean that M3 was the target magnitude of completeness (Mc). No spatio-temporal analysis of the Mc has been conducted here; the present paper is already quite long. The analysis of Mc is part of another study that will decluster the catalog, derive the Mc in space and time, obtain the G-R parameters and eventually perform seismic hazard analysis.

The conclusions rephrase the problem that was addressed and summarise the work undertaken and its results.

Further comments and questions:

Abstract: Please define upfront that “homogenizing” involves merging different catalogs and that you aim for Mw>=3.0 equivalent magnitude estimate for seismic hazard analysis. Also please provide more details on the form of the regression, i.e. linear, exponential and power law.

We added the possible functional forms in the abstract. The target minimum magnitude of Mw 3 is mentioned in the last sentence of the abstract.

Section 1 Introduction: For ease of reading and clarity, please consider stating your aim upfront and then providing details on the importance of the work.

We are happy with the conceptual flow of information in the Introduction (from generic to specific). We state the problem we are trying to solve, our motivation, a brief literature review, our conceptual goal and our case-study.

Section 2 Seismotectonic session: Figure 1 covers a very long period of time; please consider also providing a stem plot of magnitudes versus time to give the reader, who is not familiar with the area, a temporal understanding of the earthquake occurrence.

Figure 1 includes the historical catalog and thus data from another study led by the authors. We did not want to include too many figures sourced from another publication. Furthermore, the other reviewer asked us to remove this Figure altogether.

Section 3 Input datasets:

Provide details on the catalogue sources in the same order as they are listed in Table 1 - that will hugely help the reader to take in the information.

We thank the reviewer for her suggestion. We modified Table 1 to be in line with the ordering within the text.

Make sure all acronyms like EMC, CSEM, EMEC, CENEC etc are explained at first mention. Please consider adding a glossary of terms since the paper features many acronyms.

We double-checked that all acronyms are explained at first mention. We will do it again during the proof-reading process.

In section “Data availability”, we provide a link to a glossary from ISC, which we followed faithfully. We also mention this link a couple of times in the text.

Table 1: Please add the full name for each acronym

The names of these entities are very long and will make this Table super wide. We do not wish to add them there. They are mentioned in the text just above the Table.

Also please clarify what area in each catalog was searched as addressed above.

The stated spatial boundaries apply to all the data listed in Table 1. We clarified that in the caption now.

Section 4 Magnitude homogenisation

How come that this section precedes the section on Merging multiple catalogs? How can you combine the analyse the magnitudes if you have not gathered them from various sources before? Or are you initially using only magnitudes that are reported in the ISC catalogue?

Please refer to our response to a similar comment above.

Section 4.1 I enjoyed the section on magnitude scales and their limitations; however, it is not new material and could be shortened.

Indeed section 4.1 is a literature review that contains information that at first glance might seem non-essential. We internally debated whether it should be part of an appendix. However, in the end, we decided to keep it in the main body because it is cross-referenced several times during the discussion of the results (original manuscript: L343, L395, L411). We struggle to see how it can be further shortened, but we are open to specific suggestions.

Section 4.2.1 Carroll and Ruppert, 1996, is not cited correctly. Their main point is "that orthogonal regression is often misused in error-in variables linear regression because of a failure to account for equation errors". Equation error is inherent scatter in the data, other than uncertainty in the measurements. Equation error is present for different magnitude types, e.g. moment magnitude and local magnitude since they capture different physical aspects of the earthquake. Carroll and Ruppert recommend using the method of moments to derive the regression.

Carroll & Ruppert (1996) do acknowledge the presence of measurement errors and state that OR is used to address that (their equation 1). They just highlight the presence of another error variable "q" (their equation 9). In any case, we removed the citation, to avoid confusion.

The authors cite Gasperini et al. (2015) "it is widely accepted that the error in Mw is generally smaller than those of traditional magnitudes". I am not so sure what the basis for the statement is. Since Mw is calculated from the simultaneous inversion of many station recordings, it is not possible to estimate uncertainty by variation in measurements between stations. For New Zealand, the estimate of uncertainty of regional moment magnitude is around 0.2 (Section 3.8 in Christophersen et al., 2022) and 0.018 for ML in the sample set to derive regressions of Mw versus ML (Rhoades and Christophersen, 2017). It would be good to include the uncertainties that are available in the paper (for the ISC data as stated in Section 4.2.6).

We thank the reviewer for this comment and for the useful observations from NZ that she provided. We would like to clarify that our quote from Gasperini et al. (2015) is 100% accurate (first paragraph of their paper). Certainly, the numbers might differ between networks. We do agree that the uncertainty behind Mw is around 0.2 units. Perhaps NZ has a very good ML calibration and that is why the uncertainty is that low. We are pretty confident that the networks in the Middle East (perhaps with the exception of GII) have much higher uncertainties in their ML than NZ. The first author is familiar with TexNet (state network in Texas) and the ML uncertainty there is often above 0.3 even for very low magnitudes. To conclude, we are just replicating a cited quote that we feel is directionally correct for our region of interest.

Section 5.2 What was your rationale for your preferred hierarchy of Mw?

Partly expert-opinion, partly plots of each source against all others (see added plots in Figure S3). The effect on the derived relations is minimal.

Figure 6 was a bit excited and seems to have jumped into the manuscript three times. Figure 7 might be easier to interpret if you selected percentiles similar to Figure 8. Maybe combine them by a vertical line to help the eye.

We thank the reviewer for her correction. We fixed the issue with Figure 6 now. We also modified Figure 7 following her suggestion.

Figure 8: The symbols and colour for the axis on the right do not stand out enough – maybe select an open larger circle instead of a little start? Also, it might help to connect the symbols for each station with a vertical line.

We made changes to the figure following the suggestions of the reviewer.

Section 6: Catalog overview

Have you investigated the magnitude of completeness? Is your target of Mw>=3.0 achieved at completeness level, at least for the later years?

Please see our response to a similar question above.

Data availability section

Why not make the missing conversion relations available in the supplement?

Table 2 includes enough conversion relations to deal with the vast majority of relevant events. In *Data availability* we mention that “derived conversion relations missing from Table 2 can become available by the authors upon reasonable request”.

I have tried all the dataset links; https://earthquake.co.il/en/earthquake/searchEQS.php does not work for me. It would be helpful to include the search criteria for each catalog and when the dataset was last accessed.

GII changed its weblink. We have updated our link to the new one.

All catalogs were filtered to our spatial boundaries.

Additional question:

Have you thought about making the code for your procedures available?

We are open to releasing the *matlab* code for the conversion relations, together with the paper, once it is accepted. It would be an electronic supplement.

The code for the duplicate-removal process can be found here https://github.com/klunk386/CatalogueTool-Lite/tree/master/OQCatk (link added in the paper).