

## Reviewer #1

This informative paper is well constructed and explains well the issues associated with the use of time observations from the reading of analogue seismogram at a limited number of stations.

The approach proposed for introducing a time variable error for events before 1964 that reduce with later dates is well founded, and an improvement on current assumptions. The choice of working with fixed depth in the early years is reasonable even though a few events are such that reasonable depth estimates can be made.

We thank the reviewer for the comments and suggestions. Below are our answers to each point.

The presentation is generally clear though there are few unfortunate typos.

However, the very heavy fonts used for labelling in a number of figures are very jarring.

The balance achieved in Figure 8 is preferable to that in Figure 3, where the inbox labels are dwarfed by the external labelling.

We updated most Figures to provide more consistent fonts and labelling

Specific points:

"estimations" should be replaced by "estimates" throughout → Done

The Abstract has several clumsy wordings - a modified version is suggested below:

"Uncertainty in reported body-wave arrival times is a key contributor to earthquakes location error estimates, especially in the early-instrumental period (e.g., prior to the early 1960s). As such, a reliable assessment of the observational errors in the early-instrumental period is an important element of the earthquake location problem. The standard location procedures at the International Seismological Centre assume seismic arrival time picking errors as defined for the most recent decades of instrumental seismology (i.e., from the early 1960s). However, the error estimates currently used fail to capture the uncertainty in the seismic arrival time picks for earthquakes that occurred before the early 1960s (early-instrumental period). The larger observational uncertainty in earlier years arise from a range of error sources when reading arrival times on analogue seismographs. Such errors have drastically reduced since the 1960s thanks to significant improvements in seismometry and time keeping, as well as the migration from analogue to digital stations worldwide. Since observational errors play a key role in the uncertainty estimates for earthquake locations, it follows that error ellipses for early-instrumental earthquakes are underestimated in our current procedures. To address this issue, we modify the error assumptions used in the early-instrumental period by using a time dependent term that produces more reliable error ellipses for early-instrumental earthquakes.

We thank the reviewer for the suggestions, we have updated the Abstract accordingly.

Page 3 line 59: "time" should be "times" → Done

Footnote 3: "intend as" would be better as "refer to" → Done

Page 4 line 66: "underestimation" -> "underestimate" → Done

line 68: insert "assumptions" before "fail" → Done

Page 5 line 105: Some mention of the way in which the starting point for the linearised inversion is selected is desirable here.

The linearisation is only effective when close to the global minimum of fit.

We have added two sentences to describe the starting point selection (see page 5, lines 109-111 of the clean new version).

Page 6 line 122: the end of the sentence is missing, "and" should not be in italics.

We thank the reviewer for spotting this text error, sentence corrected.

Page 9 Figure 5:

Some comment on why there is a broad spread of residuals around 1990 when digital records are well established would be helpful!

We have added two sentences to this regard along with an updated Figure 5, as requested by Reviewer #2 (see text at lines 225-238 of the clean new version).

Page 10 line 240: " $\sigma_{thres}$ " is this supposed to be " $\sigma_{thres}$ "?

Note the quantity is never defined, some further explanation needed.

In the original submission *sigmathres* is defined at page 6, line 132.

Page 11 line 274:

"in order to detect" would be better as "which would allow the detection of" → Done

Page 12 line 292: "differ to" -> "differ from" → Done

Page 14 Figure 7:

The symbols and ellipses very hard to see for all but the smallest errors.

The whole scale needs to be darker (less saturated) to make this figure work.

We have updated the Figure to make it clearer.

Page 17 line 393: "dating before" -> "dating from before" → Done

Reviewer #2

In this paper the authors address the relatively large arrival time pick uncertainties and errors in early instrumental (e.g. pre 1960's) earthquake location and how to better include these pick uncertainties in ISC locations. They describe sources of large pick error in early observations and the corresponding location errors as determined in different studies. They show that current ISC-GEM procedure, using pick uncertainties derived from recent, post 1960's monitoring of ground-truth events, give an underestimation of location error estimates for early instrumental events relative to most other studies. The authors attribute this underestimation to the need for larger, more realistic pick uncertainties for early events, which they then derive through a statistical, grid-search analysis of location residuals for events in the period 1904-1963. They show that relocations using the derived, new pick uncertainties give much larger location uncertainties than current ISC-GEM

locations for example events in 1907 and 1912, and that the new location uncertainties agree well with those of other studies.

This paper reads well and is interesting and informative with regards to sources of early instrumental pick errors, the effects of such errors on locations, and to current and proposed ISC procedures for relocation of early-instrumental events.

We thank the reviewer for the feedback. Below are our answers to each point.

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

#### Comments:

Line 111: One or two sentences describing the variogram model would avoid the reader having to consult Bondár and Storchak (2011). In particular, to aid in understanding of Eq. 1, mention that the variogram level increases with station distance.

We have added two sentences in the revised version (see page 5, lines 111-119) to better explain the effects of the variogram.

Line 236: “A further refinement appears to occur after 2004” – This is not apparent, there is an increase in sigma around 2004, but overall there seems mainly a slow decrease from the start of GPS timing until ~2010. The sigma is remarkably stable between 1964 and the start of GPS.

We removed the sentence.

Line 246: “disruptions ... related to global conflicts” – these fluctuations as presented as residuals in Fig 5 often have very distinct signatures, e.g. many positive and large positive residuals and very few small-negative residuals during ~1939-1944. More description and explanation, and any references on these features might be of interest to most readers.

We added one sentence to this regard (lines 260-262 of the new clean version), but we are unable to provide a better explanation or references (most of the data analysed here was never used before). However, we point out that we intend to perform a more systematic review of the data by ISC analysts to address some of those signatures.

Line 367: “the distribution of the travel times residuals relative to the best fitting sample of the grid search, i.e., the point where the standard deviation is the lowest” – I understand this histogram as showing time differences between the residual for each P pick for each grid search location and the residual for that same pick at the best fitting location sample. So, effectively, the pick residual at the best fitting location is taken as “correct”, thus a sort of time correction due to whatever cause (e.g. clock error), and then the distribution of differential residuals for the pick around the “correct” residual are taken as an estimate of observation uncertainty for this pick. This pick observation uncertainty is thus defined by the change in residuals for the pick over many samples of lower misfit locations, and thus controlled in part by the full set of observations and their (unknown) errors.

If my understanding is correct, this procedure seems a useful, if perhaps ad-hoc and not statistically formal, way of quantifying pick uncertainty from noisy observations only. In any case, I would suggest a bit more description and explanation of the procedure to confirm,

correct and/or extend my understanding of the procedure. If there is a formal statistical methodology or precedence for the procedure, it would be advantageous and informative to cite this formal basis.

We agree with the reviewer that this method is *ad hoc*, in that it is designed specifically for this purpose and has no strong relation to other methodologies previously applied to this problem. We have intentionally stepped back from a more formal statistical approach as we felt this approach '*drove the data*' too hard. This is an empirical and data driven measure of pick uncertainty – while we accept this is not fully statistically formal, it is still a major step on from previous approached in the literature. The only misunderstanding in the reviewer's comments is that the histogram is not representing the residuals for all low misfit samples, just the one lowest standard deviation sample. We have added four sentences to this section to further clarify (see page 16, lines 393-399 of the clean new version).

Suggested minor issues to address:

Figure 2 caption: "as plotted at the same origin time" → "as plotted at the same origin time and with the same symbol type" → Done

Line 87: "a defined travel time in ak135 model" → "a defined travel time in the ak135 model" → Done

Line 115: "where the variogram levels off" → "where the variogram levels off at large station separation" → Done

Line 151: "and 30 cm per minute" → "and 3.0 cm per minute" ???

This could be a typo in the original publication, we point this out in the updated footnote 5.

Line 213 "observations at regional distances are notoriously more affected by Earth's heterogeneities" – maybe a reference or two here?

We have added the reference Myers et al., 2010.

Line 220: "where it is possible to see the occurrence of many of the errors listed in the previous section (e.g., miscounts as minute mark errors, misidentifications, typos)" – 1 min errors and other possible small errors (10 or 15 sec?) are not visible. I would strongly suggest a second panel in this figure with a zoom along the Y axis (P residual) from maybe -700 to 700 s and histogram in bins of 5-10 s, to show +/- 10 min, +/- 1 min and other such errors clearly. This would be a very valuable resource by itself!

We have added a second panel to the figure showing the residuals within 12 minutes (+/- 720 seconds).

Line 245: "we" → "were" → Done

Line 273: "not many stations operated continuously in order to detect clock drifts" – not clear the relation between operating continuously and detecting drifts

We have updated the sentence as: "However, not many stations operated continuously which would allow the detection of clock drifts and this limits significantly the possibility of

applying time-dependent station corrections when locations are based on the global network."

Line 302, line 349: "grid search" → "grid search over potential locations" or similar. → Done

Line 411: "agreement " → "agreement of our sigma\_t values" → Done

Anthony Lomax