# Reviewer Comments

## Reviewer 1

Adam RINGLER

Review of: When are non-double-couple components of seismic moment tensors reliable?

This paper compares differences in the non-double couple components of three different catalogs. This is an interesting problem and the authors do a good job investigating the initial pieces of this problem. While I think this will make a useful contribution in the field of seismology and more specifically moment tensor catalogs, I have a few suggestions before I recommend this for publication. I think addressing these could help make the paper more readable and have a higher impact overall.

Major:

I had a very hard time following the convention for the equations. There are a number of non-standard conventions and equations with potential typos. While some of these could be “correct” they do not enhance the readability of the paper. Perhaps the authors could pick a convention that is widely used (e.g., Dziewonki et al., 1981) and convert everything to these conventions. Or they could explain why their conventions deviate.

The words “reliable,” “repeatable,” “reproducible” etc. all have different meanings. It wasn’t clear to me that these were used consistently throughout. Please double check this.

Please put more detail into the conclusions and introduction about why this work is important and how it fits into seismology. While all three catalogs have different approaches, we would eventually hope we get the same answer from all three methods. While your work does a good job pointing out that they might not be consistent, it could have a higher impact of pointing out why this is important.

Minor:

Lines 26 through 28: Please double check that this sentence is correct. I believe the first systematic studies of MTs came from Dziewonski, Chou, and Woodhouse (1981). This would be analog data. If these were “automatic” or not is up to debate, but I think MTs were being systematically studied on analog records. Anyway, just double check that this is correct. I don’t know enough about MT history to say either way.

Equation 1: Please identify your variables. What is M_ij

Equation 2: Please identify your variables what is M_0 and M_w, please cite where this comes from. I believe it is Kanamori (1977).

Equation 3: Should you not remove the 2 from each side?

Equation 3: Using lambda_3 in the indexing seems a bit odd. Since you are straying from Dziewonski et al. (1981) as well as Dahlen and Tromp (1998) could you explain this or cite a reference to this notation?

Line 68: do you know want $^3$?

Equation 4: What are the indices on the epsilons? Please write out the definitions of all your variables.

Equation 4: Why do you have the 2 on both sides? Could that not be removed?

Equation 5: Why do you have the factor of 4 in here? Should it not just be 1/N with N=3 as you are using three different moment tensors?

Lines 71 and 72: I don’t see an N-axis in the paper by Frohlich 1992, they refer to it as a B-axis.

Line 73: Your less-than sign is the opposite of that in Saloor and Okal (2018). Why is this? Given the large deviation in a standard set of conventions, I don't know if this is a typo, intended to be this way, or something else.

Figure 2: What is this a percentage of? Are the vertical axes the absolute value of the epsilon from each, then converted into a percent, or are they somehow related to epsilon bar defined in equation 4?

Line 83: The correlation coefficient of what? Sorry, I don't understand what you are taking the correlation coefficient of. Please clarify

Line 85: I don't think this is reproducibility, I think this is consistency. You aren't reproducing the experiment (E.g., using the same method and checking if you get the same answer), instead you are seeing how consistent the results are between different methods. Wouldn't you expect different results? One used long-period surface waves with a number of corrections (GCMT), which W-phase uses the fastest possible method to get a solution.

Line 86: I don't think this has to do with "reliability of determination," again W-phase isn't interested in subtle differences in moment tensor results, they want to have a moment tensor to issue a Tsunami warning as quickly as possible.

Figure 3: Would this not make for "large inconsistencies between methods?" The method could have a low uncertainty, but it might not be consistent.

Line 99: What is this percentage of? All three catalogs have an average within 1% of each other?

Line 106: Is Gaussian appropriate? It very well could be as the central limit theorem is likely in play, but I can't tell.

Figure 4: Isn't the standard deviation a measure of the spread of the data, not the reproducibility?

Line 116: Isn't the smallest eigenvalue lambda_2, based on line 68?

Line 117: I thought epsilon was a number between with absolute value between 0 and 1. Why a percentage?

Line 150: Are you sure this would be reliable? I am not sure that I follow this argument. Could you explain in more detail so that the reader could better understand the argument? I am not saying it is incorrect, but I just am not sure I understand what you are saying.

Line 167: I think reproducibility is the wrong word. Perhaps consistency.

Line 174: Wouldn't reliably determined have to do with how repeatable the result is with a fixed method and changes in the data, models, inversions (but fixed waveforms)? While you are looking at how consistent the numbers are?

Line 179: I think you want "real source processes" and not just real.

Line 191: I think you want consistent, not reliable.

References: Please put DOIs next to your references.

Please reference one of the ObsPy references, as you made use of their code. If you used something like GCMT, it would be good to reference that. You might also provide references to the papers that describe the methods of the inversions. You do this for the GCMT, but it would be good to include something from the GFZ method (Dreger and Helmberger, 1993, perhaps) and Duputel et al. (2012) for NEIC? I am not sure of the "official" one, but it would be good to reference these.

I think this paper could make a valuable contribution and do encourage this work be eventually published. However, I think the very inconsistent notation detracts from this paper. I would hope the authors could help make the paper more easily read by fixing this.

Peer-review: **‘When are Non-Double-Couple Components of Seismic Moment Tensors Reliable?’**

Dear authors and Seismica Editor,

1. **Summary of the work, methods, results, and conclusions.**

The manuscript entitled ‘When are Non-Double-Couple Components of Seismic Moment Tensors Reliable?’ performs a statistical analysis comparing the seismic moment tensors reported in GMCT, GFZ, and USGS catalogs. The aim is to see the reliability of the non-double couple components (NDCC) based on how correlated those components are among the three catalogs. This is how similar or dispersed the NDCC is for the events reported in the three catalogs. The authors observe mainly that:

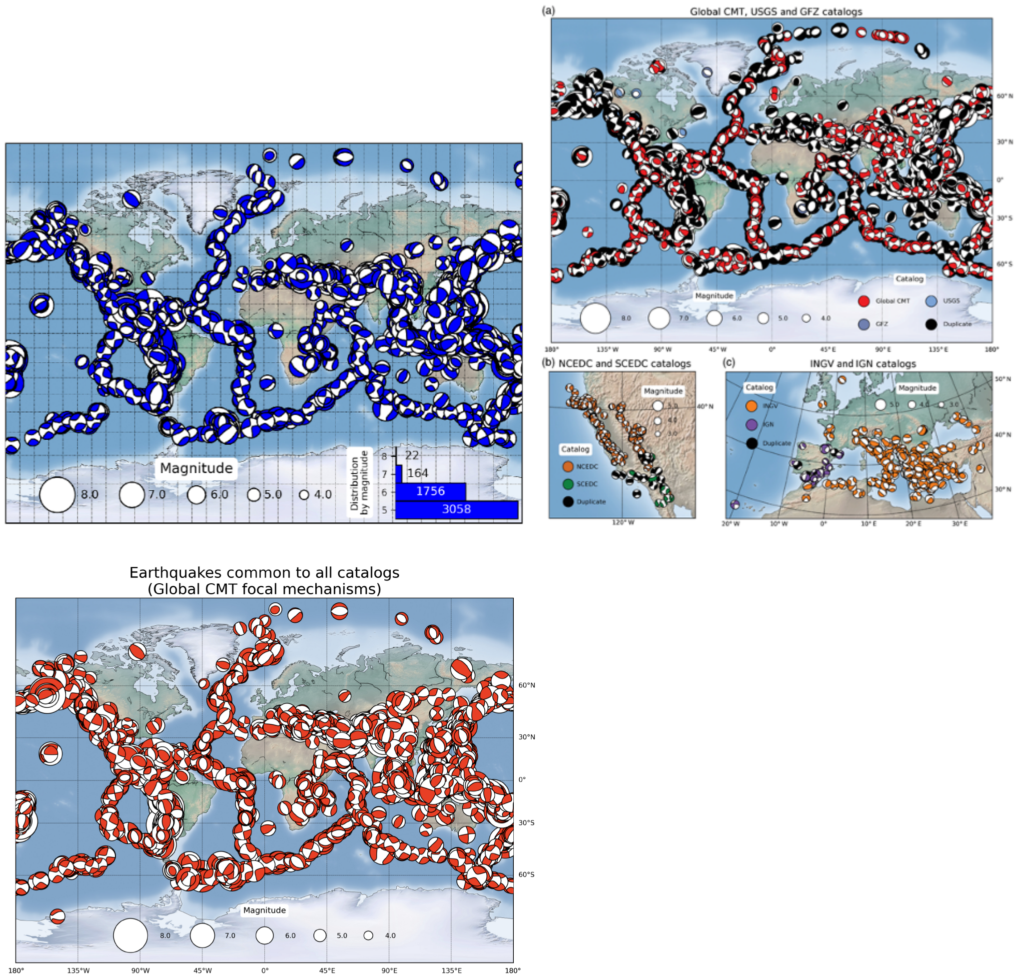
- The NDCC standard deviation decreases as the earthquake magnitude increase.
- The NDCC size decreases as the earthquake magnitude increase as well.
- There is no clear correlation between fault orientation (thrust, normal, strike-slip event) and NDCC.

In addition, the authors perform a synthetic experiment reproducing NDCCs randomly in order to verify a possible bias in the reduction of the NDCC standard deviation when this component percentage exceeds 60%. Considering that the limit for the NDCC is 100%, the highest NDCCs tend to be closer, which may result in an artifact reduction of the standard deviation.

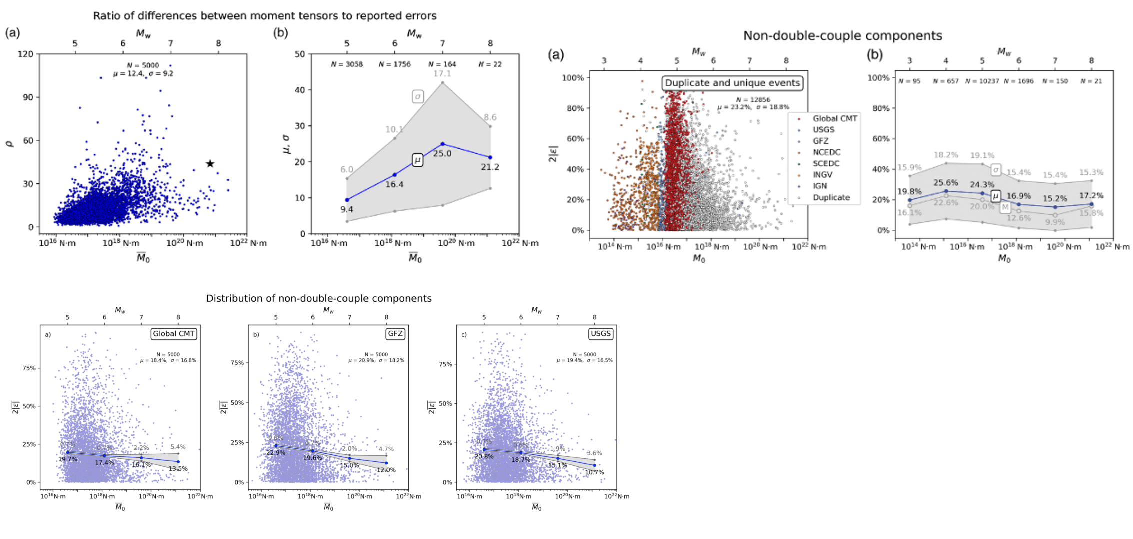
The authors conclude that only the NDDCs above 60% and the ones coming from large events are reliable.

2. **Impact and originality of the work**

The authors already explored a similar topic with similar datasets and methodologies in two published works: Rösler et al., 2021 and Rösler and Stein, 2022. The first publication also included regional moment tensor catalogs. It is striking how similar some of the figures, results, and conclusions are among the three works:




a) Rösler et al., 2021, b) Rösler and Stein, 2022 and c) current manuscript.




a) Rösler et al., 2021, b) Rösler and Stein, 2022 and c) current manuscript.

It is difficult not to think this work is repetitive compared to published ones. Indeed, many of the conclusions in previous works resemble or could lead to the conclusions presented in this manuscript.

For instance, in Rösler et al., 2021

- “The correlation between NDC components in the two catalogs suggests that NDC components of small earthquakes are less certain and often spurious. However, when NDC components are consistent, they likely represent real source processes, either a deviation from a DC or source complexity producing an apparent NDC component.”

- “For larger earthquakes, we expect a decreased difference in moment tensors found using different Earth models for both elastic and anelastic structures. Thus, the source mechanisms reported in different catalogs should agree better for large earthquakes, despite different inversion procedures.”

- “Hence, small earthquakes often show large NDC components, but many have large uncertainties and are likely to be artifacts of the inversion.”

Again, in Rösler and Stein, 2022

- “Here, we take a complementary approach by examining a large moment tensor dataset to assess how NDC components vary with earthquake magnitude, mechanism type, and geologic environment.”

- “Furthermore, there are at most small differences in NDC components between earthquakes with different faulting mechanisms and in different geologic environments.”

- “Hence, although some earthquakes have real NDC components, it appears that for most earthquakes, especially smaller ones, the NDC components are likely to be artifacts of the inversion.”

If in Rösler et al., 2021 and Rösler and Stein, 2022 the methods, datasets, and conclusions seem to overlap, I cannot see how this manuscript adds something extra (beyond the random generation of the NDCCs) or complementary to the aforementioned manuscripts. The conclusions of this study can be derived or are explicit in the authors' previous work.

## 3. Introduction

Some important references are missing in this section: The papers of Shuler et al., 2013 a and b reporting vertical CLVD components in active volcanoes and providing some plausible physical mechanisms that explain them.
In addition, when mentioning NDC moment tensors reported in volcanic environments, the authors mention some historical and well-known papers but omit to present the most recent works related to volcano events:

- Moment tensors for the Bárðarbunga caldera collapse reported in Gudmundsson 2016, and Rodriguez-Cardozo et al., 2021.
- The Sandambata et al., 2021 moment tensors reported for Sierra Negra caldera and Hawaii.

## 4. Methodology

The core of this study is the statistical analysis comparing the difference in the moment tensor. I am unsure how robust the variance reductions and averages calculated for a single event are. Each calculation consists of 3 points (the epsilon reported in the three catalogs), which may not be a significant number of samples for a reliable calculation. I would prefer to average the differences in the epsilon value (or even try to see the differences in the moment tensor angle) and the variance reduction of such differences in a certain magnitude range (e.g., the differences for all events between Mw. 4.8 and Mw. 5.0).

I am also surprised about not seeing a single mention in the methodology about the generation of random NDCCs for dealing with the bias in the standard deviation for large NDCCs. This is put in the results section without any prior advice that this will be done and an explanation about why it is necessary.

## 5. Results

The results section is hard to follow and seems to encompass contents that correspond to other sections, like the methodology, discussion, and conclusions, making this section too dense and unstructured.

Some sentences resemble a discussion. For instance, when the authors try to contrast their results with other studies that are their own previous works (lines 77-80, 89-90, 154-156).

In lines 149- 150, an explicit conclusion is mentioned. Therefore, this could go into the conclusion section.

The paragraph about how and why it is required to perform an experiment for generating aleatory NDCCs mentioned in lines 103 - 112 should go in the methodology.

## 6. Discussion and conclusions

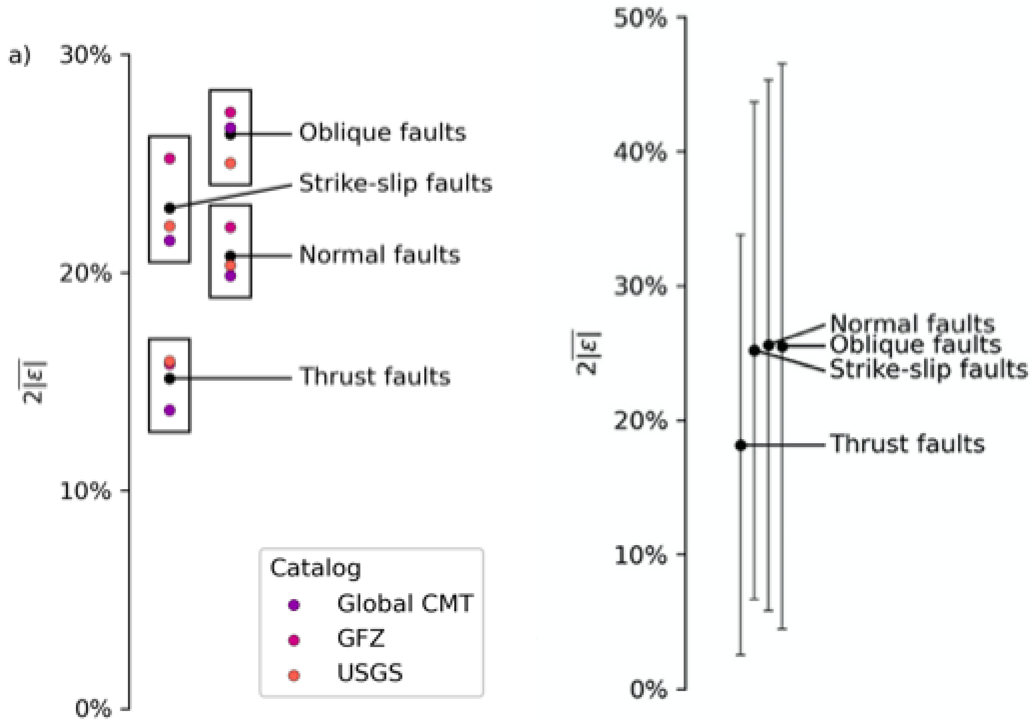
In my mind, the discussion is the soul of a manuscript because it gives the opportunity to the authors to interpret their results and contrast their perspective with other related works either to confirm, refute or complement previous findings. What is written in the discussion is not what I used to expect in that section. It is more a recall of the dataset employed, the results, and the conclusions already presented in the results. For instance, I do not see any specific reason for mentioning figure 7 in the discussion, but figure 5, which belongs to the same synthetic experiment, is mentioned in the results.

I will try to address what I consider are the discussions and conclusions written in the results section despite of considering that they are placed in the wrong place.

It is very odd that there is no single reference in addition to autoreferences in the entire discussion. In line 153 they wrote, ‘ This value is much larger than proposed 153 in earlier studies’, referring to the threshold percentage for considering reliable NDCC. However, they never mention o put a reference to the earlier studies.

The authors consider their findings consistent with their previous results using overlapping methods and data. Lines 77-80, 89-90, 154-156 are examples of autoreferences for confirming the validity or consistency of the results presented in this work. However, if current and past conclusions came from repeated results performed by the same author, it is difficult not to think such consistency may be biased.

The same issue mentioned previously about very similar or repeated graphics is present again in figure 6 in both, [Rösler and Stein., 2022]() and this work:




Left: this work and right [Rösler and Stein, 2022]() .

However, in this case, something more surprising occurred. [Rösler and Stein, 2022]() concluded: *‘Furthermore, there are at most small differences in NDC components between earthquakes with different faulting mechanisms and in different geologic* *environments. This consistency suggests that most NDC components do not reflect actual source processes, which would likely cause variability'*. Therefore, they claimed there is no clear trend between source orientation and NDCC. Conversely, the authors wrote in this manuscript: '[Rösler and Stein, 2022] *noticed that the size of NDC components varies with faulting type. Consistent with their observation, thrust-faulting earthquakes have the smallest NDC components on average of all different faulting types in our dataset.".* In conclusion, the authors cited themselves wrongly and concluded oppositely despite using similar datasets.

In line 195, they mention the size, but it is unclear what size they are referring to. Is it the size of the earthquake? If so, mentioning the magnitude is enough, and the size is wordy.

Finally, the conclusion is that only large events with NDCCs and events with large NDCCs are reliable (60%). So what is the threshold magnitude? Any NDCC larger than 60%, no matter the magnitude of the earthquake, is reliable? What about induced NDDC because of unmodeled waveforms because of the absence of a 3D velocity model? [Hejrani et al., 2017] showed a remarkable reduction in NDCCs for some events that GCMT and other catalogs reported as having high NDC (larger than 60%). This demonstrates that even if all the global catalogs agree with reporting important NDDC for a given large event, they may face the same limitation of using Green Functions based on a 1D velocity model, and hence, the NDDC is artificial.

## Reviewer 3

## Michael Pasyanos Pasyanos

Review of “When are Non-Double-Couple Components of Seismic Moment Tensors Reliable?” by authors Boris Rösler, Seth Stein, and Bruce Spencer for submission to Seismica. This manuscript uses comparisons of moment tensor solutions from several global earthquake catalogs to explore the reliability of these non-double-couple components, and conclude that they are only so when NDC components exceed 60%. In general, the study is straightforward, the analysis is clear, and the manuscript is well-written. However, I think there are a few items to address and opportunities to improve the manuscript, which I have listed below.

Particularly in the Introduction, you need to be more specific in qualifying that this is an analysis of NDCs for (presumed) earthquakes. In my research, we look at explosions and collapses which have significant NDC (both CLVD and isotropic) components. For example, line 30. "Because the isotropic components of *earthquake* MTs are generally small ..."

Full (six-element) moment tensor solutions (Minson and Dreger, 2008 GJI) are now being routinely calculated. In fact, using full MTs (allowing both ISO and CLVD) has been found to be a reliable way of identifying explosions (e.g. Ford et al., 2020 GJI; Pasyanos and Chiang, 2022 BSSA) and collapses (current research). I think there needs to be a discussion of the assumption that the NDC components in all of the catalogs utilized are limited to CLVD. How would relaxing this assumption to include full moment tensor solutions affect your results?

line 38. Is this Ross et al. (1999) that is in the Reference List.

It would be useful to define the terms used in the equations in the in-line text. For example, NDC component (<epsilon>), eigenvalues (<lambda>'), NDC standard deviation (<sigma>), etc. Otherwise, it is not always clear how the variables in the equations relate to the text.

lines 92-93. "... indicating that NDC components of large earthquakes are more reliable than the ones of small earthquakes." Please state why this is. I would speculate that the analysis for large earthquakes is performed at longer periods, where the velocity model is less variable, but that the analysis of smaller earthquakes is performed at shorter periods, where the velocity models are more variable, and often a poorer fit.

Figure 4b and lines 94-95. Why is it significant that the standard deviation decreases for NDC components larger than 50%, but not significant that it *decreases* for NDC components smaller than 50%. What does this mean? Also, I am think that a finer sampling of the statistics (mean, sd) with Mo and 2e in Figures 2, 4, 5ab, and 7ab might shed more light on the trends in the data.

lines 164-165. To the list of differences among the various catalogs should be added the velocity models that are used for the Green's functions, which is quite important.

Michael Pasyanos

I would recommend accepting with minor revisions.

Michael Pasyanos

## Response to Reviews

We recognize that our paper addresses a complex topic using a new multi-catalog “big data” approach, based on the fact that different moment tensor inversions give different estimates of the earthquake source, so comparing inversions gives insight into how well these quantities are determined. We appreciate the fact that two reviewers (#1 and 3) like this approach, and have addressed their suggestions that will clarify the paper. We have also clarified how this paper builds on its predecessors to address reviewer #2’s concern, and their specific points.

**Reviewer 1:**
Review of: When are non-double-couple components of seismic moment tensors reliable?
This paper compares differences in the non-double couple components of three different catalogs. This is an interesting problem and the authors do a good job investigating the initial pieces of this problem. While I think this will make a useful contribution in the field of seismology and more specifically moment tensor catalogs, I have a few suggestions before I recommend this for publication. I think addressing these could help make the paper more readable and have a higher impact overall.

***Major:***
I had a very hard time following the convention for the equations. There are a number of non-standard conventions and equations with potential typos. While some of these could be “correct” they do not enhance the readability of the paper. Perhaps the authors could pick a convention that is widely used (e.g., Dziewonki et al., 1981) and convert everything to these conventions. Or they could explain why their conventions deviate.

Unfortunately, there are no universally accepted conventions. Our definition of the NDC component (equation 3) follows Giardini (1984, page 903). Some later papers (Kuge and Kawakatsu (1993, equation 1), Frohlich and Davis (1999, equation 1) have changed the notation to the negative ratio. There seems to be no agreement about the sign of the NDC component. We have added the reference to our manuscript.

We have further clarified our equations by removing the factor of two from the definition of the NDC components (equation 1) and their mean (equations 2).

The words “reliable,” “repeatable,” “reproducible” etc. all have different meanings. It wasn’t clear to me that these were used consistently throughout. Please double check this.

Thank you for pointing this out. We have removed all instances of „reproducibility“ and replaced them with „consistency“. Because the consistency of NDC components between catalogs is a measure for their reliability, we continue to distinguish these.

Please put more detail into the conclusions and introduction about why this work is important and how it fits into seismology. While all three catalogs have different approaches, we would eventually hope we get the same answer from all three methods. While your work does a good job pointing out that they might not be consistent, it could have a higher impact of pointing out why this is important.

We have clarified the importance of our work at the end of the introduction.

***Minor:***
Lines 26 through 28: Please double check that this sentence is correct. I believe the first systematic studies of MTs came from Dziewonski, Chou, and Woodhouse (1981). This would be analog data. If these were "automatic" or not is up to debate, but I think MTs were being systematically studied on analog records. Anyway, just double check that this is correct. I don't know enough about MT history to say either way.

Dziewonski et al.'s (1981) paper describes the methodology of a moment tensor inversion process that minimizes the misfit between observed and synthetic waveforms. They explain that "This scale of the analysis has become feasible because of the availability of the high-quality data from the recently established global digital seismic network ." They describe their procedure as automatic, using digital (SRO & ASRO) data, but in practice requiring initial visual examination of records due to glitches and polarity issues. Their approach has been adopted by most moment tensor catalogs and is still used by the Global CMT Project (Ekström et al., 2012). Dziewonski et al.'s paper ignores NDC components. Giardini's (1984) analysis of 200 earthquakes between 1977 and 1980 may be the first study that highlights NDC components in seismic moment tensors. However, catalogs were still too small for a systematic analysis of NDC components in moment tensor catalogs, which is why we believe Frohlich (1994) to be the first example of the analysis of NDC components in moment tensor catalogs.

Equation 1: Please identify your variables. What is M_ij

We have added a definition for M_ij in line 68.

Equation 2: Please identify your variables what is M_0 and M_w, please cite where this comes from. I believe it is Kanamori (1977).

Kanamori (1977) defines the scalar moment, whereas the cited Silver and Jordan (1982) describes the calculation of the scalar moment from the moment tensor components. We have added the reference to Kanamori (1977) for the definition of the moment magnitude.

Equation 3: Should you not remove the 2 from each side?

We have removed the factor of 2 from both sides for clarity.

Using lambda_3 in the indexing seems a bit odd. Since you are straying from Dziewonski et al. (1981) as well as Dahlen and Tromp (1998) could you explain this or cite a reference to this notation?

We use Stein and Wysession's notation where eigenvalues are ordered by absolute value, which followed Jost and Herrmann (1989, page 42), and Hudson et al.'s (1989, equation 1) definition of the eigenvalues for source-type plots, and indicate this in the manuscript. This is a common mathematical usage in many applications involving eigenvalues (e.g. principal stresses) and consistent with our previous paper, Rösler and Stein (2022).

Line 68: do you know want ³?

This definition of the eigenvalues is consistent with Stein and Wysession (2003), Jost and Herrmann (1989), and Hudson et al. (1989).

Equation 4: What are the indices on the epsilons? Please write out the definitions of all your variables.

We have changed the formula and added an explanation for the index of summation.

Equation 4: Why do you have the 2 on both sides? Could that not be removed?

We have removed the factor of 2 from both sides for clarity.

Equation 5: Why do you have the factor of 4 in here? Should it not just be 1/N with N=3 as you are using three different moment tensors?

NDC components are generally reported as 2ε, for example in the USGS catalog, ranging from -100% to 100%, as proposed by Jost and Herrmann (1989). The factor of 4 is a consequence of calculating the standard deviation among the values for 2ε. The equation is correct.

Lines 71 and 72: I don't see an N-axis in the paper by Frohlich 1992, they refer to it as a B-axis.

There is no agreement in the name for this axis. Some papers use „N" for the *null* axis, others „B". We use „N" for consistency with our previous paper (Rösler and Stein , 2022) and more recent studies, and note in the text that "B" is sometimes used in line 80.

Line 73: Your less-than sign is the opposite of that in Saloor and Okal (2018). Why is this? Given the large deviation in a standard set of conventions, I don't know if this is a typo, intended to be this way, or something else.

Thank you for catching this. This is a typo and should have been a greater-or-equal sign. We have corrected it.

Figure 2: What is this a percentage of? Are the vertical axes the absolute value of the epsilon from each, then converted into a percent, or are they somehow related to epsilon bar defined in equation 4?

2ε ranges from -100% to 100%. Therefore, its mean absolute value ranges from 0 to 100%.

Line 83: The correlation coefficient of what? Sorry, I don't understand what you are taking the correlation coefficient of. Please clarify

In line 81 (now 90), we write: „the values of the NDC components for earthquakes in the three catalogs are only weakly correlated between catalogs". „Correlation coefficients" in line 83 (now 91/92) refer to these values. We have clarified the text in lines 92 and 93.

Line 85: I don't think this is reproducibility, I think this is consistency. You aren't reproducing the experiment (E.g., using the same method and checking if you get the same answer), instead you are seeing how consistent the results are between different methods. Wouldn't you expect different results? One used long-period surface waves with a number of corrections (GCMT), which W-phase uses the fastest possible method to get a solution.

Good point. We have changed the wording.

Line 86: I don't think this has to do with "reliability of determination," again W-phase isn't interested in subtle differences in moment tensor results, they want to have a moment tensor to issue a Tsunami warning as quickly as possible.

Despite using different frequency ranges for their inversions, the moment tensor catalogs should yield the same source mechanisms for the same earthquakes in order to be reliable. If one of the moment tensor catalogs was interested only in tsunami warnings, it could be constrained to DC source mechanisms. However, the USGS which uses the W-phase explicitly reports deviatoric moment tensors which include NDC components, making it possible to compare them to the NDC components of other moment tensor catalogs and to assess their reliability of determination.

Figure 3: Would this not make for "large inconsistencies between methods?" The method could have a low uncertainty, but it might not be consistent.

The methods yield inconsistent results in the NDC components. The methods are not necessarily inconsistent, which we cannot assess without knowing the details of the inversion process (see Rösler et al, 2022). The weak correlations between NDC components are therefore an expression of large uncertainties in their determination because they are not consistently determined among the three catalogs.

Line 99: What is this percentage of? All three catalogs have an average within 1% of each other?

We expanded the text to clarify that this paragraph illustrates the ceiling effect - as the mean NDC component gets close to its maximum of 100%, the standard deviation of the different estimates is required to be small.

Line 106: Is Gaussian appropriate? It very well could be as the central limit theorem is likely in play, but I can't tell.

In the absence of constraints for a random distribution, we use the common Gaussian distribution.

Figure 4: Isn't the standard deviation a measure of the spread of the data, not the reproducibility?

This is correct. As stated in line 84, „the standard deviation of the NDC components for each earthquake in the catalogs is a measure of the NDC component's reproducibility and can be used to assess the reliability of its determination."

Line 116: Isn't the smallest eigenvalue lambda_2, based on line 68?

Correct. We meant the „absolutely smallest eigenvalue" lambda 3 and have corrected this.

Line 117: I thought epsilon was a number between with absolute value between 0 and 1. Why a percentage?

We follow the USGS catalog in reporting the fraction of DC component as a percentage, as proposed by Jost and Herrmann (1989).

Line 150: Are you sure this would be reliable? I am not sure that I follow this argument. Could you explain in more detail so that the reader could better understand the argument? I am not saying it is incorrect, but I just am not sure I understand what you are saying.

The argument of our paper is based on our statement in line 84 (now 93): „Hence the standard deviation of the NDC components for each earthquake in the catalogs is a measure of the NDC component's reproducibility and can be used to assess the reliability of its determination." The more consistently an NDC component is determined among the three catalogs, the higher its reliability.

Line 167:  I think reproducibility is the wrong word.  Perhaps consistency.

We have changed the wording.

Line 174: Wouldn't reliably determined have to do with how repeatable the result is with a fixed method and changes in the data, models, inversions (but fixed waveforms)? While you are looking at how consistent the numbers are?

Moment tensor catalogs currently do not report on the reliability of their results. The only catalog that reports errors is the Global CMT Project, and those errors are based only on the misfit between observed and synthetic waveforms, which vastly underestimates the true uncertainty. Measures for the uncertainty of the NDC components, and the stability of the inversion, are missing completely. Therefore, we assume the consistency of the NDC components to be a measure for their reliability.

Line 179: I think you want "real source processes" and not just real.

Ok.

Line 191: I think you want consistent, not reliable.

In this case, we refer to the reliability as derived from the consistency.

References: Please put DOIs next to your references.

Done.

Please reference one of the ObsPy references, as you made use of their code.  If you used something like GCMT, it would be good to reference that.  You might also provide references to the papers that describe the methods of the inversions.  You do this for the GCMT, but it would be good to include something from the GFZ method (Dreger and Helmberger, 1993, perhaps) and Duputel et al. (2012) for NEIC?  I am not sure of the "official" one, but it would be good to reference these.

Through personal communication with Joachim Saul from the GFZ, we learned that there is no peer-reviewed paper about the methodology of the moment tensor inversion used by the GFZ: „There is no paper about the methodology. We use a software developed in collaboration with Gempa": https://www.gempa.de/

We have added Hayes et al. (2009) as reference for the methodology of the USGS catalog, and Beyreuther et al. (2010) as reference for ObsPy.

Recommendation: Resubmit for Review

------------------------------------------------------

**Reviewer 2:**
The comments are sent in the attached file named revision_Rosler_et_al_2022.pdf.
Recommendation: Decline Submission

***1. Summary of the work, methods, results, and conclusions.***

The manuscript entitled 'When are Non-Double-Couple Components of Seismic Moment Tensors Reliable?' performs a statistical analysis comparing the seismic moment tensors reported in GMCT, GFZ, and USGS catalogs. The aim is to see the reliability of the non-double couple components (NDCC) based on how correlated those components are among the three catalogs. This is how similar or dispersed the NDCC is for the events reported in the three catalogs. The authors observe mainly that:

- The NDCC standard deviation decreases as the earthquake magnitude increase.
- The NDCC size decreases as the earthquake magnitude increase as well.
- There is no clear correlation between fault orientation (thrust, normal, strike-slip event) and NDCC.

In addition, the authors perform a synthetic experiment reproducing NDCCs randomly in order to verify a possible bias in the reduction of the NDCC standard deviation when this component percentage exceeds 60%. Considering that the limit for the NDCC is 100%, the highest NDCCs tend to be closer, which may result in an artifact reduction of the standard deviation.

The authors conclude that only the NDDCs above 60% and the ones coming from large events are reliable.

***2. Impact and originality of the work***

The authors already explored a similar topic with similar datasets and methodologies in two published works: Rösler et al., 2021 and Rösler and Stein, 2022. The first publication also included regional moment tensor catalogs. It is striking how similar some of the figures, results, and conclusions are among the three works:

…

It is difficult not to think this work is repetitive compared to published ones. Indeed, many of the conclusions in previous works resemble or could lead to the conclusions presented in this manuscript.

…

If in Rösler et al., 2021 and Rösler and Stein, 2022 the methods, datasets, and conclusions seem to overlap, I cannot see how this manuscript adds something extra (beyond the random generation of the NDCCs) or complementary to the aforementioned manuscripts. The conclusions of this study can be derived or are explicit in the authors' previous work.

We are sorry if we have not made clear how this study differs from our previously published papers. In our previous studies (Rösler et al., 2021; Rösler and Stein, 2022), we showed that most NDC components, especially of smaller earthquakes, are likely to be artifacts of the inversion. In this study, we derive objective criteria for when NDC components likely represent real source processes. This study was motivated by the fact that various authors have proposed different thresholds for when NDC components should be considered real.

This paper differs in methodology and data from our previous papers: In Rösler et al. (2021), we used 5000 moment tensors for the same earthquakes in the GCMT and USGS catalogs to derive an estimate for their uncertainties. In Rösler and Stein (2022), the moment tensors in our dataset contained both unique earthquakes and earthquakes for which several catalogs provided a moment tensor. Neither dataset made it possible to calculate the standard deviation for the NDC components and hence derive criteria for their reliability. Thus, in this study, we use the consistency of NDC components between catalogs as a measure for their reliability. The conclusions confirm what we speculated about in our previous papers.

Because this paper builds on our previous papers, we have used a similar style for the figures for clarity.

### *3. Introduction*

Some important references are missing in this section: The papers of Shuler et al., 2013 a and b reporting vertical CLVD components in active volcanoes and providing some plausible physical mechanisms that explain them.

In addition, when mentioning NDC moment tensors reported in volcanic environments, the authors mention some historical and well-known papers but omit to present the most recent works related to volcano events:

- Moment tensors for the Bárðarbunga caldera collapse reported in Gudmundsson 2016, and Rodriguez-Cardozo et al., 2021.

- The Sandambata et al., 2021 moment tensors reported for Sierra Negra caldera and Hawaii.

Thank you for pointing us to those recent studies. We have added them to the references.

### *4. Methodology*

The core of this study is the statistical analysis comparing the difference in the moment tensor. I am unsure how robust the variance reductions and averages calculated for a single event are. Each calculation consists of 3 points (the epsilon reported in the three catalogs), which may not be a significant number of samples for a reliable calculation. I would prefer to average the differences in the epsilon value (or even try to see the differences in the moment tensor angle) and the variance reduction of such differences in a certain magnitude range (e.g., the differences for all events between Mw. 4.8 and Mw. 5.0).

I am also surprised about not seeing a single mention in the methodology about the generation of random NDCCs for dealing with the bias in the standard deviation for large NDCCs. This is put in the results section without any prior advice that this will be done and an explanation about why it is necessary.

We think that this suggested procedure is equivalent to our approach of calculating the standard deviation and presents the same information.

***5. Results***

The results section is hard to follow and seems to encompass contents that correspond to other sections, like the methodology, discussion, and conclusions, making this section too dense and unstructured.
Some sentences resemble a discussion. For instance, when the authors try to contrast their results with other studies that are their own previous works (lines 77-80, 89-90, 154-156).

In lines 149- 150, an explicit conclusion is mentioned. Therefore, this could go into the conclusion section.

The paragraph about how and why it is required to perform an experiment for generating aleatory NDCCs mentioned in lines 103 - 112 should go in the methodology.

No other reviewer has requested changes to the structure of our manuscript. Reviewer 1 said "the authors do a good job investigating the initial pieces of this problem… this will make a useful contribution in the field of seismology.“ Reviewer 3 noted that “the study is straightforward, the analysis is clear, and the manuscript is well-written”. Therefore, we have retained the present structure.

***6. Discussion and conclusions***

In my mind, the discussion is the soul of a manuscript because it gives the opportunity to the authors to interpret their results and contrast their perspective with other related works either to confirm, refute or complement previous findings. What is written in the discussion is not what I used to expect in that section. It is more a recall of the dataset employed, the results, and the conclusions already presented in the results. For instance, I do not see any specific reason for mentioning figure 7 in the discussion, but figure 5, which belongs to the same synthetic experiment, is mentioned in the results. I will try to address what I consider are the discussions and conclusions written in the results section despite of considering that they are placed in the wrong place.

It is very odd that there is no single reference in addition to autoreferences in the entire discussion. In line 153 they wrote, ' This value is much larger than proposed  in earlier studies', referring to the threshold percentage for considering reliable NDCC. However, they never mention o put a reference to the earlier studies.

The authors consider their findings consistent with their previous results using overlapping methods and data. Lines 77-80, 89-90, 154-156 are examples of autoreferences for confirming the validity or consistency of the results presented in this work. However, if current and past conclusions came from repeated results performed by the same author, it is difficult not to think such consistency may be biased.

Our approach of using large moment tensor catalogs to derive general properties of earthquake source mechanisms is the same as in our previous papers, but the specific methods and data used here differ, as mentioned before.

The same issue mentioned previously about very similar or repeated graphics is present again in figure 6 in both, Rösler and Stein., 2022 and this work:

<figure>

Left: this work and right Rösler and Stein, 2022 .

However, in this case, something more surprising occurred. Rösler and Stein, 2022 concluded: 'Furthermore, there are at most small differences in NDC components between earthquakes with different faulting mechanisms and in different geologic environments. This consistency suggests that most NDC components do not reflect actual source processes, which would likely cause variability'. Therefore, they claimed there is no clear trend between source orientation and NDCC. Conversely, the authors wrote in this manuscript: 'Rösler and Stein, 2022 noticed that the size of NDC components varies with faulting type. Consistent with their observation, thrust-faulting earthquakes have the smallest NDC components on average of all different faulting types in our dataset.". In conclusion, the authors cited themselves wrongly and concluded oppositely despite using similar datasets.

We have corrected our statement for consistency with our previous paper in lines 165-166.

In line 195, they mention the size, but it is unclear what size they are referring to. Is it the size of the earthquake? If so, mentioning the magnitude is enough, and the size is wordy.

We refer to „the variation of NDC components with earthquake magnitude and size“ as stated in the manuscript.

Finally, the conclusion is that only large events with NDCCs and events with large NDCCs are reliable (60%). So what is the threshold magnitude?

It is not possible to determine a magnitude threshold from our current dataset as the standard deviation decreases monotonically with magnitude. Defining a threshold for the standard magnitude would mean to speculate about how small the standard deviation must be in oder for the NDC components to represent real source processes.

Any NDCC larger than 60%, no matter the magnitude of the earthquake, is reliable?

This is a misunderstanding of our statement that „ an NDC component greater than 60% is *likely* to reflect a real source process“. We made a general statement about the reliability of NDC components, not about the NDC component of the MT of every earthquake.

What about induced NDDC because of unmodeled waveforms because of the absence of a 3D velocity model?

We attributed the appearance of spurious NDC components to the inversion based on Green's function generated for 1D Earth models in another manuscript submitted to Seismica by Rösler et al. (2022), which is currently in review. There, we generate synthetic seismograms for the DC component of earthquakes for one Earth model, and invert them using Green’s functions generated for a different Earth model. The NDC components obtained during that study have a distribution (mean and standard deviation) similar to the ones observed in MT catalogs, confirming our argument that most NDC components in MT catalogs are artifacts of the inversion.

Hejrani et al., 2017 showed a remarkable reduction in NDCCs for some events that GCMT and other catalogs reported as having high NDC (larger than 60%). This demonstrates that even if all the global catalogs agree with reporting important NDDC for a given large event, they may face the same limitation of using Green Functions based on a 1D velocity model, and hence, the NDDC is artificial.

To avoid such bias, we use a large dataset containing MTs of 5000 earthquakes in three different catalogs.

------------------------------------------------------

**Reviewer 3:**

Review of “When are Non-Double-Couple Components of Seismic Moment Tensors Reliable?” by authors Boris Rösler, Seth Stein, and Bruce Spencer for submission to Seismica. This manuscript uses comparisons of moment tensor solutions from several global earthquake catalogs to explore the reliability of these non-double-couple components, and conclude that they are only so when NDC components exceed 60%.

In general, the study is straightforward, the analysis is clear, and the manuscript is well-written. However, I think there are a few items to address and opportunities to improve the manuscript, which I have listed below.

Particularly in the Introduction, you need to be more specific in qualifying that this is an analysis of NDCs for (presumed) earthquakes. In my research, we look at explosions and collapses which have significant NDC (both CLVD and isotropic) components. For example, line 30. "Because the isotropic components of *earthquake* MTs are generally small …"

Thank you for pointing it out. We have added the word.

Full (six-element) moment tensor solutions (Minson and Dreger, 2008 GJI) are now being routinely calculated. In fact, using full MTs (allowing both ISO and CLVD) has been found to be a reliable way of identifying explosions (e.g. Ford et al., 2020 GJI; Pasyanos and Chiang, 2022 BSSA) and collapses (current research). I think there needs to be a discussion of the assumption that the NDC components in all of the catalogs utilized are limited to CLVD. How would relaxing this assumption to include full moment tensor solutions affect your results?

Good point. We have added a discussion of this effect to our introduction.

line 38. Is this Ross et al. (1999) that is in the Reference List.

Our reference was wrong. We wanted to cite Ross, A., Foulger, G. R., & Julian, B. R. (1996). *Non-double-couple earthquake mechanisms at The Geysers geothermal area, California*. Geophysical Research Letters, **23**, no. 8, 877-880 and have updated the reference list.

It would be useful to define the terms used in the equations in the in-line text. For example, NDC component (<epsilon>), eigenvalues (<lambda>'), NDC standard deviation (<sigma>), etc. Otherwise, it is not always clear how the variables in the equations relate to the text.

We now define these terms repeatedly.

lines 92-93. "... indicating that NDC components of large earthquakes are more reliable than the ones of small earthquakes." Please state why this is. I would speculate that the analysis for large earthquakes is performed at longer periods, where the velocity model is less variable, but that the analysis of smaller earthquakes is performed at shorter periods, where the velocity models are more variable, and often a poorer fit.

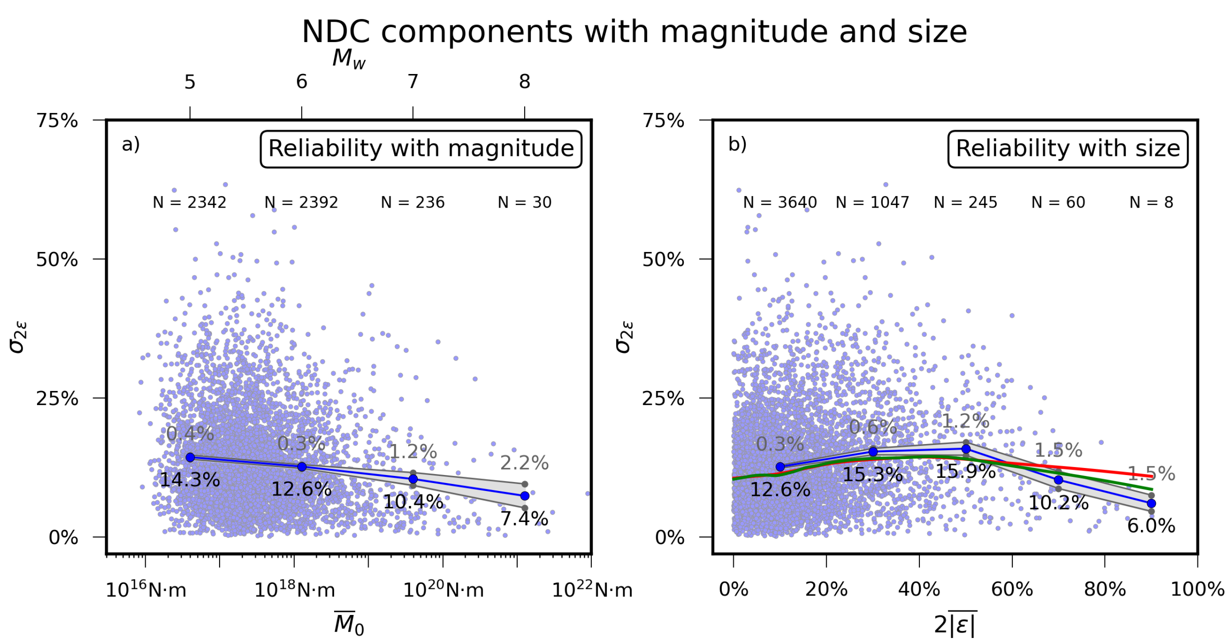
We have clarified our reasoning, which is even simpler: Because the size of the NDC components vary only slightly for earthquakes of different magnitudes (Fig. 2), the decrease in their standard deviation cannot be due to their size, and we instead attribute this decrease to the magnitude of the earthquake in the absence of a ceiling constraints. Therefore, NDC components of large earthquakes are more reliably determined than the ones of small earthquakes. We agree with the reviewer that this is likely an effect of the longer periods used for larger earthquakes and have added the paragraph in lines 102-104.

Figure 4b and lines 94-95. Why is it significant that the standard deviation decreases for NDC components larger than 50%, but not significant that it *decreases* for NDC components smaller than 50%. What does this mean?

The decrease of standard deviation for smaller NDCs is about 3 percentage points, whereas the decrease for large NDCs is about 10 percentage points, and we thus do not consider this decrease significant. Nevertheless, it may be a real phenomenon for which we currently can only speculate about an explanation.

Also, I am think that a finer sampling of the statistics (mean, sd) with Mo and 2e in Figures 2, 4, 5ab, and 7ab might shed more light on the trends in the data.

A LOWESS regression (see figure below) shows a very similar behavior of the standard deviation with size of NDC component as the binned values. The red curve shows the LOWESS regression with the standard value of 2/3 of all values considered, the green curve with 1/3 of all values. The inclusion of less points makes variations more visible, which is why the green curve shows a stronger variation with NDC size. Further reduction of this fraction shows only very subtle changes in the variation of the standard deviation with NDC size. The difference between the LOWESS regression and the binned values is so small that we decided not to include the LOWESS regression in the figure for clarity. The same argument applies for smaller bin sizes.



lines 164-165. To the list of differences among the various catalogs should be added the velocity models that are used for the Green's functions, which is quite important.

We now note that all three catalogs use PREM for their Green's functions and have added this statement to our manuscript.

Recommendation: Revisions Required