**Round 1**

**Reviewer A**

In this paper the authors consider modern station distributions, location catalogs and location algorithms to update the ISC criteria for identifying and certifying ground truth events for the IASPEI Reference Events List. They suggest several modifications over the current criteria which they show increases the number and improve the geographic distribution of certified ground truth events.

The paper is concise, well organized and generally clearly written, and forms an important contribution to identifying and cataloging ground-truth seismicity.

I have only a few comments and also indicate some minor issues.

Comments:

Introduction, History of Ground Truth Events – All of the references are for the authors' work. Are there no other important identification procedures and catalogs concerning ground-truth events?

Section 1.2 – I suggest adding some description the ISCloc procedure and inversion algorithm in general, plus short, algorithmic description of all specific features important tp the current study. This would give more context for later analyses and results in the paper, and avoid that the reader needs to be familiar with ISCloc or consult the 2011 reference.

The semi-major axis of the error ellipse gives posterior information on the nominal epicenter constraint. A standard error on depth (e.g. from the full error ellipsoid) would give equivalent information on depth constraint. Instead, it seems that information on depth constraint in this study is only binary - whether the event depth is resolved (free) or not resolved (fixed). In addition, this information is not posterior to location, but mixed with the location process. Some mention and discussion of these issues (beyond the last two sentences in the conclusions), and why some kind of posterior information on depth constraint is not used or not available, would be of interest and a valuable addition to the study.

Line 44-46: You could mention that GT events are also very important as reference or anchor for calibrating event absolute locations (epicenter and/or depth). Maybe there are other key uses for GT events that could be listed.

Line 60: "ΔU" has not yet been defined in the main text – at least a short phrase summarizing this measure is probably needed here, like lines 78-79.

Line 110: “A feature of the ΔU criteria…” - at this point, and for this and following paragraphs, I think a full definition and mathematical description of ΔU is needed.

Line 113: “(e.g. Figure 1, bottom)” – I think there is not yet enough information on ΔU, uniform azimuths and CPQ for the reader to make much sense of the meaning of Figure 1.

Line 147: “In contrast, the values of CPQ are very similar for the two events and requires no stations to be removed.” – It seems here also that a full definition and mathematical description of CPQ is needed before discussing results based on CPQ.

Line 179: It would be interesting to have a very concise explanation of the terms in this equation and the mechanics of the calculation. Not at all evident from a quick glance.

Figure 4: You might explain why some of the linear structures in the CPQ plot show increased mislocation with increasing CPQ. This seems an odd and potentially interesting and relevant relation.

Line 311: “An additional constraint for selecting GT events after relocation is a requirement for the locator to be able to resolve the event depth. If this is not the case and the depth is set to a fixed depth, then the event is rejected.” and Line 323: “It is clear from this figure that having one or more stations within 10 km of the event significantly improves the potential depth is resolution.” – This conclusion depends on how the locator determines to set a fixed depth, since fixed depth is taken as “unresolved” depth. So a description of how/when depth is fixed seems necessary. In particular, there would be a problem with the conclusions if fixing the depth depends directly on number of stations within 10 km or on number of stations with P and S readings.

Suggested minor issues to address:

Line 47-51: A bit choppy, I suggest modifying to something like:
“GT events are typically obtained in one of three ways: 1) explosions or other anthropogenic events, where the location and timing of the event is known or well documented by nearby observations; 2) seismic events that are well located by a dense and well distributed near-field seismic network; 3) seismic events that are well located by a multi-event location technique.”

Line 133: I suggest more information on “uniform event to station azimuth” is needed. Is it the corresponding uniform azimuth in a sorted list of N uniform azimuths? Maybe add the definition from Bondár and McLaughlin (2009a): unifi = 360i / N for i = 0,..., N – 1

Figure 3 caption: “CPQ is given by the ratio enclosed by the red polygon, and the area enclosed by the black circle.” → “CPQ is given by the ratio of the area enclosed by the red polygon to the area enclosed by the black circle.”

Figure 6 caption: Description of Top and Bottom should use the same ordering, e.g. “Semi-major axis of error ellipse with respect to CPQ (left) and ΔU (right)”

Line 291: “ΔU” → “secondary gap”

Figure 7 caption: All panels might best be described as “Semi-major axis of the error ellipse with respect to XXX”

Line 336: “bottom” → “bottom left”

Line 343: “For these events there are no stations reporting P & S phases and no stations within 10 km only ...” → “For these events, when there are no stations reporting P & S phases only ...”

Line 370: “this” → “the above criteria”

Line 400: “algorithms” → “distributions”

Line 411: “resolving for the event depth” → “that event depth is resolved (not fixed)”

Anthony Lomax

**Reviewer B**

Review for the Seismica submission: Revising the Seismic Ground Truth Reference Event Identification Criteria.

This paper proposes an updated criterion, Cyclic Polygon Quotient (CPQ), for assessing the Ground Truth (GT) seismic event locations. The authors compare CPQ with ΔU in terms of mislocated explosion distributions, semi-major axes of error ellipses, and categorized event distributions, demonstrating that CPQ is more flexible and effective for identifying GT events. Additionally, the manuscript incorporates the semi-major axis of the error ellipse from the ISCloc location program as a metric for location quality and explores how the number of stations reporting both P and S phases correlates with depth uncertainty.

The manuscript is generally clear, well-written, and accessible. It addresses a topic of significant interest to the Seismica community by presenting a catalog of precise, high-quality global earthquake locations. However, there are several issues and suggestions the authors should address before the manuscript is accepted. The most significant concern is that the use of the semi-major axis of the uncertainty ellipse and the role of P and S phase stations in constraining depth are not novel contributions to the field. To maximize the manuscript's impact, given the new proposed criterion and the convenience of accessing global dataset, I would recommend the authors switch the focus more on the new GT event catalog, illustrate more about the difference between new and old catalog, provide access to the catalog and code example of calculating the CPQ.

Suggestions:

1. The introduction of the \Delta U and its limitation is well-written; but the structure of the *Introduction*and *Quantifying Unbalanced Station Distributions* sections could be improved. Introducing the equation for ΔU early in the *Introduction* would provide better context and help readers understand why ΔU becomes less effective as the number of stations increases.
2. A brief explanation of how ISCloc locates events and calculates uncertainty ellipses is needed. Since the semi-major axis of the uncertainty ellipse is proposed as a quality metric, it is important to clarify the confidence level of this metric, e.g. 95%, 99%. The authors mention "90% semi-major axis" in Figure 5, but it is unclear whether this refers to 90% of the axis length or a 90% confidence interval. Yu et al. (2024) highlighted that most earthquake location programs do not output 95% confidence ellipses, and various statistical assumptions are used to derive them. Clarifying this for ISCloc would strengthen the manuscript.
3. If the ellipse is of 95% confidence interval, it is unsurprising that the 93.2% of the known mislocated explosions falls under the curve in Figure 5., as this aligns with how confidence interval function. This finding instead supports the validity of ISCloc's uncertainty evaluation and its use as a metric for location quality. The authors should rephrase this discussion to avoid overstating the result.

4. Gomberg et al., 1990 has demonstrated the effectiveness of S-wave arrival times on the depth determination from both observation and numerical experiment. This work aligns with the authors' argument that the stations with P&S phases help constrain the depth. The authors should refer to this article in that part at least. Indeed, the paper would benefit a lot from a comprehensive engagement with existing literatures. More references in Introduction, Section 3-5, would better contextualize this paper with the existing body of works.
5. Per Seismica's data policy, the new GT catalog produced using the CPQ criterion should be made public available. It would be better if you can share a demo code that generate the catalog using ISC's data.
6. The figures in the manuscript are low resolution. If this is due to preprint formatting, it should be clarified. If not, the authors need to produce high-resolution versions of all figures for publication.

Minor issues:

1. Some acronyms (e.g., NEIC, ISC, GNS) are missing their full names.
2. Line 133: typo - easz_i should be esaz_i.
3. Line 179-180: typo – "easz" is misused.

Reference:

Yu, Y., Ellsworth, W. L. & Beroza, G. C. Accuracy and Precision of Earthquake Location Programs: Insights from a Synthetic Controlled Experiment. *Seismological Research Letters* (2024) doi:10.1785/0220240354

Gomberg, J. S., Shedlock, K. M. & Roecker, S. W. The effect of S-wave arrival times on the accuracy of hypocenter estimation. *Bulletin of the Seismological Society of America* **80**, 1605–1628 (1990).

Reviewer 1:

Introduction, History of Ground Truth Events – All of the references are for the authors' work. Are there no other important identification procedures and catalogs concerning ground-truth events?

We have focused on highlighting the papers which show the development of single event ground-truth event criteria over time. The need for the development of ground truth criteria initially arose from the need to assess the performance of 3D velocity models in event location accuracy, especially in the context of the Comprehensive-Test-Ban-Treaty verification. We have neglected to include other methodologies which establish ground-truth events using either waveforms, multi-event base processing, or analysis of InSAR imagery. We agree that this is an oversight and that including references to these provides useful context for the reader. We have thus amended the text as follows:

*Lines 44-57*

***"Seismic events with well constrained locations and origin times, referred to as "Ground Truth" (GT) events, are an important resource allowing seismologists to test and validate new techniques relating to seismic event locations, such as location algorithms and velocity models (e.g. Begnaud et al. 2021). As we demonstrate in this study GT events are of particular use for calibrating the absolute event locations (e.g. Belinić & Markušić, 2017). Additionally GT events can be used as the seed locations for multi-event relocation techniques (e.g. Bergman et al. 2022; Bondár et al., 2024).***

***GT events can be defined in many ways: by explosions or other anthropogenic events, where the location and timing of the event is known or well documented by nearby observations (e.g. Bennett et al. 2010; Bittner et al. 2023); seismic events that are well located by a dense and well distributed near-field seismic network (e.g. Bondár and McLaughlin, 2009a; Boomer et al 2010); seismic events well located by a multi-event location technique (e.g. Bondár et al. 2008; Bergman et al. 2022). Additionally there are less well established methods of obtaining GT events including using ambient seismic noise (e.g. Zeng et al. 2015) and InSAR (e.g. Zhu et al. 2021)."***

*Lines 81-84*

***"The standardisation of GT event criteria was achieved by an iterative process of criteria refinement and publication (e.g. Sweeney 1996; Sweeney 1998; Bondár et al. 2001; Bondár et al. 2004a; Bondár et al. 2004b) leading to a broadly accepted set of definitions (Bondár et al. 2004a)."***

We have thus added references to the following papers:

*Begnaud, M. L., Myers, S. C., Young, B., Hipp, J. R., Dodge, D., & Phillips, W. S. (2021). Updates to the regional seismic travel time (RSTT) model: 1. Tomography. Pure and Applied Geophysics, 178(7), 2475-2498, DOI: 10.1007/s00024-020-02619-5*

*Belinić, T., & Markušić, S. (2017). Empirical criteria for the accuracy of earthquake locations on the Croatian territory. Geofizika, 34(1), 1-17, DOI: 10.15233/gfz.2017.34.5*

*Bennett, T. J., Oancea, V., Barker, B. W., Kung, Y.-L., Bahavar, M., Kohl, B. C., Murphy, J. R., Bondár, I. (2010). The Nuclear Explosion Database (NEDB): A New Database and Web Site for Accessing Nuclear Explosion Source Information and Waveforms.* Seismological Research Letters, *81(1), 12–25, DOI: 10.1785/gssrl.81.1.12*

*Bergman, E. A., Benz, H. M., Yeck, W. L., Karasözen, E., Engdahl, E. R., Ghods, A., Hayes, G. P., Earle, P. S. (2022). A Global Catalog of Calibrated Earthquake Locations. Seismological Research Letters, 94(1), 485–495, DOI: 10.1785/0220220217*

*Bittner, P., Le Bras, R., Mialle, P., & Nielsen, P. (2023). International Data Centre Bulletin Events Triggered by Controlled Underwater Explosions of World War 2 Ordnances. Pure and Applied Geophysics, 180, 1303–1315, DOI: 10.1007/s00024-022-03146-1*

*Boomer, K. B., Brazier, R. A., and Nyblade, A. A. (2010). Empirically Based Ground Truth Criteria for Seismic Events Recorded at Local Distances on Regional Networks with Application to Southern Africa.* Bulletin of the Seismological Society of America, *100 (4), 1785–1791, DOI: 10.1785/0120090237*

*Bondár, I., Myers, S.C., Engdahl E. R., and Bergman, E. A. (2004b). Epicentre accuracy based on seismic network criteria, Geophysical Journal International, 156, 483-496, DOI: 10.1111/j.1365-246X.2004.02070.x*

*Bondár, I., Bergman, E., Engdahl, E.R., Kohl, B., Kung, Y-L., and McLaughlin, K. (2008). A hybrid multiple event location technique to obtain ground truth event locations, Geophysical Journal International, 175, 185-201, DOI: 10.1111/j.1365-246X.2008.03867.x*

*Bondár, I., Godoladze, T., Cowgill, E., Yetirmishli, G., Myers, S. C., Gunia, I., Buzaladze, A., Czecze, B., Onur, T., Gök, R., and Chiang, A. (2024) Relocation of the Seismicity of the Caucasus Region, Bulletin of the Seismological Society of America, 114, 857-872, DOI: 10.1785/0120230155*

*Gomberg, J. S., Shedlock, K. M., and Roecker, S. W. (1990). The effect of S-wave arrival times on the accuracy of hypocenter estimation. Bulletin of the Seismological Society of America, 80 (6A), 1605-1628, DOI: 10.1785/BSSA08006A1605*

***Sweeney, J.J. (1996). Accuracy of teleseismic event locations in the Middle East and North Africa, Lawrence Livermore National Laboratory, UCRLID-125868.***

***Sweeney, J.J. (1998). Criteria for selecting accurate event locations from NEIC and ISC bulletins, Lawrence Livermore National Laboratory, UCRL-JC-130655.***

***Zeng, X., Xie, J. & Ni, S. (2015). Ground Truth Location of Earthquakes by Use of Ambient Seismic Noise From a Sparse Seismic Network: A Case Study in Western Australia. Pure Applied Geophysics, 172, 1397–1407, DOI: 10.1007/s00024-014-0993-6***

***Zhu, C., Wang, C., Zhang, B., Qin, X., and Shan, X. (2021). Differential Interferometric Synthetic Aperture Radar data for more accurate earthquake catalogs. Remote Sensing of Environment, 266, 112690, DOI: 10.1016/j.rse.2021.112690***

Section 1.2 – I suggest adding some description the ISCloc procedure and inversion algorithm in general, plus short, algorithmic description of all specific features important tp the current study. This would give more context for later analyses and results in the paper, and avoid that the reader needs to be familiar with ISCloc or consult the 2011 reference.

Good point – agreed that this would add to the clarity of the argument. We have added the below text:

*Lines 110 – 143*

***"The correlated errors are accounted for by the non-diagonal elements of the covariance matrix as defined by Bondár & Storchak (2011), with the covariance for a given station pair depending on the station separation. The covariance matrix can vary at every iteration of the linear relocation, as phases are redefined or even rejected.***

***The posterior data covariance matrix, calculated for the final converged hypocentre is used to define the error ellipse. The remaining phase residuals are combined with the posterior data covariance matrix to inform the 4D error ellipse (e.g. equation 8 of Bondár & Storchak, 2011). This 4D error ellipse (latitude, longitude, depth and origin time) is then used to define the 2D horizontal error ellipse, described by the semi-major axis, semi-minor axis, and the orientation of the semi-major axis, as well as 1D errors for event depth and origin time. The reported error ellipse is scaled to a 90 % confidence level, through benchmarking with the original GT list (Bondár & Storchak, 2011).***

***The single event GT earthquakes considered in this study are seeded with well recorded events from the ISC Bulletin, that fulfil the criteria discussed and refined in this paper (see Table 1 and 2). The events are then relocated using ISCloc. which employs an iterative linear relocation procedure, where the hypocentre is refined from a given starting point through the linearised reduction of travel time residuals between the observed seismic phases, and those predicted by ak135 (Kennett et al 1995).***

***ISCloc attempts to solve for the event depth where one of the following is true; there is at least one reported station within 0.2°, there are at least five stations reporting P & S phases within 3°, there are at least five reported depth phases or there are at least five core reflection phases. We note that the last two of these are irrelevant for constraining depths for GT events defined using local data. If the linear inversion fails to converge using a resolved depth from the above criteria then the inversion is repeated with a "fixed depth" which is taken from a geographic grid of user defined depths. In the case of GT qualifying events, the free depth criteria within ISCloc will almost certainly be met for all events considered, as the equivalent GT criteria are much stricter. Fixed, or unresolved depths can still occur however when the linear relocation procedure fails to converge. This may occur if the available phases have elevated degrees of error (e.g. pick errors resulting from noisy waveform data), or if the travel times predicted from the 1D velocity model account for the arrival times of the observed phases poorly. If the depth is unresolved and thus set to a fixed depth, the event is rejected as a GT event.***

***The improvements to the location procedures implemented in ISCloc by Bondár & Storchak (2011) provide an opportunity to update the GT criteria."***

The semi-major axis of the error ellipse gives posterior information on the nominal epicenter constraint. A standard error on depth (e.g. from the full error ellipsoid) would give equivalent information on depth constraint. Instead, it seems that information on depth constraint in this study is only binary - whether the event depth is resolved (free) or not resolved (fixed). In addition, this information is not posterior to location, but mixed with the location process. Some mention and discussion of these issues (beyond the last two sentences in the conclusions), and why some kind of posterior information on depth constraint is not used or not available, would be of interest and a valuable addition to the study.

While the posterior depth error is calculated, we believe it is strongly biased by the use of ak135 and at this stage do not want to make definitive statements when there are known limitations to the location procedure. We do not feel able to fully investigate/address this issue without being able to properly benchmark the depths output by ISCloc. This is currently not possible as the GT0 events used in this study to benchmark the locations are all shallow or surface explosions. Other benchmarks are needed, and we feel this is beyond the scope of this paper.

We have added the following to clarify our reasoning for the reader:

*Line 374 – 383*

***"In this study, we continue to require a free or resolved depth for an earthquake to be considered a GT event. At this stage, we do not discriminate GT events based on the size of the depth error, as the error in depth may be as much controlled by the deviation of the unknown local velocity structure from ak135, as by the station distribution geometries that are primarily considered in this study. In addition, we currently have no reliable benchmark data set that is appropriate for testing depth resolution. The assertions made in this work concerning horizontal location errors are based on the well constrained locations of explosions, with***

***sources located very close to the surface, making them ill-suited for testing depth resolution. We therefore consider only if the event depth can be resolved given the available phase data."***

Line 44-46: You could mention that GT events are also very important as reference or anchor for calibrating event absolute locations (epicenter and/or depth). Maybe there are other key uses for GT events that could be listed.

We agree that more key uses of GT events should be listed and we have amended the text as follows:

*Lines 44-49*

***"Seismic events with well constrained locations and origin times, referred to as "Ground Truth" (GT) events, are an important resource allowing seismologists to test and validate new techniques relating to seismic event locations, such as location algorithms and velocity models (e.g. Begnaud et al. 2021). As we demonstrate in this study GT events are of particular use for calibrating the absolute event locations (e.g. Belinić & Markušić, 2017) and additionally GT events can be used as the seed locations for multi-event relocation techniques (e.g. Bergman et al. 2022; Bondár et al., 2024)."***

Line 60: "ΔU" has not yet been defined in the main text – at least a short phrase summarizing this measure is probably needed here, like lines 78-79.

We agree that the reader requires a short explanation of ΔU at this point. We have amended the text as follows:

*Lines 66-67*

***"and a ΔU (network quality metric, describing station distribution as proposed by Bondár and McLaughlin 2009a) of less than"***

Line 110: "A feature of the ΔU criteria…" - at this point, and for this and following paragraphs, I think a full definition and mathematical description of ΔU is needed.

We agree that a better definition of ΔU is required in this section, however, we feel moving the full mathematical description from Section 2 to here would disruption the flow of the introduction. We have remedied this by including a qualitative description of ΔU including a simple example of its operation:

*Lines 150-154*

***"The ΔU criteria is a measure of the deviation of the azimuthal station distribution from a perfect azimuthal distribution. For example, a perfect distribution of five points would be described by the corners of a pentagon. This means that adding a single station can decrease the value of ΔU by altering the perfect distribution that the station azimuths are compared to, thus causing a GT candidate event to fail while all other criteria are improved."***

Line 113: “(e.g. Figure 1, bottom)” – I think there is not yet enough information on ΔU, uniform azimuths and CPQ for the reader to make much sense of the meaning of Figure 1.

We link this to the previous point and think that the qualitative description of ΔU which we have added, allows the reader to be more informed when looking at Figure 1.

Line 147: “In contrast, the values of CPQ are very similar for the two events and requires no stations to be removed.” – It seems here also that a full definition and mathematical description of CPQ is needed before discussing results based on CPQ.

In this line we only intend for the reader to note the large change in ΔU relative to the consistent value of CPQ. We do not feel that it is necessary to move the full description of CPQ to this section to make this clear.

Line 179: It would be interesting to have a very concise explanation of the terms in this equation and the mechanics of the calculation. Not at all evident from a quick glance.

We agree that a description of this equation is required to assist the reader. We have amended the equation and text as follows:

$$CPQ = \frac{\left|\sum_{i=1}^{n-1} x_i\, y_{i+1} + x_n y_1 - \sum_{i=1}^{n-1} x_{i+1}\, y_i - x_1 y_n\right|}{2\pi}$$

***Lines 224-233***

***“where x and y are the Cartesian coordinates of the event to station azimuths (esazi) ordered from 0 to 360 degrees, the subscript refers to the number of the vertex in clockwise order. The Cartesian coordinates are found using x = cos[esazi] & y = sin[esazi]) with a radius of 1. Once the vertices of the cyclic polygon are in Cartesian coordinates, we can apply the “Shoelace Formula” where in a clockwise (or counter-clockwise) direction we calculate the sum of the product of the x coordinate value with the y coordinate value of the next vertex and the subtraction of the product of the y coordinate value with the x coordinate value of the next vertex. We can subtract these two sums and divide by two to get the area of the cyclic polygon. By dividing this value by π, we get the ratio of the area of the cyclic polygon to the area of the unitary circle.”***

Figure 4: You might explain why some of the linear structures in the CPQ plot show increased mislocation with increasing CPQ. This seems an odd and potentially interesting and relevant relation.

We agree that the linear structures in the bottom plot of Figure 4 are of interest and offer the following explanation as to a potential source of these.

If there are multiple stations with poor quality picks then adding them an event may make the mislocation worse even as the value of CPQ improves. For most events where there are many possible stations then the random selection of stations are unlikely to include a “bad” station, hence the mislocation generally improves with more stations (i.e. better CPQ). If there are events

with fewer stations then the "bad" stations will have a larger effect on the mislocation of the event as well as being randomly selected more often. Further due to the use of a global 1-D velocity model (ak135) in ISCLoc it is possible that the best seismically derived location will still have a significant absolute mislocation.

We have added the following text to outline this explanation in the paper:

*Lines 276-282*

***"The majority of these linear structures demonstrate that increased CPQ values are correlated with decreasing mislocation, consistent with the general trend observed. In a minority of linear groupings, the opposite is observed, with increasing CPQ correlated with greater mislocation. This can be explained by an increasing proportion of stations with poor quality picks contributing to the constrained hypocentre, or by the greater chance of the location procedure being influenced by unmodelled local velocity perturbations."***

Suggested minor issues to address:

Line 47-51: A bit choppy, I suggest modifying to something like:
"GT events are typically obtained in one of three ways: 1) explosions or other anthropogenic events, where the location and timing of the event is known or well documented by nearby observations; 2) seismic events that are well located by a dense and well distributed near-field seismic network; 3) seismic events that are well located by a multi-event location technique."

We have amended the text to mirror this structure and added additional references. We have not included numbers as we have added additional GT methodologies. The text is amended as follows:

*Lines 51-57*

***"GT events can be defined in many ways: by explosions or other anthropogenic events, where the location and timing of the event is known or well documented by nearby observations (e.g. Bennett et al. 2010; Bittner et al. 2023); seismic events that are well located by a dense and well distributed near-field seismic network (e.g. Bondár and McLaughlin, 2009a; Boomer et al 2010); seismic events well located by a multi-event location technique (e.g. Bondár et al. 2008; Bergman et al. 2022). Additionally, there are less established methods of obtaining GT events including using ambient seismic noise (e.g. Zeng et al. 2015) and InSAR (e.g. Zhu et al. 2021)."***

Line 133: I suggest more information on "uniform event to station azimuth" is needed. Is it the corresponding uniform azimuth in a sorted list of N uniform azimuths? Maybe add the definition from Bondár and McLaughlin (2009a): unifi = 360i / N for i = 0,..., N – 1

We agree that adding the definition from Bondár and McLaughlin (2009a) is required for clarity:

*Line 177*

*“event to station azimuth (unifi = 360i / N for i = 0,..., N – 1)”*

Figure 3 caption: “CPQ is given by the ratio enclosed by the red polygon, and the area enclosed by the black circle.” → “CPQ is given by the ratio of the area enclosed by the red polygon to the area enclosed by the black circle.”

Changed as suggested.

Figure 6 caption: Description of Top and Bottom should use the same ordering, e.g. “Semi-major axis of error ellipse with respect to CPQ (left) and ΔU (right)”

Changed as suggested.

Line 291: “ΔU” → “secondary gap”

Changed as suggested.

Figure 7 caption: All panels might best be described as “Semi-major axis of the error ellipse with respect to XXX”

Changed as suggested.

Line 336: “bottom” → “bottom left”

Changed as suggested.

Line 343: “For these events there are no stations reporting P & S phases and no stations within 10 km only ...” → “For these events, when there are no stations reporting P & S phases only …”

Changed as suggested.

Line 370: “this” → “the above criteria”

Changed as suggested.

Line 400: “algorithms” → “distributions”

Changed as suggested.

Line 411: “resolving for the event depth” → “that event depth is resolved (not fixed)”

Changed as suggested.

---

Reviewer B:

The manuscript is generally clear, well-written, and accessible. It addresses a topic of significant interest to the Seismica community by presenting a catalog of precise, high-quality global earthquake locations. However, there are several issues and suggestions the authors should address before the manuscript is accepted. The most significant concern is that the use of the semi-major axis of the uncertainty ellipse and the role of P and S phase stations in constraining depth are not novel contributions to the field. To maximize the manuscript's impact, given the new proposed criterion and the convenience of accessing global dataset, I would recommend the authors switch the focus more on the new GT event catalog, illustrate more about the difference between new and old catalog, provide access to the catalog and code example of calculating the CPQ.

We agree that access should be provided to the catalogue and as such we have made the 2018-2020 test dataset available through the ISC website and have included the link to this in the Data and code availability section. We have also made available a python code for calculating the value of CPQ from a list of event to station azimuths and place a link to this in the Data and code availability section.

We focus on the development of the new criteria CPQ and it's improvement over ΔU. This requires us to evaluate the parameters of semi-major axis of the error ellipse and the role of P and S phase stations in relation to CPQ and ΔU. Further, we have endeavoured to be clear that the while the semi-major axis of the error ellipse and the role of P and S phase stations in constraining depth are not new or novel contributions in the field of earthquake location, they are being applied to Ground Truth criteria in a new way. To my knowledge stations reporting both P & S phases have not been used as an exclusionary criteria for any previous set of GT criteria. We also demonstrate that the semi-major axis of the error ellipse is sufficient on its own (when accounting for uneven stations distributions) to be an exclusionary criteria for GT events. To my knowledge this is the first time that this has been demonstrated.

We are hesitant to apply this set of criteria to the whole GT list and to detail that here as we first want the new criteria to be published for the community review, ahead of proposing the new criteria to the IASPEI / CoSOI Working Group on Reference Events for Improved Locations. After this work and future innovations of the GT methodology (adding pics, manual review of events, providing greater constrain on depth etc.) has been completed we intend to publish a complete description of the new updated GT list. In the intervening period we will make a provisional version of the GT list available through the ISC website. This provisional version will be sign-posted on the same page that the links in the Data and code availability will point to.

Suggestions:

The introduction of the \Delta U and its limitation is well-written; but the structure of the *Introduction* and *Quantifying Unbalanced Station Distributions* sections could be improved.

Introducing the equation for ΔU early in the *Introduction* would provide better context and help readers understand why ΔU becomes less effective as the number of stations increases.

While we agree that moving the definition of ΔU earlier in the text would help clarify some of the concepts relating to ΔU, however, we feel that this would break the flow of the argument laid out in the introduction. To address this we have included a qualitative description of ΔU including a simple example of its operation:

*Lines 150-154*

***"The ΔU criteria is a measure of the deviation of the azimuthal station distribution from a perfect azimuthal distribution. For example, a perfect distribution of five points would be described by the corners of a pentagon. This means that adding a single station can decrease the value of ΔU by altering the perfect distribution that the station azimuths are compared to, thus causing a GT candidate event to fail while all other criteria are improved."***

A brief explanation of how ISCloc locates events and calculates uncertainty ellipses is needed. Since the semi-major axis of the uncertainty ellipse is proposed as a quality metric, it is important to clarify the confidence level of this metric, e.g. 95%, 99%. The authors mention "90% semi-major axis" in Figure 5, but it is unclear whether this refers to 90% of the axis length or a 90% confidence interval. Yu et al. (2024) highlighted that most earthquake location programs do not output 95% confidence ellipses, and various statistical assumptions are used to derive them. Clarifying this for ISCloc would strengthen the manuscript.

Good point, we have added a section detailing the relevant features of ISCloc, including how the error ellipses are derived below:

*Lines 110-141*

***"The correlated errors are accounted for by the non-diagonal elements of the covariance matrix as defined by Bondár & Storchak (2011), with the covariance for a given station pair depending on the station separation. The covariance matrix can vary at every iteration of the linear relocation, as phases are redefined or even rejected.***

***The posterior data covariance matrix, calculated for the final converged hypocentre is used to define the error ellipse. The remaining phase residuals are combined with the posterior data covariance matrix to inform the 4D error ellipse (e.g. equation 8 of Bondár & Storchak, 2011). This 4D error ellipse (latitude, longitude, depth and origin time) is then used to define the 2D horizontal error ellipse, described by the semi-major axis, semi-minor axis, and the orientation of the semi-major axis, as well as 1D errors for event depth and origin time. The reported error ellipse is scaled to a 90 % confidence level, through benchmarking with the original GT list (Bondár & Storchak, 2011).***

***The single event GT earthquakes considered in this study are seeded with well recorded events from the ISC Bulletin, that fulfil the criteria discussed and refined in this paper (see Table 1***

***and 2). The events are then relocated using ISCloc. which employs an iterative linear relocation procedure, where the hypocentre is refined from a given starting point through the linearised reduction of travel time residuals between the observed seismic phases, and those predicted by ak135 (Kennett et al 1995). ISCloc attempts to solve for the event depth where one of the following is true; there is at least one reported station within 0.2°, there are at least five stations reporting P & S phases within 3°, there are at least five reported depth phases or there are at least five core reflection phases. We note that the last two of these are irrelevant for constraining depths for GT events defined using local data. If the linear inversion fails to converge using a resolved depth from the above criteria then the inversion is repeated with a "fixed depth" which is taken from a geographic grid of user defined depths.***

***In the case of GT qualifying events, the free depth criteria within ISCloc will almost certainly be met for all events considered, as the equivalent GT criteria are much stricter. Fixed, or unresolved depths can still occur however when the linear relocation procedure fails to converge. This may occur if the available phases have elevated degrees of error (e.g. pick errors resulting from noisy waveform data), or if the travel times predicted from the 1D velocity model account for the arrival times of the observed phases poorly. If the depth is unresolved and thus set to a fixed depth, the event is rejected as a GT event."***

If the ellipse is of 95% confidence interval, it is unsurprising that the 93.2% of the known mislocated explosions falls under the curve in Figure 5., as this aligns with how confidence interval function. This finding instead supports the validity of ISCloc's uncertainty evaluation and its use as a metric for location quality. The authors should rephrase this discussion to avoid overstating the result.

We have confirmed in the text that the error ellipse reported is the 90 % error ellipse.

*Lines 120-122*

***"The reported error ellipse is scaled to a 90 % confidence level, through benchmarking with the original GT list (Bondár & Storchak, 2011)."***

*Lines 307-309*

***"We find that for ~87% of the explosion data set the known location is within the calculated (90% confidence interval) error ellipse"***

We fully agree with the reviewer's statement that "This finding …. supports the validity of ISCloc's uncertainty evaluation and its use as a metric for location quality". We report this finding in the original text in the following way (now line 314-316 of revised manuscript):

*"Based on this result we propose that the semi-major axis of the error ellipse is the best control on whether an event has a mislocation of 5 km or less and that other criteria are secondary."*

In addition, we have now clarified that this is a re-assessment of the error ellipses in ISCloc, but rather than assessing this with the full set of global observations, we use just the local phases

used in GT relocation. We therefore feel this test is needed to justify the use of the semi-major axis as our measure of location accuracy in the rest of the paper.

*Lines 309-311*

***This corroborates the results from Bondár and Storchak (2011) who defined the 90% confidence interval when using ISCloc with global data, however, in our case we have replicated this result using only local data (e.g. seismic phases reported less than 150 km from the event).***

Gomberg et al., 1990 has demonstrated the effectiveness of S-wave arrival times on the depth determination from both observation and numerical experiment. This work aligns with the authors' argument that the stations with P&S phases help constrain the depth. The authors should refer to this article in that part at least. Indeed, the paper would benefit a lot from a comprehensive engagement with existing literatures. More references in Introduction, Section 3-5, would better contextualize this paper with the existing body of works.

We have added a reference to the Gomberg et al. (1990) paper and credited them with showing an improvement in the trade-off between event depth and origin time where both a P and S phase are reported at a local station:

*Lines 388-393*

***"Another way of addressing the trade-off between event depth and origin time is by requiring S and P phases recorded at a local station. This has been shown to significantly reduce the trade-off between depth and origin time (e.g. Gomberg et al. 1990)."***

We agree that the reference list required expanding to provide the reader with greater context. We have added references to the following papers in the text:

***Begnaud, M. L., Myers, S. C., Young, B., Hipp, J. R., Dodge, D., & Phillips, W. S. (2021). Updates to the regional seismic travel time (RSTT) model: 1. Tomography. Pure and Applied Geophysics, 178(7), 2475-2498, DOI: 10.1007/s00024-020-02619-5***

***Belinić, T., & Markušić, S. (2017). Empirical criteria for the accuracy of earthquake locations on the Croatian territory. Geofizika, 34(1), 1-17, DOI: 10.15233/gfz.2017.34.5***

***Bennett, T. J., Oancea, V., Barker, B. W., Kung, Y.-L., Bahavar, M., Kohl, B. C., Murphy, J. R., Bondár, I. (2010). The Nuclear Explosion Database (NEDB): A New Database and Web Site for Accessing Nuclear Explosion Source Information and Waveforms.*** **Seismological Research Letters,** ***81(1), 12–25, DOI: 10.1785/gssrl.81.1.12***

*Bergman, E. A., Benz, H. M., Yeck, W. L., Karasözen, E., Engdahl, E. R., Ghods, A., Hayes, G. P., Earle, P. S. (2022). A Global Catalog of Calibrated Earthquake Locations. Seismological Research Letters, 94(1), 485–495, DOI: 10.1785/0220220217*

*Bittner, P., Le Bras, R., Mialle, P., & Nielsen, P. (2023). International Data Centre Bulletin Events Triggered by Controlled Underwater Explosions of World War 2 Ordnances. Pure and Applied Geophysics, 180, 1303–1315, DOI: 10.1007/s00024-022-03146-1*

*Boomer, K. B., Brazier, R. A., and Nyblade, A. A. (2010). Empirically Based Ground Truth Criteria for Seismic Events Recorded at Local Distances on Regional Networks with Application to Southern Africa.* **Bulletin of the Seismological Society of America,** *100 (4), 1785–1791, DOI: 10.1785/0120090237*

*Bondár, I., Myers, S.C., Engdahl E. R., and Bergman, E. A. (2004b). Epicentre accuracy based on seismic network criteria, Geophysical Journal International, 156, 483-496, DOI: 10.1111/j.1365-246X.2004.02070.x*

*Bondár, I., Bergman, E., Engdahl, E.R., Kohl, B., Kung, Y-L., and McLaughlin, K. (2008). A hybrid multiple event location technique to obtain ground truth event locations, Geophysical Journal International, 175, 185-201, DOI: 10.1111/j.1365-246X.2008.03867.x*

*Bondár, I., Godoladze, T., Cowgill, E., Yetirmishli, G., Myers, S. C., Gunia, I., Buzaladze, A., Czecze, B., Onur, T., Gök, R., and Chiang, A. (2024) Relocation of the Seismicity of the Caucasus Region, Bulletin of the Seismological Society of America, 114, 857-872, DOI: 10.1785/0120230155*

*Gomberg, J. S., Shedlock, K. M., and Roecker, S. W. (1990). The effect of S-wave arrival times on the accuracy of hypocenter estimation. Bulletin of the Seismological Society of America, 80 (6A), 1605-1628, DOI: 10.1785/BSSA08006A1605*

*Sweeney, J.J. (1996). Accuracy of teleseismic event locations in the Middle East and North Africa, Lawrence Livermore National Laboratory, UCRLID-125868.*

*Sweeney, J.J. (1998). Criteria for selecting accurate event locations from NEIC and ISC bulletins, Lawrence Livermore National Laboratory, UCRL-JC-130655.*

*Zeng, X., Xie, J. & Ni, S. (2015). Ground Truth Location of Earthquakes by Use of Ambient Seismic Noise From a Sparse Seismic Network: A Case Study in Western Australia. Pure Applied Geophysics, 172, 1397–1407, DOI: 10.1007/s00024-014-0993-6*

***Zhu, C., Wang, C., Zhang, B., Qin, X., and Shan, X. (2021). Differential Interferometric Synthetic Aperture Radar data for more accurate earthquake catalogs. Remote Sensing of Environment, 266, 112690, DOI: 10.1016/j.rse.2021.112690***

Per Seismica's data policy, the new GT catalog produced using the CPQ criterion should be made public available. It would be better if you can share a demo code that generate the catalog using ISC's data.

We agree that both the new GT list of 2018-2020 detailed in this paper and the Provisional GT list developed with these criteria should be made public. Links are now on the ISC website and have been added to the Data and code availability section.

It is currently not possible to make the full workflow available as the results are achieved using multiple different codes linked to the internal ISC database. However the results could be repeated using the open access data via the ISC website, and the open access ISCloc code (again available on the ISC Website, links to which are in the Data and code availability section of the revised manuscript).

We also provide access to a python code for calculating the value of CPQ from a list of event to station azimuths and placed a link to this in the Data and code availability section.

The figures in the manuscript are low resolution. If this is due to preprint formatting, it should be clarified. If not, the authors need to produce high-resolution versions of all figures for publication.

We agree that this is not ideal and related to embedding figures in the .docx document format. We will provide separate high resolution jpeg files alongside our corrections.

Minor issues:

Some acronyms (e.g., NEIC, ISC, GNS) are missing their full names.

Acronyms have now been given their full names.

Line 133: typo - easz_i should be esaz_i.

Changed as suggested.

Line 179-180: typo – "easz" is misused.

Changed as suggested.

**Round 2**

**Reviewer A**

The revised version of the paper is much improved and addresses well the reviewers' comments. In particular, I find that the authors' added description the ISCloc procedure and inversion algorithm, explanation of why a very simple depth criteria is used to define GT events, and their solutions to the tricky issues of where and how to introducing explanation of the ΔU criteria, are all clear, informative and beneficial changes to the manuscript.

I only note only one minor issues.

Suggested minor issues to address:

1. Line 304: "top panel of Figure 7" → "top panel of Figure 6"

Anthony Lomax

**Reviewer B**

I have carefully reviewed the revised version of the manuscript and found that the authors have addressed my comments. I have no further comments or questions, and I recommend accepting the manuscript.

Yifan Yu