doi:10.26443/seismica.v4i1.1499

SEISMICA

On the location uncertainty of early-instrumental earthquakes

Domenico Di Giacomo * [1], **Tom Garth** [1], **István Bondár** [2], **Natalia Poiata** [1,3], **Ryan Gallacher** [1], **James Harris** [1], **Dmitry A. Storchak** [1]

[1]International Seismological Centre, Pipers Lane, RG19 4NS, Thatcham, UK, [2]Seismic Location Services, Lagos, Portugal, [3]Now at National Institute for Earth Physics, Măgurele, Romania

Author contributions: *Conceptualization*: Domenico Di Giacomo, Tom Garth, István Bondár, Natalia Poiata, Ryan Gallacher, James Harris, Dmitry A. Storchak. *Formal Analysis*: Domenico Di Giacomo, Tom Garth, Natalia Poiata, James Harris. *Writing - Original draft*: Domenico Di Giacomo, Tom Garth, István Bondár, Dmitry A. Storchak, Natalia Poiata. *Funding acquisition*: Dmitry A. Storchak.

Abstract Uncertainty in reported body-wave arrival times is a key contributor to earthquake 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 arises from a range of error sources when reading arrival times on analogue seismographs. Such errors have drastically reduced since the 1960s thanks to the 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.

Production Editor: Andrea Llenos
Handling Editor: Giuseppe Petrillo
Copy & Layout Editor: Kirsty Bayliss

Received: October 29, 2024
Accepted: February 5, 2025
Published: February 25, 2025

1 Introduction

From early-instrumental seismological observations at the turn of the 20$^{\text{th}}$ century (e.g., Wiechert, 1903) to modern digital seismic networks (Ammon et al., 2010), one of the fundamental tasks performed by seismologists is to interpret seismic recordings and provide arrival times of various phases in order to locate seismic events. Since the least-square procedure introduced by Geiger (1910), location algorithms are typically based on: a) an Earth model (usually 1-D) or travel time tables; b) a broad idea of the likely location (or starting point); c) arrival times of seismic phases $t_i = T(x_i, y_i, z_i, x_0, y_0, z_0) + t_0$, where T is the seismic phase travel time, x_i, y_i, z_i are the station coordinates of the i_{th} station, x_0, y_0, z_0 and t_0 are the hypocentre coordinates and event origin time, respectively.

Numerical approaches for earthquake location are notoriously affected by several sources of error. Two important components of uncertainty stem from deviations of the real Earth's structure from the adopted model and errors in the observations (seismic phase arrival times, t_i). These two components of location error are difficult to separate (Billings et al., 1994). Nevertheless, modern location algorithms have procedures in place to minimise biases due to Earth's heterogeneities and take into account uncertainties in t_i (more details in the next section). Consequently, measurements and model errors are intrinsic features of the earthquake location problem and are linked to the resulting location uncertainties (e.g., Flinn, 1965; Evernden, 1969; Buland, 1976; Jordan and Sverdrup, 1981).

Considering the significant changes in seismology and seismometry since the beginning of global observations at the turn of the 20$^{\text{th}}$ century (e.g., Kanamori, 1988; Agnew, 2002; Wielandt, 2002), improvements in time keeping (e.g., Storchak et al., 2015; Agnew, 2020, and references therein) and the ever growing worldwide seismic network and associated data (Bondár and Storchak, 2011; Bondár et al., 2015), it is natural to expect a decrease over time in global location errors. To corroborate this, one should use a dataset that includes consistently calculated instrumental location uncertainties (i.e., using the same Earth model and algorithm) over a long period of time. To our knowledge, the only dataset that fulfils such requirements is the ISC-GEM Catalogue (International Seismological Centre, 2023b). The advantage of this catalogue is that it uses modern location techniques (Bondár et al., 2015) and provides uncertainties for the entire period of global instrumental seismology (i.e., since 1904). The location error estimates in the ISC-GEM Catalogue are ultimately computed by the ISC location algorithm (Bondár and Stor-

*Corresponding author: domenico@isc.ac.uk

chak, 2011), hereafter referred to as ISCloc. Fig. 1 shows the annual error ellipse semi-major axis ($Smajax$) distribution[1] (interquartile range) as extracted from the catalogue version released in March 2023 (see Data availability).

Despite the presence of minor fluctuations, a general decrease over time of the location errors is observed, with the largest location uncertainties occurring in the early years of the 20th century. The annual median value of $Smajax$ is just above 30 km before 1910, between 20 and 30 km up to the late 1920s, then between 10 and 20 km up to 1952 and below 10 km afterwards. The small error ellipses in the more recent instrumental period[2] are a result of the nature of the ISC-GEM Catalogue, which includes only large earthquakes (down to magnitude 5.0-5.5 since 1964). This means that, in general, earthquakes in the more recent period are well-recorded worldwide with good station azimuthal coverage.

Although $Smajax$ in the first two decades of the 20th century is about 4 to 7 times larger than in the 21st century (Fig. 1), we find that ISC-GEM error ellipses for early-instrumental[3] earthquakes can differ significantly from estimates by other authors, as outlined in Fig. 2.

Even if comparing results from different location techniques and their associated error estimates is not straightforward, it appears that early-instrumental earthquake location error estimates from the ISC-GEM Catalogue are consistently smaller than those from other studies considering the same period. In the recent period we observe a much better fit between literature results and the ISC-GEM Catalogue, although the number of earthquakes depicted in Fig. 2 is relatively small. We show that the underestimate of the error ellipses in the ISC-GEM Catalogue for early-instrumental earthquakes is due to the use of prior error assumptions originally defined for the more recent period of instrumental seismology (Bondár and Storchak, 2011). These assumptions fail to capture the real uncertainty of the reported picks in the first part of the 20th century. To address this issue, we derive a time dependent uncertainty term from the analysis of travel time residuals calculated for relatively well recorded earthquakes in the ISC-GEM Catalogue. We show that the use of such uncertainty term allows us to obtain more realistic error ellipses for early-instrumental earthquakes.

2 Observational errors

The challenges of locating early-instrumental earthquakes are well documented (Adams, 2004). They originate mainly from a combination of sparse and uneven station distribution (Di Giacomo et al., 2018) and uncertainties associated with the readings of the seismic arrival times. With regard to the latter, it is fitting for the purpose of this work to quote Kanamori et al. (2010)[4] regarding the location of the 4 January 1907 Sumatra earthquake: *"We conclude that the available arrival time data in 1907 are so poor (probable picking error of about 10 s for Milne seismograms or about 3 s for Omori or Wiechert seismograms; in addition, there were large clock errors, exceeding tens of seconds at some station) that epicentre location has an error of at least 1°"*. It follows that a reliable assessment of the uncertainty associated with seismic arrival times is necessary. Before discussing our approach to account for the magnitude of early-instrumental observational errors, we first recall the setup in our procedures, provide a brief overview of the observational error assessments in the literature (both for the early-instrumental and recent period), and discuss features of travel time residuals in ISC-GEM locations.

2.1 Summary of the ISC location procedures

Here we give a brief overview of key features of the ISC location procedures as described more fully in Bondár and Storchak (2011). ISCloc uses all phases with a defined travel time in the ak135 model (Kennett et al., 1995) and performs an iterative linearised inversion using an *a priori* estimate of the full data covariance matrix to account for correlated model errors (Bondár and McLaughlin, 2009a). Bondár and Storchak (2011) inspected the travel time residuals computed with respect to ground truth (GT) locations (Bondár and McLaughlin, 2009b), also known as IASPEI Reference Event List (International Seismological Centre, 2023c, hereafter referred to as GT List), to set *a priori* picking errors for specific phases in different distance ranges. The *a priori* picking errors range from a minimum of 0.8 s set for the first arriving P-type phases (Pg, Pb, Pn, P), generally picked more accurately than later phases, to 2.8 s for phases diffracted along the core-mantle boundary (e.g., $Pdif$, $Sdif$, $PKPdif$). In-between there is a multitude of seismic phases with *a priori* errors of 1.2 s for P at distances between 15° and 28° and Pn at distances between 15° and 20°, 1.3 s for core phases and a number of depth phases (e.g., pP, sP, pwP), 1.5 s for the first arriving S-type phases (Sg, Sb, Sn, S) and 1.8 s and 2.5 s for free surface reflections, conversions and less reliable depth phases. Fig. 3 shows the *a priori* errors for commonly picked phases (the full table is available in the Supplementary Material). The relative differences in the *a priori* errors described above means that phases that are expected to have greater uncertainty in the picked onset are down-weighted in the location algorithm. Assuming that seismic arrival times are picked reliably and ignoring dependency on signal-to-noise ratio (magnitude), distance and variations between seismic networks (or data reporters) or even from station to station, the *a priori* error measurements are a statistical quantity representing observational arrival time error estimates of the global seismic network (here in-

[1] The same trends observed in Fig. 1 are seen if the area of the error ellipses is considered instead of $Smajax$. Nevertheless, $Smajax$ is a good proxy for the size of the error ellipse as ISCloc, by taking into account correlated errors, makes the error ellipses more circular. It reduces, therefore, the number of long, elongated error ellipses.

[2] For sake of simplicity, we distinguish between early-instrumental and recent period of instrumental seismology if an earthquake occurred before and since 1964, respectively, where 1964 is the first year of the reviewed ISC Bulletin (International Seismological Centre, 2023a).

[3] For early-instrumental we here refer to the period where earthquakes were recorded by analogue instruments such as the Wiechert, Galitzin, Mainka, Bosh-Omori and other instruments operative in the first half of the 20th century (see, e.g., Kanamori, 1988).

[4] Quote from the Appendix, page 371.

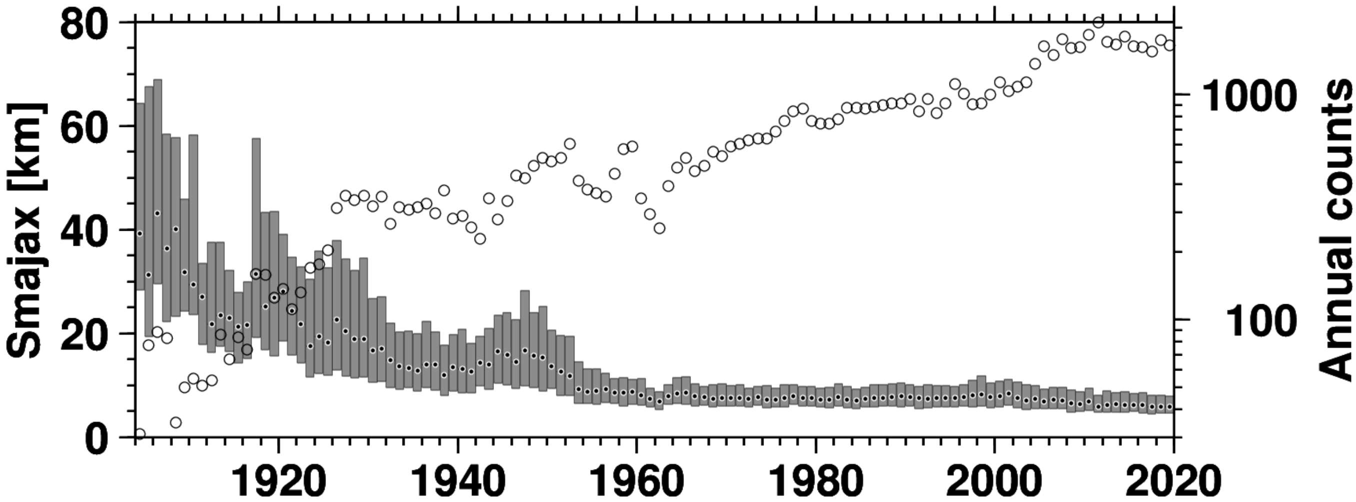

Figure 1 Annual median (small filled circle with white outline) and interquartile range (filled grey box) of the semi-major axis of the error ellipses ($Smajax$) in the ISC-GEM Catalogue, version 10, covering the period 1904-2019. We only show the interquartile range (i.e., instead of the classic box-and-whisker plot) because we want to highlight the general trends in location uncertainty over the years (outliers would require a y axis range of 0-2000 km). The large open circles represent the annual number of earthquakes (right-hand y-axis).

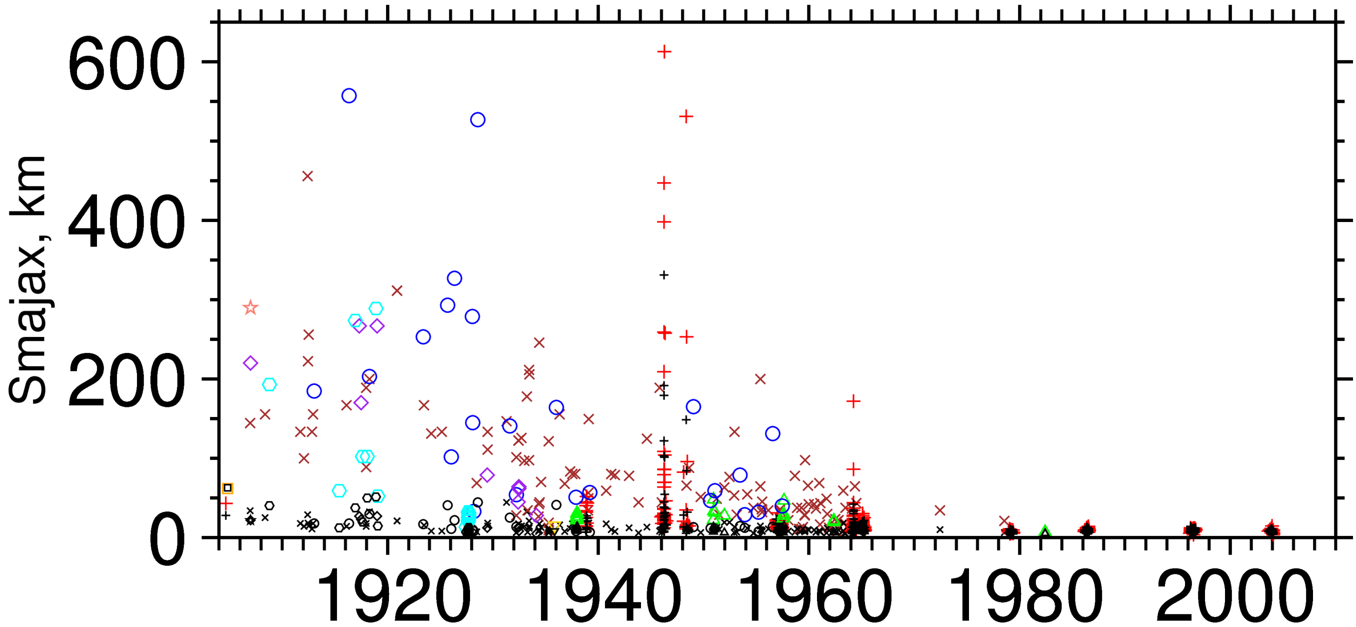

Figure 2 Timeline of $Smajax$ estimates from different works: brown crosses from studies authored and co-authored by D. Doser (Doser, 2001, 2004, 2005, 2006; Fernández Arce and Doser, 2009; Suleiman and Doser, 1995) for earthquakes in different regions and using the location technique of Petroy and Wiens (1989); purple diamonds from studies authored and co-authored by E. Okal (Martin et al., 2019; Okal and Hartnady, 2009; Okal et al., 2010; Okal and Borrero, 2011; Okal et al., 2011) for the 4 January 1907 Sumatra earthquake, three large events during 1917-1919 in the Kermadec-Tonga-Samoa region, three large earthquakes of the 1932 Manzanillo (Mexico) sequence, the 27 June 1929 South Sandwich Islands earthquake and the 14 February 1934 China Sea earthquake. These relocations were obtained with the procedure described in Wysession et al. (1991); red plus symbols from Tape et al. (2017) and Tape and Lomax (2022) as obtained from the NonLinLoc location algorithm (Lomax et al., 2000) for a broad region encompassing Alaska and the Aleutians; cyan hexagons from studies authored by A. Morozov and co-authors (Morozov et al., 2019a,b, 2020) for earthquakes in the Arctic and the 1927 Crimea sequence; green triangles are from Nishenko and Singh (1987) for earthquakes in the Acapulco-Ometepec, Mexico, region; blue circles from Petroy and Wiens (1989) for earthquakes in the northeast Indian Ocean; orange square from Bungum et al. (2009) for the 23 October 1904 Oslofjord earthquake using the HYPOSAT algorithm (Schweitzer, 2001); gold inverted triangle from Niemz and Amorèse (2016) (NonLinLoc solution) for the 10 November 1935 near Monserrat earthquake; pink star from Kanamori et al. (2010) for the 4 January 1907 Sumatra earthquake. The smaller black symbols are the $Smajax$ values in the ISC-GEM Catalogue corresponding to the coloured symbols, as plotted at the same origin time and with the same symbol type. Readers are referred to the works cited here for details about the location procedures adopted by different authors. We only point out that NonLinLoc error ellipses are scaled at 68% and that most of the other $Smajax$ values derive from 90% or 95% confidence levels.

tended as all stations with seismic arrival times available to the ISC). These uncertainty estimates (i.e., errors in the data) clearly impact the uncertainty expected in the location solution.

The inversion procedure of the ISC locator solves for $\mathbf{Gm} = \mathbf{d}$, where $\mathbf{G}$ denotes the matrix containing the partial derivatives of N data by M model parameters, $\mathbf{m}$ the model adjustment vector and $\mathbf{d}$ the vector of time residuals (see for details Bondár and Storchak, 2011). The starting point (latitude, longitude, origin time and depth) can be provided as an instruction parameter when running ISCloc. For ISC-GEM locations (Bondár et al., 2015) we use as starting point the location obtained by the Engdahl et al. (1998) location alghorithm (known as EHB).

When similar raypaths are present, the assumption of independent errors is violated, and the data covariance matrix is no longer diagonal. To account for correlated model structure error (i.e., errors arising when the station separation is less than the scale of the unmodeled 3-D Earth's velocity structure), the ISC locator follows the approach of Bondár and McLaughlin (2009a) and assumes that the similarity between ray paths is well approximated by the station separation. This allows the inter-station covariance to be estimated by using the variogram prescribed by Bondár and Storchak (2011). Note that phases picked at co-located stations are considered fully correlated by the variogram, and the correlation gradually decreases with increasing station separation along the variogram curve. Once the variogram flattens out at the sill value, the pairs of phase picks are no longer correlated. The *a priori* errors for each phase are then added to the diagonal elements of the data covariance matrix
$C_D(i,j) = \sigma^2_{sill} - \gamma(h_{ij}) + \delta_{ij}\sigma^2_{phase}$,
where σ^2_{sill}, $\gamma(h_{ij})$, δ_{ij} and σ^2_{phase} denote the background variance where the variogram levels off at large station separation (the pairs become independent), the variogram value for the distance h_{ij} between the ith and jth stations, the Kronecker delta and the *a priori* estimate of the measurement error covariance for an observed phase, respectively. Once a convergent solution is found, the error ellipses are obtained from
$(\mathbf{r} - \mathbf{r}_{loc})^T \mathbf{C}_M (\mathbf{r} - \mathbf{r}_{loc}) = \kappa^2_\alpha$,
where $\mathbf{r}_{loc}$ denotes the location vector of the epicentre, κ^2_α is defined following Jordan and Sverdrup (1981) and $\mathbf{C}_M$ is the posteriori model covariance matrix $\mathbf{C}_M = \mathbf{G}^{-1}\mathbf{C}_D\mathbf{G}^{-1^T}$. All ISC-GEM error ellipses are scaled to α = 90 confidence level. Since we can assume that the correlated travel time errors are time-independent (i.e., unmodeled 3-D heterogeneities in the Earth's structure do not change over the instrumental period of seismology), the components that can be affected over this time period are solely the observational errors, accounted for in the diagonal elements of the covariance matrix.

Along with the error ellipses, the ISC locator also computes various metrics (Bondár et al., 2004; Bondár and McLaughlin, 2009a) that help to assess the quality of a solution. These include maximal secondary azimuthal gap ($sgap$) computed not only for the entire network but also for different distance ranges (local: 0-150 km; near regional: 3°-10°; teleseismic: 28°-180°), number of defining phases ($ndef$), number of defining stations ($ndefsta$) and number of independent defining phases ($nrank$). A seismic phase arrival time is set as defining if its residual is within $sigmathres$ times that of the corresponding *a priori* error ($sigmathres$ = 6 in our configuration). As guideline, we can expect that a more robust location is obtained when $sgap$ is low (ideally down to 180° or less) and it fits as many observations as possible (i.e., $ndef$, $ndefsta$ and $nrank$ are as high as possible).

2.2 Literature review

Here we provide examples of the observational error estimates from the literature. The works considered are not intended as a comprehensive review but rather to serve as an independent assessments of the errors in the observations both the in early-instrumental and modern period.

For the early-instrumental period we have to consider that seismograms were recorded on paper (with time marks) moving at variable speed depending on the seismograph setup. In this context, observational error sources have been summarised well by Freedman (1968) as:

- miscounts (e.g., misread seconds, minutes or even hours);
- misidentifications (e.g., when due to poor signal-to-noise ratio the first arriving P signal may be overlooked);
- instrumental errors (e.g., clock drift, variable paper speed);
- reading errors (e.g., depending on the analyst picking experience).

To these error sources we can add typos (e.g., wrong minute, hour etc.) in parametric data originating either from original data sources (e.g., printed station bulletins) or from the digitisation of arrival times before reports to the ISC became fully electronic.

Jeffreys (1931) was probably the first to identify observational error sources by analysing the 1923-1927 time residuals in the ISS (1918–1963) and provided the first assessment of time instrumental errors:
"*On machines with speeds of paper of 8 mm, 1.5 cm, and 30 cm per minute respectively, this corresponds to uncertainties in timing of 4 s, 2 s, and 1 s. Errors exceeding 4 s on this ground are therefore not to be expected*"[5]. Jeffreys (1931) also notes that "*the clock error at most stations is probably trustworthy within a second or so, though there are some stations where the consistency of P and S residuals of large amount points to clock errors of 30 s or more*".

Relocations of significant earthquakes have been performed in many works authored and co-authored by E. Okal (some of which are included in Fig. 2 with purple diamonds). In such studies relocations have been done for early-instrumental (and to some extent also

[5] Jeffreys (1931, page 333). Note that we quote the 30 cm speed as in the original paper, however this is probably a typo and most likely a speed of 3.0 cm was intended.

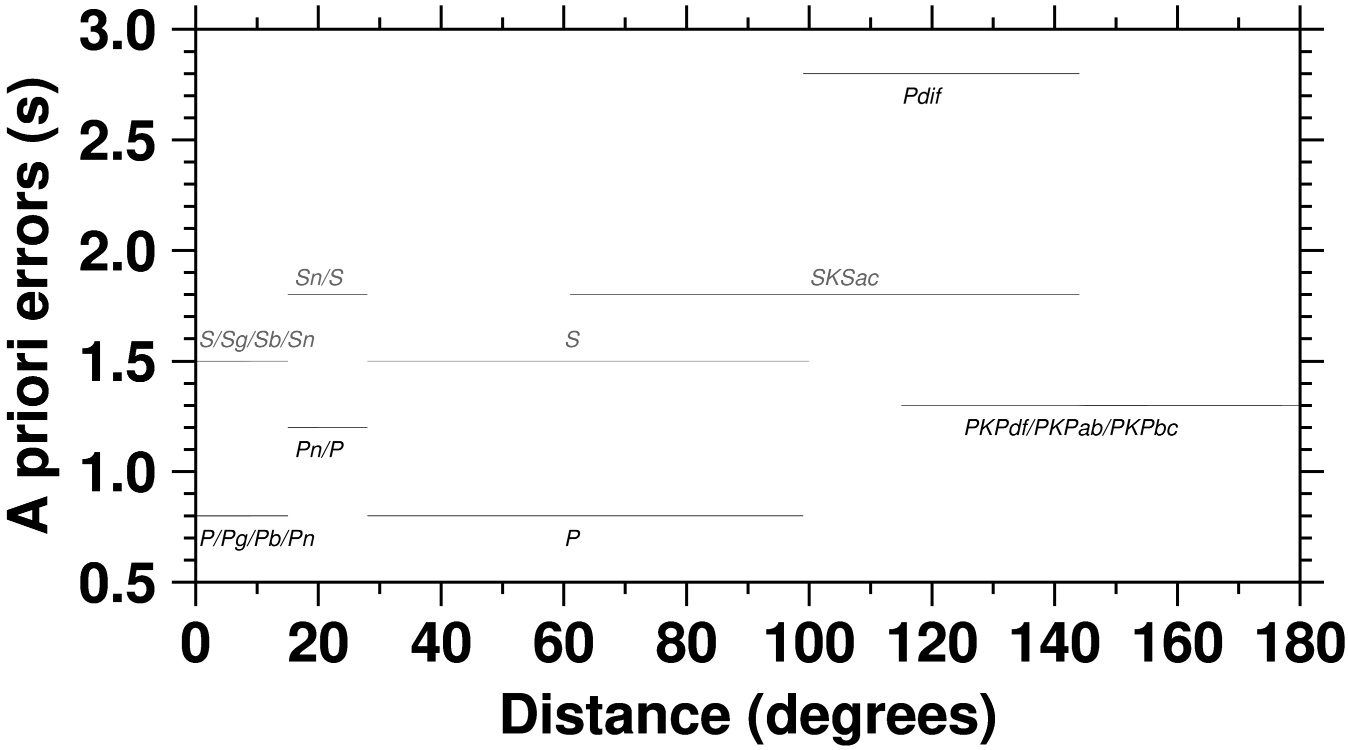

Figure 3 *a priori* errors for commonly picked phases as function of distance (P waves in black and S waves in dark grey).

modern) earthquakes using the interactive method of Wysession et al. (1991), which allows the injection of Gaussian noise (σ_G) into the arrival times. The choice of σ_G appears to be based on the assessed reliability of the data in a given year or on expert-judgement. Fig. 4 summarises the σ_G values for specific earthquakes (see the Supplementary Material for details about events and references used to create the plot). At times the value of σ_G is set following special circumstances, as it is for the unique case of the great Aleutian and Chilean earthquakes of 17 August 1906, which are separated by 30 minutes, where Okal (2005) used a σ_G value of 35 s due to the challenging situation created by the overlapping recordings resulting from these two earthquakes. As such, a σ_G of 35 s should be considered an outlier and not representative of the error measurements in 1906. Indeed, even for the Sumatra earthquake of 1907 σ_G goes down to 12.5 s (Martin et al., 2019). Nevertheless, the important feature shown in Fig. 4 is the overall decrease of σ_G with time. The minimum of 1 s is reached for relocations of earthquakes occurring in 1963 (Okal et al., 2011).

Relocations of early-instrumental earthquakes in Alaska and the Aleutians have recently been performed by Tape et al. (2017) and Tape and Lomax (2022) using NonLinLoc (Lomax et al., 2000) (red plus symbols in Fig. 2). For the 1904 Central Alaska earthquake Tape et al. (2017) use "*observation uncertainties*" of 10 s whereas for earthquakes belonging to Aleutian-Alaska megathrust early-instrumental (from 1938) sequences Tape and Lomax (2022) use 1 and 4 s for P and S arrivals, respectively. Earthquake relocations done by Morozov et al. (2019a,b, 2020) (cyan hexagons in Fig. 2) were obtained with "*onset picking errors*" of 2 or 3 s.

In the more recent instrumental period we see an important reduction in the uncertainty of the observations due to the significant improvements in instrumentation and time keeping that took place from the early 1960s. Freedman (1968) suggested that high quality instruments, such as the ones belonging to World-Wide Standardized Seismograph Network (WWSSN, Oliver and Murphy, 1971; Peterson and Hutt, 2014), make the instrumental error negligible compared to other error sources. Therefore, considering the increased number of reliable stations deployed around the world and the consequent reduction in earthquake location uncertainty, it became possible to use residuals of seismic arrival times as a reliable and meaningful input to infer features the Earth structure (e.g., Poupinet, 1979). After decades of observations this allowed the derivation of improved 1-D Earth models (Kennett and Engdahl, 1991; Kennett et al., 1995). In this context picking error measurements have not been a main focus of research (Zeiler and Velasco, 2009) as their contribution in travel time residuals has been generally considered (understandably) of secondary importance compared to the error related to Earth's heterogeneities. Yet some studies considered the uncertainty related to picking errors. For example, Freedman (1966), in a pioneering study to quantify picking errors from seismograms recorded at Mineral, California, for earthquakes in 1962, reports that reading errors for P-waves are expected to be within 0.2 s and 1 s in cases of clear or poor signals, respectively. High and low quality picks made by Jordan and Sverdrup (1981) were assigned standard errors of 0.7 s and 1 s, respectively, in line with findings by Freedman (1967) and Evernden (1969). Billings et al. (1994) assumed a standard deviation of 0.25 s and 0.5 s for P-waves and S-waves, respectively, in their investigation on errors in location. Douglas et al. (1997) used

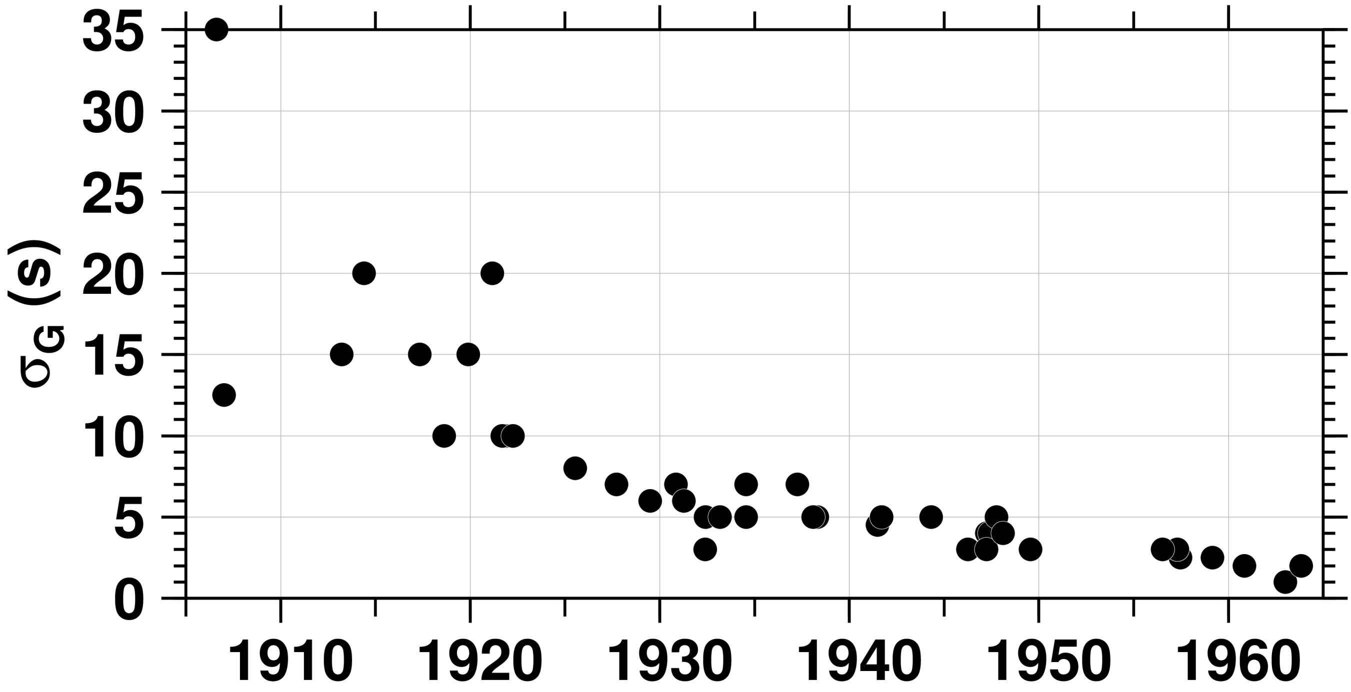

Figure 4 Gaussian noise (σ_G) added to the arrival times for relocating specific earthquakes in works authored and co-authored by E. Okal. The details about events and corresponding papers are in the Supplementary Material.

both explosions and earthquakes to suggest that the true picking error of P-wave onsets can be over 0.5 s (up to several seconds). A recent investigation on the picking error problem from digital waveforms has been done by Zeiler and Velasco (2009). They show that picking errors depend on distance, noise and even the institution reporting picks. Among other findings, they estimated a standard deviation of about 0.4 s and 0.3 s for regional and teleseismic picks of P-waves, respectively. At regional distances, however, errors may exceed 3 s for other phases.

2.3 Travel time residuals from ISC-GEM locations

The station data used to obtain the locations listed in the ISC-GEM Catalogue allows us to explore features of the travel time residuals over the entire instrumental period of seismology since, as mentioned earlier, the locations have been obtained using the same 1-D model (ak135, Kennett et al., 1995) and algorithm (Bondár et al., 2015). In the version released in March 2023 (V10, see Data availability), the catalogue covers the period 1904-2019 and includes moderate earthquakes down to magnitude 5.5 and 5 in the early-instrumental and recent period, respectively (see catalogue overview page at www.isc.ac.uk/iscgem/overview.php).

The selection criteria to include an earthquake in the ISC-GEM Catalogue are detailed in Di Giacomo et al. (2018). The important aspect to recall here is that magnitude is the driving parameter (i.e., all earthquakes with magnitude 5-5.5 are considered). It follows that if an earthquake of moderate size is poorly recorded (only a handful of stations are available) and the network geometry is far from desirable (e.g., $sgap$ > 270°), we still try to relocate it and include it in the catalogue. For this work, however, we are interested in describing prominent time-variable trends of the travel time residuals. To capture such trends one would ideally require that the starting input is free, to the largest extent possible, of mislocated events (as done by Bondár and Storchak, 2011, with the GT List). As we cannot use the GT List in the early-instrumental period, we only consider those earthquakes in the ISC-GEM Catalogue that have $ndefsta \geq 10$ and teleseismic $sgap \leq 220°$ up to 1912 and $\leq 180°$ from 1913. We use the teleseismic $sgap$ instead of the whole network $sgap$ since a) in the early-instrumental period the majority of arrival times are in the teleseismic distance range; b) the model error contribution to the time residuals is lower (i.e., observations at regional distances are notoriously more affected by Earth's heterogeneities, see, e.g., Myers et al., 2010). The larger $sgap$ criterion for the early years of the 20th century is necessary in order to have a larger number of earthquakes (N=58) selected (only 14 earthquakes have teleseismic $sgap \leq 180°$ before 1913). Such criteria do not guarantee the exclusion of potentially mislocated earthquakes but we assume that their bias is negligible considering the size of the whole dataset (Table 1).

From the selected earthquakes we consider only phases identified in the relocation as P and in the teleseismic distance range $28° \leq \Delta \leq 95°$. Their travel time residuals with respect to the ISC-GEM locations are shown in Fig. 5, 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).

Table 1 provides some statistics of the travel time residuals shown in Fig. 5.

About 10% and 1% of the teleseismic P-wave residuals are larger than $\pm$ 10 s in the early-instrumental and recent period, respectively. The bottom panel of Fig. 5 also shows that one minute errors are among the most

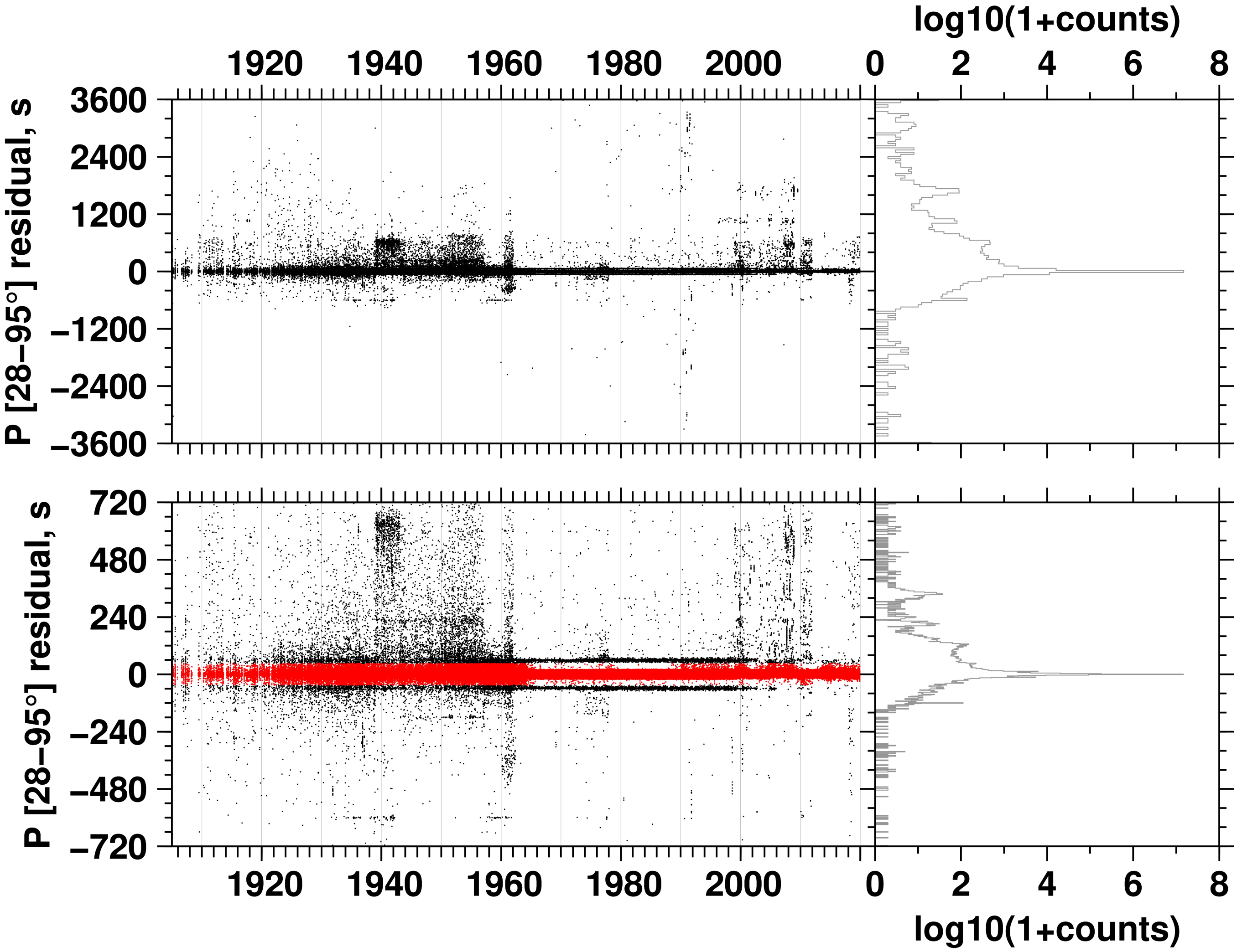

Figure 5 Top, from left to right: travel time residuals of teleseismic P-waves for the selected earthquakes within $\pm$ 1 hour and corresponding histograms of the travel time residuals in bins of 45 s. Bottom, from left to right: same as top but zooming in for travel time residuals within $\pm$ 12 minutes and histogram bins of 10 s; In red are shown the residuals within $\pm$ 45 s.

Period	N(selected earthquakes)	N(P 28°-95°)	±60 s	±30 s	±10 s	±5 s
1905-1963	6699	433K	97.72	95.83	90.15	82.80
1964-2019	57853	14.5M	99.95	99.89	99.41	98.18

Table 1 Summary of the travel time residuals of teleseismic P-waves for the selected earthquakes. Each row shows statistics for the two periods (early-instrumental and recent): number of earthquakes and teleseismic P-wave travel time residuals, percentage of the residuals within different limits. Note the much larger dataset for 1964-2019 compared to 1905-1963.

common large outliers, even in the recent period. In addition, in the 2000s it appears that a larger number of large positive residuals occur with respect to preceding years. We attribute these features in the recent period to a combinations of the following factors:

- minute mark errors may still be present since the 1960s before analogue stations were upgraded with digital setup;
- phase type misidentifications, resulting in large residuals since ISCloc cannot rename a P-type phase to a S-type phase, and vice versa;
- ISC-GEM locations for earthquakes during 1964-2009 are based on the phase information available before the rebuild work of the ISC Bulletin started (Storchak et al., 2017, 2020), where many of the previous errors have been corrected.

Nevertheless, in the recent period the percentage of teleseismic P-wave large residuals is below 1% (Table 1) and they do not not affect the ISCloc final solutions.

Starting from the dataset summarised in Fig. 5 and Table 1, we compute an annual uncertainty as quantified by the standard deviation, particularly for the early-instrumental period. To remove the effects of large outliers and to keep the range of travel time residuals within a meaningful interval for earthquake location (e.g., 60 or 120 s residuals are likely due to miscounts and can eventually be fixed during manual review), we tested both outlier removal (Z-score and interquartile range) and cut-off threshold of 35, 45 and 55 s. We found that the P-wave residuals within 45 s (red points in Fig. 5) bring the most satisfactory results since they minimise the contribution of minute marks or larger errors as well as gross misidentifications/reading errors, but they are still large enough to capture a wide range of travel time residuals.

Fig. 6 illustrates the annual standard deviation (σ_P) of the travel time residuals of the subset shown in red in Fig. 5. Keeping in mind that our intention is not to estimate model error contributions, we emphasise features in Fig. 6 relevant to this work.

Considering that σ_P can be interpreted as a proxy of the state of health of the global network, it is not surprising to observe an overall decrease with time. A significant step can be identified in 1964 (σ_P drops from 4.1 in 1963 to 3.1 in 1964), coinciding with the significant improvements in earthquake monitoring technology, the establishment of the ISC (i.e., arrival times are reviewed by ISC analysts). We can consider the σ_P in the recent period as not characterised by large steps, and the σ_{phase} depicted in Fig. 3 with a $sigmathres$ = 6, as set in ISC procedures, acceptably captures the σ_P for the period 1964 onwards, suggesting that *a priori* errors describe the observational errors well in this period.

In the early-instrumental period, however, variation in σ_P is far more significant. In general, a tendency of σ_P to decrease over time is observed (σ_P = 19.4 in 1905, 6.0 in 1950, 2.0 in 1999, and between 1.3 and 1.4 since 2009). The decrease is not entirely steady over the years with the presence of fluctuations more pronounced from 1908 up to the early 1920s and in some years during the 1940s. These time periods were affected by disruptions in data collection and seismic station operations related to global conflicts. Some of strongest residual signatures (often positive) depicted in Fig. 5 occur in these years and are likely due to data issues and phase type misidentifications. We intend to investigate thoroughly these features in future work (as outlined in the Conclusions). In addition, other issues affecting the years 1908-1912 are due to a combination of dips both in the record of global earthquakes and bulletins (details in Di Giacomo et al., 2018). All these features can be seen in the bottom panel of Fig. 6. For each year we also checked for travel time residual variations as a function of distance (e.g., as found in Zeiler and Velasco, 2009), but no conspicuous feature appeared in our early-instrumental teleseismic dataset.

We suggest that the large σ_P (i.e., $>$ 10-15) observed in the first decades of the 20th century are mostly due to instrumental errors, with clock drifts being dominant. It follows that model error contributions to travel time residuals in the early-instrumental period are dwarfed by observational errors (i.e., early-instrumental observational errors are on a different scale compared to the recent period). Consequently, the early-instrumental period requires larger errors than the ones currently in use (Fig. 3). We will discuss how we take into account the large nature of the observational errors by modifying the assumed errors in the data covariance matrix.

2.4 Remarks regarding observational errors

The literature review and the features of the travel time residuals discussed above allow us to highlight the following aspects of the observational errors:

- they are due to several components, with the instrumental error being significant before the early 1960s. For the early-instrumental period the estimates (or assumptions) by different authors vary from 1 to tens of seconds, also depending on the year an earthquake occurred. However, the instrumental component (e.g., time keeping) of observational errors decreases with time;
- although estimates by various authors may differ, there is a general consensus that observational errors are minimal since the early 1960s (coinciding with upgrades in instrumentation and time keeping, e.g., as a result of the introduction of the WWSSN and ESSN). In this respect, we point out that the *a priori* error measurements set in standard ISC procedures (i.e., since 1964) and adopted also for the ISC-GEM Catalogue locations, appear to agree with findings in the literature, if not being more conservative. Hence we focus on the early-instrumental period in the following sections;
- Clock errors are likely to be the most important source of observational errors for the first part of the 20th century. Their presence introduces bias even if arrivals have been picked precisely. Clock errors may be detected by investigating the timeline of the travel time residuals of a seismic station (or by waveform cross-correlation, but this option is not feasible in the early-instrumental period). 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.

We therefore seek to approximate the observational errors for early-instrumental earthquakes with a parameter that we call timing uncertainty (σ_t). This uncertainty affects P and S phase observations equally and is added to σ_{phase} in the diagonal elements of the data covariance matrix:
$C_D(i,j) = \sigma^2_{sill} - \gamma(h_{ij}) + \delta_{ij}(\sigma_{phase} + \sigma_t)^2$.
In this way we aim to better account for the timing errors introduced due to clock drifts, and assume that these timing errors become negligible in the mid 20th century onwards as the number of stations with improved clocks and timing setup progressively improves.

3 Assessing σ_t for early-instrumental earthquakes

To allow a systematic approach to the relocation of early-instrumental earthquakes, we require a measure of the time varying observational uncertainty σ_t that results from the various sources discussed above. The approach taken by Bondár and Storchak (2011) (summarised in Section 2.1) cannot be applied for this time

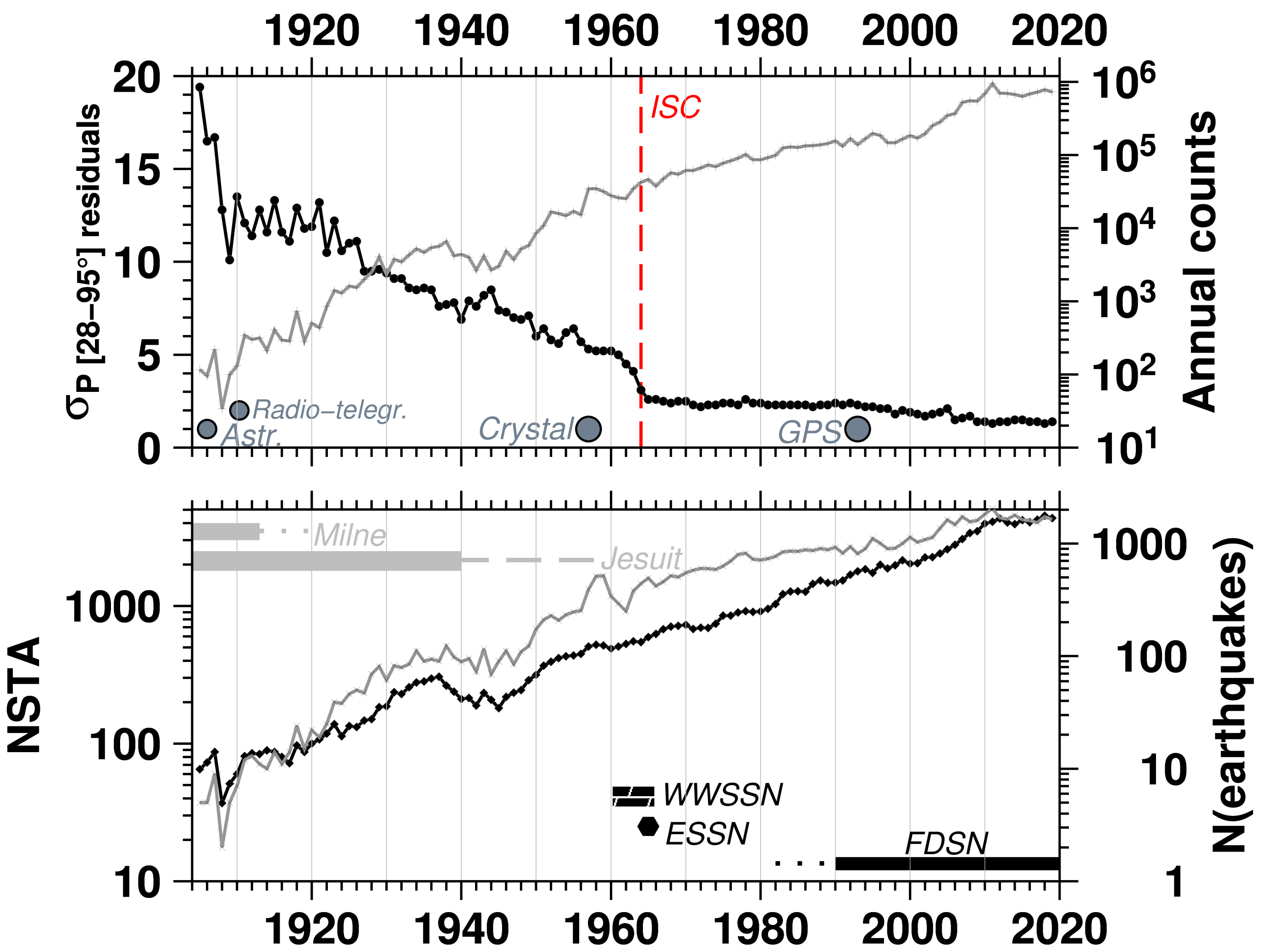

Figure 6 Top panel: annual standard deviation (σ_P) for teleseismic P travel time residuals (black solid line and filled circles, left-hand y axis) along with the number of residuals per year (grey line and plus symbols, right-hand y axis); Bottom panel: annual number of stations providing teleseismic P-wave arrivals (black solid line and filled diamonds, left-hand y axis) along with annual number of earthquakes (grey line and cross symbols, right-hand y axis). Inspired by Storchak et al. (2015), the top panel also shows important changes in time keeping setup as well as the ISC establishment in 1964 (red dashed line), and the bottom panel shows the operational time of early seismic networks such as the Milne and Jesuit (grey segments, including dotted and dashed parts) as well as the start of the WWSSN (black hachure pattern), the Unified System of Seismic Observations in former USSR (ESSN acronym is from Russian translation, black hexagon) and the Federation of Digital Seismograph Networks, FDSN (dotted and thick solid black segment). More information about seismic networks can be found, e.g., in Udias and Stauder (1996); Ammon et al. (2010); Storchak et al. (2015), and references therein.

period, as we do not have GT events from which to derive the phase uncertainty. Here we outline two approaches where we have attempted to independently quantify these errors.

In the first approach, we test a range of values of observed phase uncertainty (σ_{obs}) for the selected earthquakes (Section 3.1). We propose that $\sigma_{obs} = (\sigma_{phase} + \sigma_t + k)$, where σ_{phase} is the uncertainty for a specific phase (Fig. 3), σ_t quantifies the expected increase in uncertainty in the early instrumental period, and k is a correction factor. As the methods we propose to estimate the timing related uncertainties differ from those used by Bondár and Storchak (2011) to derive the phase uncertainties in the modern instrumental period, we do not expect an exact match between the values of σ_{obs} derived in this study and the values σ_{phase} derived by Bondár and Storchak (2011). We contend that the modern day values of σ_{phase} are more robust than the values derived for the early-instrumental period, in the absence of GT events. Therefore we introduce the constant k to our definition of σ_{obs}, to ensure that the derived values of σ_t tend to zero in the recent period.

The year at which σ_t can be considered to be zero is a matter of interpretation, but in this study it is assumed to be 1963. This year is chosen as the timing uncertainties stabilise in the late 1950s and early 1960s (as quantified in Section 3.2). It is also after several GT events that were used to derive the values of σ_{phase} for the modern period, and therefore is within the period where the GT list based analysis of Bondár and Storchak (2011) can be considered valid. In the second approach, we utilise a grid search to take a broad view of the different plausible fits to the data (Section 3.2). From here we derive the likely standard deviation in phase residuals for each earthquake and then in each year in the early-instrumental period, and use this as a measure of σ_{obs}.

FFF

3.1 Impact of varying σ_t for early-instrumental earthquake relocations

Increasing the assumed prior error in the early-instrumental period not only has the effect of increasing the area of the horizontal error ellipse, but also impacts other metrics such as the percentage of phases that contribute to the location of a given earthquake and ultimately controlling the derived earthquake location. This is demonstrated in Fig. 7 with six events from different decades (three in the early instrumental and three in the recent period) and seismic regions. In this demonstration the locations and corresponding error ellipses have been computed using ISCloc while varying the value of σ_{obs} so that its value for all phases ranges between 2 and 30 s and 2 and 10 s for early-instrumental and recent earthquakes, respectively, and increasing in steps of 1 s for each relocation run. The locations for this demonstration were run with the earthquake depth fixed to the ISC-GEM Catalogue depth, as most early-instrumental earthquakes lack depth resolution (Bondár et al., 2015), and we are primarily interested in assessing the impact on the horizontal component of error ellipses. We also performed the same demonstration while allowing the locator to solve for the depth, but this did not greatly affect our results, as the lack of depth resolution in the early-instrumental period causes ISCloc to adopt a default depth for these events anyway. The assumed default depth does not usually change between runs, hence the location changes are similar to the ones obtained by fixing the depth.

Increasing σ_{obs} also increases the residual threshold where arrival times become time-defining, as a phase pick is used in the relocation if its travel time residual is less than six times σ_{obs}. Therefore we expect to see improvements in the percentage of time defining phases. Fig. 7 shows a clear contrast between early-instrumental and recent period earthquakes. For the recent earthquakes we observe only small location changes that are within the ISC-GEM error ellipse, and the higher values of σ_{obs} have a limited and largely insignificant effect on the earthquake location, and mainly serve to increase the estimated error ellipse. In contrast, for the early-instrumental earthquakes we see much broader variations in earthquake location that encompass areas much wider than the current ISC-GEM error ellipses.

Over a certain level of assumed error in the seismic phase arrivals the locations do not vary any more, and only the error ellipses become larger. For this reason it is not possible to see red diamonds in Fig. 7 for the largest σ_{obs} values, only the corresponding error ellipses. This saturation in location changes is also observed for parameters such as the percentage of associated phases that are used in the relocation and secondary azimuthal gap.

To further demonstrate the effects of adopting a varying σ_t in ISCloc, Fig. 8 shows the summary of the test (as described for Fig. 7) on the 29 sets of locations for the 13 selected earthquakes in 1912. As demonstrated by the examples shown in Fig. 7, the resolved earthquake locations saturate and no longer move with increasing σ_{obs} at around 10, while the error ellipse continues to scale with increasing σ_{obs}.

Increasing the value of σ_{obs} directly impacts the size of the inferred error ellipse, as it directly informs the diagonal elements of the posterior covariance matrix used to derive the error ellipse, as is demonstrated in the top left panel of Fig. 8. The bottom left and right-side panels of Fig. 8 show the location differences relative to the ISC-GEM locations. In the example year of 1912 the earthquake locations are altered between 20 and 220 km for different events. The shift in locations stops when the assumed error is approximately 10 s. At this point increasing σ_{obs} further does not allow any further phase observations to be accounted for in the relocation and the locations therefore remain stable from this point. This test therefore gives an indication of the likely value of σ_t needed to account for the distribution of observations in this period. We have not, however, found a satisfactorily stable way of quantifying this for each year of the early-instrumental period and suggest that this approach, based on the linear relocation inversion, may be overly sensitive to radical outliers in the arrival time observations, the many potential sources of which are summarised in Section 2. We therefore explore a more stable approach below.

3.2 Assessing the standard deviation of early-instrumental arrival times

To obtain location uncertainty estimates that reflect the observational errors associated with early-instrumental earthquakes, we require a rigorous measure of the standard deviation of travel time residuals which is independent of the adopted location. To achieve this we analyse the distribution of travel time residuals at each point produced by an adaptive grid search based on the neighbourhood algorithm (Sambridge, 1999) and implemented in ISCloc. Here we seed the grid search over potential locations with 4500 randomly distributed samples that are within 10° distance of the ISC-GEM location of each selected earthquake. As with the linear results described above (Section 3.1), the earthquake depth is fixed to the ISC-GEM depth throughout the grid search. The broad starting assumption allows a large range of locations to be considered and is designed to allow multiple local minima that fit different portions of the associated phase data to be sampled and explored (i.e., the large radius of the grid search should include all likely locations). The initial distribution of 4500 points is then resampled 1000 times, with the Voronoi cells surrounding the 250 lowest misfit points being resampled, meaning each Voronoi cell is resampled 4 times. This process is repeated for 25 iterations, producing a total of 29,500 samples for each of the selected early-instrumental earthquakes.

In this analysis, we are primarily interested in computing the standard deviation of the travel times residuals at each point of the grid search, rather than optimis-

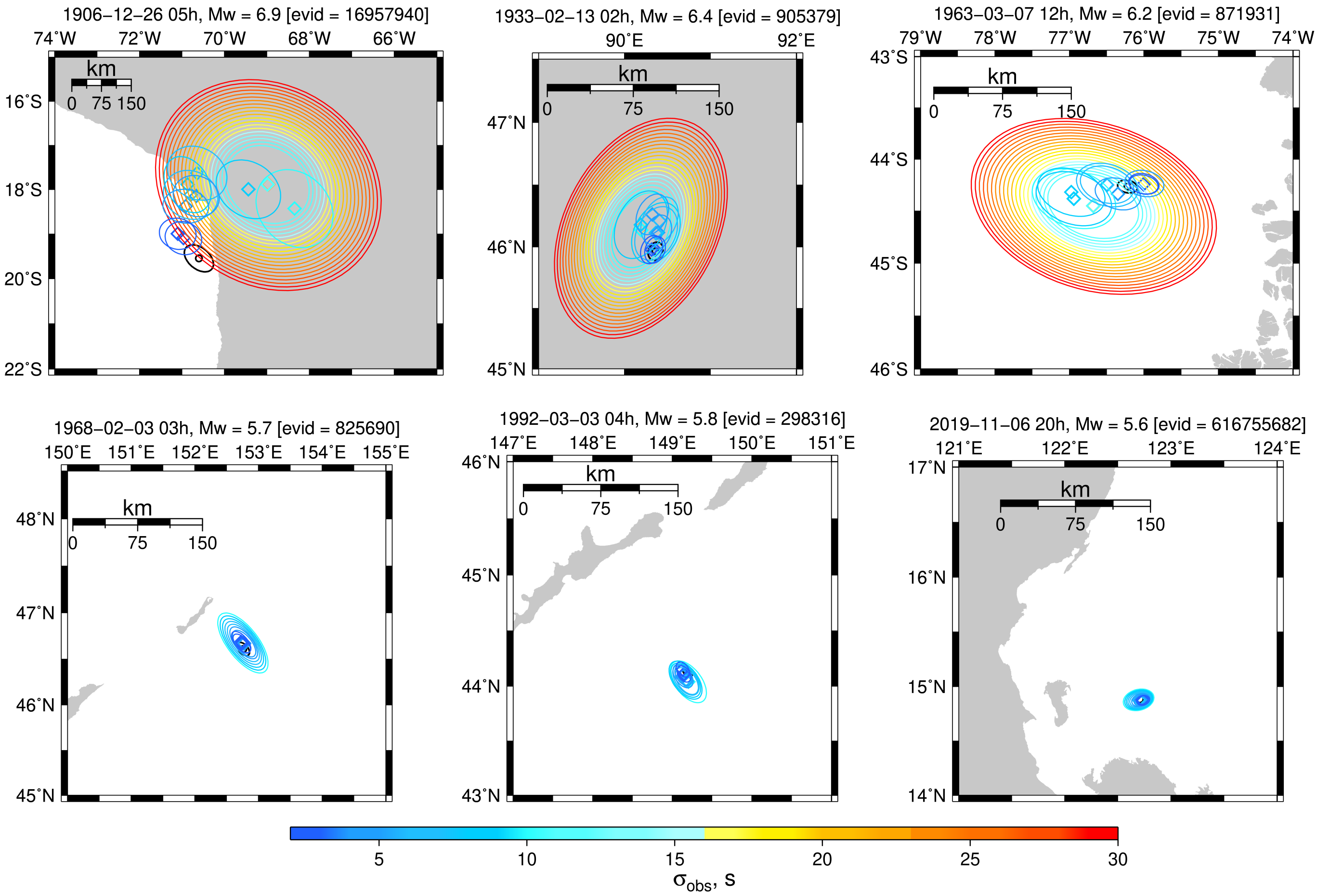

Figure 7 Test results for six earthquakes in different parts of the world and decades. In each plot we show the current ISC-GEM location (black open circle with corresponding error ellipse in black) and the locations obtained in the test with $\sigma_{obs} = (\sigma_{phase} + \sigma_t + k)$ between 2 and 30 s (diamonds and corresponding error ellipses, colour-coded by σ_{obs}). For each earthquake we provide the event identifier ($evid$), which can be used to retrieve the full solution from the ISC-GEM Catalogue or ISC Bulletin. Compared to early-instrumental earthquakes, this test confirms that σ_t is not required for earthquakes in the recent period (hence we only show the test up to σ_{obs} = 10 s for the three recent earthquakes).

ing or minimising the absolute values[6]. At each sample obtained from this extensive grid search, we calculate the travel time residuals for each of the reported P-wave picks associated to the event. For an event with 20 P-wave observations we compute the travel time residuals for all 20 observations at each of the 29,500 grid search samples. The standard deviation of the travel time residuals is then calculated from a trimmed dataset where the largest 5% and lowest 5% of travel time residuals are removed. This is designed to minimise the impact of spurious outlying observations that can lead to an overestimation of the standard deviation.

Fig. 9 summarises this approach for a single event, with the left panel demonstrating the spatial distribution of the grid search results and the standard deviation measured from the P-wave observations. The right panel of Fig. 9 shows 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. The majority of the travel time residuals are distributed around zero and approximated well by a Gaussian distribution, while a smaller number of outlying incoherent travel time phases are scattered in the outlying lobes of the histogram.

Through this estimation of the optimised source position, we derive the standard deviation in the travel time residuals. These residuals are due to a combination of timing errors and other sources of uncertainties. This process is performed for all selected early-instrumental earthquakes. The inferred standard deviation for each event is shown in Fig. 10, along with the annual median values. By considering many well recorded early-instrumental earthquakes, the analysis is not overly biased by individual events, stations or regions. This is an empirical and data driven measure of pick uncertainty, with the underlying statistical assumptions that the phase residuals have a Gaussian distribution, and the seismological assumption that the location can be optimised.

A 5th order spline (R Core Team, 2024) is fitted to the annual median values, effectively averaging over the earliest (1907-1920) period, and smoothing some of the sharper variations in measured σ_{obs}. We propose this spline fit as our estimate of σ_{obs} for the early-

[6]Since the calculated standard deviation has no resolution to the earthquake origin time or depth, only to the earthquake location, it further justifies the strategy of sampling the model space with a fixed earthquake depth.

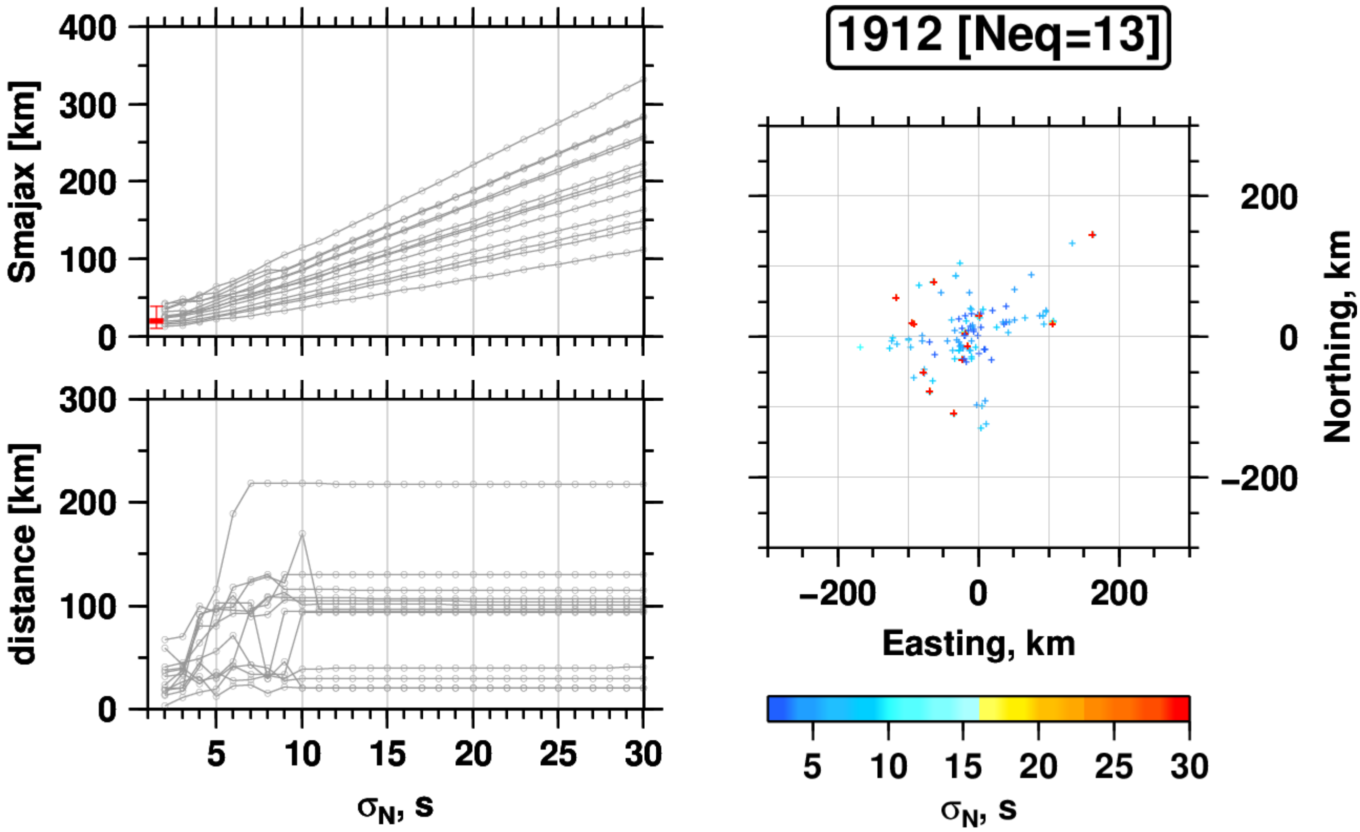

Figure 8 Summary of the location differences for the 13 selected earthquakes in 1912. Top and bottom left: $Smajax$ resulting from our test for each earthquake (along with the box-and-whisker in red for $Smajax$ in the ISC-GEM Catalogue, as computed for the selected earthquakes considered here), and the distance between each location run and the ISC-GEM location, respectively (on both subplots each grey line and open circles corresponds to results for an individual earthquake); Right: Northing-Easting (N-E) plot of the locations out of the test using the ISC-GEM location as reference. As such the N-E subplot shows 29 symbols (colour-coded by σ_{obs}) for each earthquake.

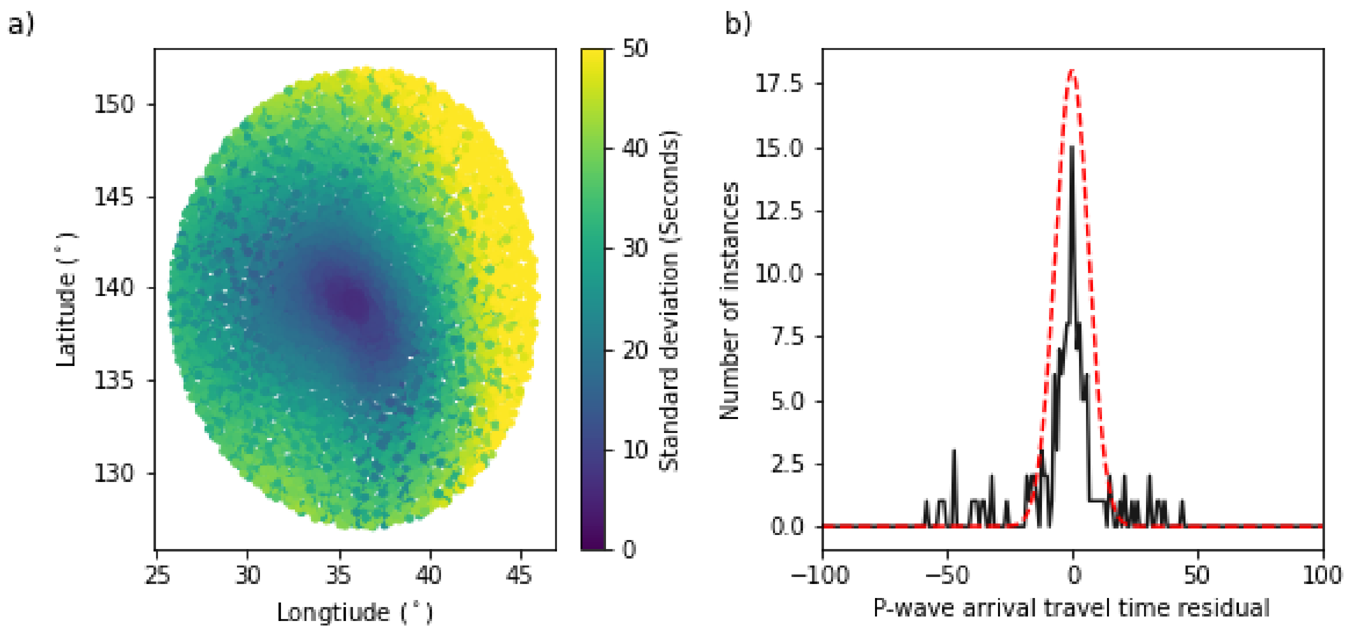

Figure 9 Example of the P-wave standard deviation computation for ISC event = 910800 (1924-01-14 20:50 UTC, Near the South Coast of Eastern Honshu). In this case the standard deviation is obtained from the travel time residuals of 158 P-wave arrivals. Left: spatial distribution of the travel time residual standard deviation (each point is from the grid search described in the text) colour-coded according to the palette on the right-hand side. Right: histogram of the travel time residuals with respect to the point in the grid search with the lowest standard deviation (the dashed red line is the Gaussian distribution fit). From such distributions we compute the event standard deviation after removing the largest 5% and lowest 5% of the residuals.

instrumental period[7]. In order to obtain values of σ_t needed for our revised locations for this period, the most recent measure of σ_{obs} must be corrected to the corresponding measure of σ_{phase} in the recent period. For our proposed values the correction factor $k = \sigma_{obs}(1963) = 0.6$. Values of σ_t for this period can then be determined by $\sigma_t = \sigma_{obs} - k$.

The observed scatter in the standard deviations suggests that while the dominant signal is characterised by a decreasing uncertainty due to better time keeping, large variations can still occur. This possibly reflects a variation in observational quality in different parts of the globe throughout this time period, as well as variations in event-station configurations or event specific data issues.

3.3 Impact of revised travel time uncertainties on location uncertainty estimates

To assess the effect of σ_t on location uncertainties, we relocate the selected early instrumental earthquakes using the revised phase uncertainty assumptions.

Examples for individual earthquakes are shown in Fig. 11 for five earthquakes in Alaska in 1912 ($\sigma_t = 18.9s$) and the 1907 Sumatra earthquake ($\sigma_t = 24.3s$). The new location uncertainty estimates are much larger than previous ones and tend to include current ISC-GEM epicentres. Location changes can be significant in some case (in line with the σ_{obs} demonstration results of Fig. 8) and will be investigated in future work as outlined in the Conclusions.

The results summarising the overall effect of σ_t on the $Smajax$ values for the selected earthquakes in the early-instrumental-period are shown in Fig. 12. The major pattern observed here is predominately controlled by our revised estimates of phase timing uncertainty. There are significant differences with current ISC-GEM location uncertainties, particularly for earthquakes dating from before the mid-1920s. In the following decades the differences with the new location uncertainties are less prominent as σ_t becomes smaller, reflecting improvements in time keeping and general setup of seismic stations around the world.

4 Conclusions

With this contribution we aimed to derive more meaningful location uncertainty estimates for early-instrumental earthquakes as values obtained with the ISC setup for the modern period are significantly underestimated. This is due to large uncertainties in arrival time observations associated with early instrumental earthquakes, before significant technological improvements in seismic networks and observatories around the world took place in the 1960s, significantly improving the accuracy of arrival time picks. To take into account the large uncertainty in the instrumental observations of early instrumental earthquakes we introduce a new term (timing uncertainty, σ_t) to the data covariance matrix that accounts for the variation in timing uncertainty over the early decades of the 20th century. Through a comprehensive analysis of the standard deviation of P-wave travel time residuals, we derive a set of values for σ_t that better captures this source of uncertainty.

The inferred values of σ_t are characterised by high standard deviations and volatility in the earliest portion of the instrumental period, from 1904 to approximately 1924. With these we derived error ellipses with median $Smajax$ of 100 km and above in the first years of the 20th century. We then see a gradual decline in the uncertainty of reported P-wave arrival times, and increasing stability of the measure, and the scale of the location uncertainties decreasing over time. The exceptions to this improving picture regard the years encompassing global conflicts. There is a general agreement of our σ_t values with the Gaussian noise used in various publications by E. Okal and summarised in Fig. 4. The major difference is that our σ_t values are notably higher in the years before 1924. This could be due to our global approach, while published works focus on a limited number of significant earthquakes or specific areas.

Furthermore, we compare the range of derived error ellipses, as quantified by the major axis of the error ellipse shown in Fig. 12, with the range of values from the literature shown in Fig. 2. While the error ellipse from the ISC-GEM catalogue form a lower end of the error ellipse in the early instrumental period, the revised error ellipse calculated using our proposed phase error assumptions are in the centre of the range of values reported in the early instrumental earthquake literature. These variations are easily explained by variations in the data and methods used, as well as plausible regional variations in earthquake location precision in this period.

The results of this work will be used for the planned project of rebuilding the ISC Bulletin for 1904-1963, where we aim to apply modern location and review procedures to early-instrumental earthquakes, as has already been done for the period 1964-2010 (Storchak et al., 2017, 2020). With the 1904-1963 review of the ISC Bulletin the earthquake relocations produced by implementing the error assumptions proposed here will be manually assessed by ISC analysts, ultimately resulting in the Reviewed ISC Bulletin spanning the entire period of instrumental seismology.

Acknowledgements

Emile Okal first pointed out the small size of the ISC-GEM Catalogue error ellipses. Similar observations were made by Phil Cummins and Stacey S. Martin whilst reassessing locations of early-instrumental earthquakes in Indonesia with the first author of this work. We thank them all for drawing our attention to this issue. We are grateful to Anthony Lomax and an anonymous reviewer for the valuable feedback and suggestions provided that helped us to improve this work. The ISC is able to continue its mission thanks to the support of 76 Member-Institutions (http://www.isc.ac.uk/members/ last access: October 2024), including the National Science Foundation (NSF) Award 2414178, Royal

[7]File listing the annual σ_{obs} values so derived is included in the Supplementary Material.

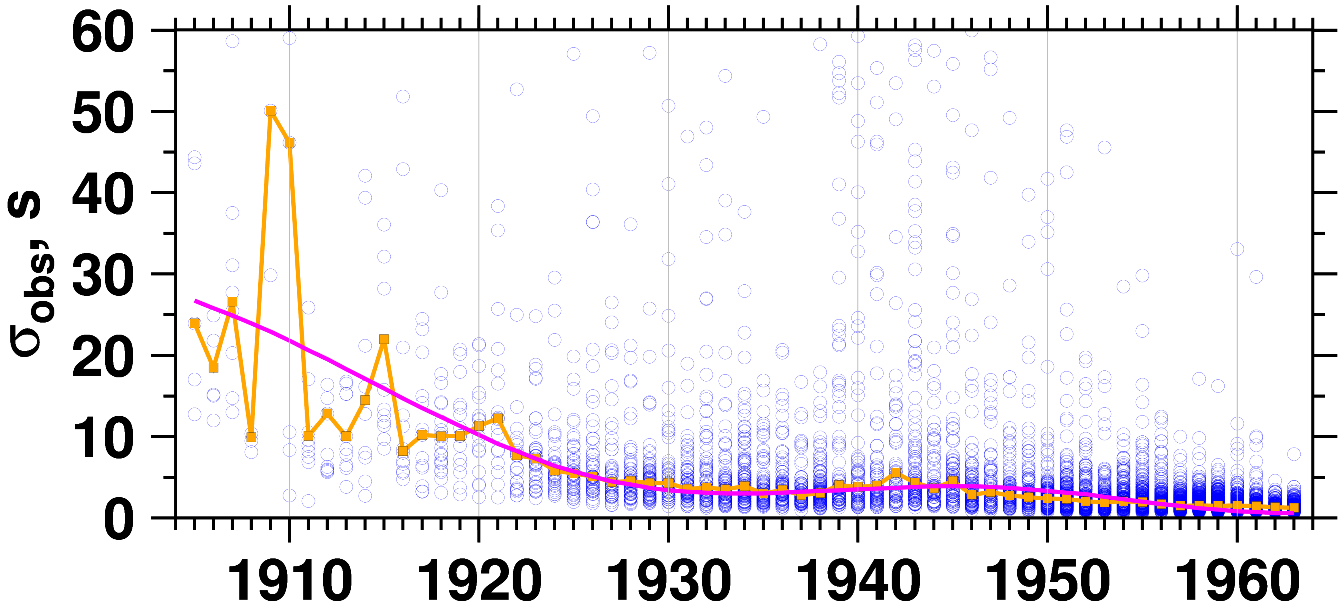

Figure 10 Timeline of the standard deviation of the P-wave travel time residuals obtained for all selected earthquakes. In blue are represented the individual event standard deviations (large outliers are excluded from the plot), in orange their annual median values. Magenta curve represents the 5th order spline (R Core Team, 2024) fitted to these median values.

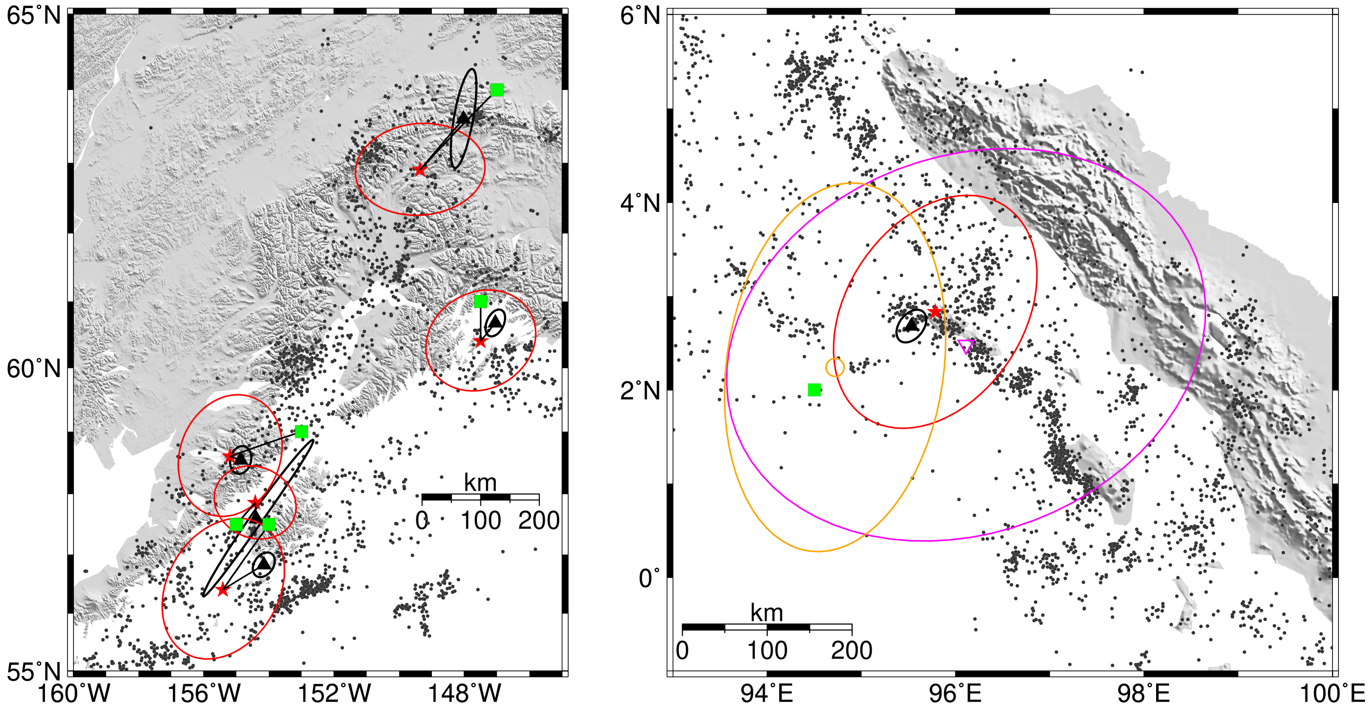

Figure 11 Maps showing current ISC-GEM locations (black triangles) and locations obtained after including σ_t in ISCloc (red stars). Left: solutions for five earthquakes that occurred in 1912 in Alaska along with the Gutenberg locations (green squares). Black triangles and green squares are linked to the red star symbols with thin black segments to identify locations belonging to each earthquake. Right: locations for the 1907 Sumatra earthquake, where the ISC-GEM and Gutenberg locations are depicted as in the left panel, and with the addition of the locations by Kanamori et al. (2010) as inverted magenta triangle and by Martin et al. (2019) as orange circle. Error ellipses are coloured according to their epicentres (except for Gutenberg locations that do not have reported error ellipses). In both maps the small dark grey circles are the ISC-EHB locations (Engdahl et al., 2020) between 1964 and 2021, shown to provide a context of the well-recorded seismicity in the recent period. Topography from NOAA National Geophysical Data Center (2009).

Society Grant INT004 and USGS EHP Award G24 AP 00440-00. The authors are also grateful to sponsors that help to fund the ISC development projects: CTBTO, FM Global, the Lighthill Risk Network, AXA XL, SARA, GeoSIG, GaiaCode and MS&AD InterRisk Research & Consulting. Most figures were drawn using the Generic Mapping Tools (Wessel et al., 2013).

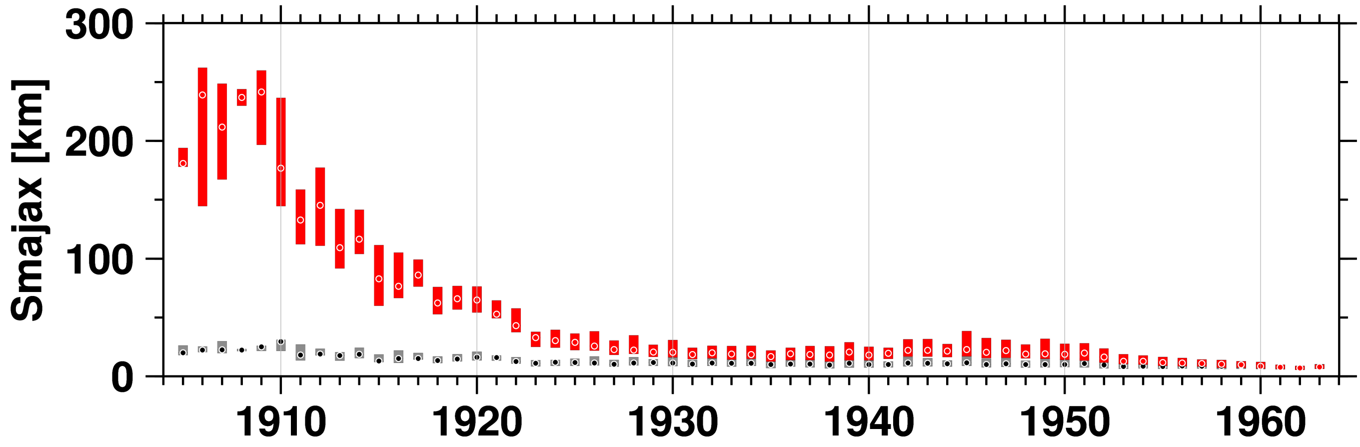

Figure 12 Annual interquartile ranges of $Smajax$ ISC-GEM Catalogue (grey) and relocations done by using σ_t (red). For both sets we only considered the selected earthquakes pre-1964 (hence the interquartile range for the ISC-GEM Catalogue is smaller than in Fig. 1).

Data availability

The ISC-GEM Catalogue is available at http://doi.org/10.31905/D808B825. We used version 10 of the catalogue which was released in March 2023 and covered the time period 1904-2019. The station arrival times for the early-instrumental period are not entirely available in the ISC Bulletin yet. We plan to make them fully available after rebuilding the period 1904-1963, as explained in the main text. The location algorithm used in this work, ISCloc, is available at http://www.isc.ac.uk/iscbulletin/iscloc/. Files referred in the text are included in the Supplementary Material.

Competing interests

The authors have no competing interests.

Supplementary Material

The zip file attached to this submission includes the files we refer in the text.

References

Adams, R. D. Re-evaluation of early instrumental earthquake locations: methodology and examples. *Annals of Geophysics*, 47 (2-3):859–871, 2004. doi: 10.4401/ag-3339.

Agnew, D. C. History of seismology. In Lee, W., Kanamori, H., Jennings, J., and Kisslinger, C., editors, *International Handbook of Earthquake and Engineering Seismology*, volume A, chapter 1, pages 3–11. Academic Press, San Diego, 2002.

Agnew, D. C. Time Marks and Clock Corrections: A Century of Seismological Timekeeping. *Seismological Research Letters*, 91(3): 1417–1429, 2020. doi: 10.1785/0220190284.

Ammon, C. J., Lay, T., and Simpson, D. W. Great Earthquakes and Global Seismic Networks. *Seismological Research Letters*, 81(6): 965–971, 2010. doi: 10.1785/gssrl.81.6.965.

Billings, S. D., Sambridge, M. S., and Kennett, B. L. N. Errors in hypocenter location: Picking, model, and magnitude dependence. *Bulletin of the Seismological Society of America*, 84(6): 1978–1990, 1994. doi: 10.1785/BSSA0840061978.

Bondár, I. and McLaughlin, K. Seismic Location Bias and Uncertainty in the Presence of Correlated and Non-Gaussian Travel-Time Errors. *Bulletin of the Seismological Society of America*, 99 (1):172–193, 2009a. doi: 10.1785/0120080922.

Bondár, I. and McLaughlin, K. L. A New Ground Truth Data Set For Seismic Studies. *Seismological Research Letters*, 80(3):465–472, 2009b. doi: 10.1785/gssrl.80.3.465.

Bondár, I. and Storchak, D. A. Improved location procedures at the International Seismological Centre. *Geophysical Journal International*, 186(3):1220–1244, 2011. doi: 10.1111/j.1365-246x.2011.05107.x.

Bondár, I., Myers, S. C., Engdahl, E. R., and Bergman, E. A. Epicentre accuracy based on seismic network criteria. *Geophysical Journal International*, 156(3):483–496, 2004. doi: 10.1111/j.1365-246x.2004.02070.x.

Bondár, I., Engdahl, E. R., Villaseñor, A., Harris, J., and Storchak, D. A. ISC-GEM: Global Instrumental Earthquake Catalogue (1900–2009), II. Location and seismicity patterns. *Physics of the Earth and Planetary Interiors*, 239:2–13, 2015. doi: 10.1016/j.pepi.2014.06.002.

Buland, R. The mechanics of locating earthquakes. *Bulletin of the Seismological Society of America*, 66(1):173–187, 1976. doi: 10.1785/BSSA0660010173.

Bungum, H., Pettenati, F., Schweitzer, J., Sirovich, L., and Faleide, J. I. The 23 October 1904 MS 5.4 Oslofjord Earthquake: Reanalysis Based on Macroseismic and Instrumental Data. *Bulletin of the Seismological Society of America*, 99(5):2836–2854, 2009. doi: 10.1785/0120080357.

Di Giacomo, D., Engdahl, E. R., and Storchak, D. A. The ISC-GEM Earthquake Catalogue (1904–2014): status after the Extension Project. *Earth System Science Data*, 10(4):1877–1899, 2018. doi: 10.5194/essd-10-1877-2018.

Doser, D. I. A Study of Historic Earthquakes of the Prince William Sound, Alaska, Region. *Bulletin of the Seismological Society of America*, 91(4):842–857, 2001. doi: 10.1785/0120000241.

Doser, D. I. Seismicity of the Denali-Totschunda Fault Zone in Central Alaska (1912-1988) and Its Relation to the 2002 Denali Fault Earthquake Sequence. *Bulletin of the Seismological Society of America*, 94(6B):S132–S144, 2004. doi: 10.1785/0120040611.

Doser, D. I. Historical Seismicity (1918-1964) of the Kodiak Island Region. *Bulletin of the Seismological Society of America*, 95(3): 878–895, 2005. doi: 10.1785/0120040175.

Doser, D. I. Relocations of Earthquakes (1899–1917) in South-Central Alaska. *Pure and Applied Geophysics*, 163(8):1461–1476,

2006. doi: 10.1007/s00024-006-0085-3.

Douglas, A., Bowers, D., and Young, J. B. On the onset of *P* seismograms. *Geophysical Journal International*, 129(3):681–690, 1997. doi: 10.1111/j.1365-246x.1997.tb04503.x.

Engdahl, E. R., van der Hilst, R., and Buland, R. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bulletin of the Seismological Society of America*, 88(3):722–743, 1998. doi: 10.1785/BSSA0880030722.

Engdahl, E. R., Di Giacomo, D., Sakarya, B., Gkarlaouni, C. G., Harris, J., and Storchak, D. A. ISC-EHB 1964–2016, an Improved Data Set for Studies of Earth Structure and Global Seismicity. *Earth and Space Science*, 7(1):e2019EA000897, 2020. doi: 10.1029/2019ea000897.

Evernden, J. F. Precision of epicenters obtained by small numbers of world-wide stations. *Bulletin of the Seismological Society of America*, 59(3):1365–1398, 1969. doi: 10.1785/BSSA0590031365.

Fernández Arce, M. and Doser, D. I. Relocation and waveform modeling of the 1924 Orotina, Costa Rica, earthquake (MS 7.0). *Tectonophysics*, 479(3–4):197–202, 2009. doi: 10.1016/j.tecto.2009.08.010.

Flinn, E. A. Confidence regions and error determinations for seismic event location. *Reviews of Geophysics*, 3(1):157–185, 1965. doi: 10.1029/RG003i001p00157.

Freedman, H. W. The “little variable factor” a statistical discussion of the reading of seismograms. *Bulletin of the Seismological Society of America*, 56(2):593–604, 1966. doi: 10.1785/bssa0560020593.

Freedman, H. W. A statistical discussion of P residuals from explosions. *Bulletin of the Seismological Society of America*, 57(3): 545–561, 1967. doi: 10.1785/BSSA0570030545.

Freedman, H. W. Seismological measurements and measurement error. *Bulletin of the Seismological Society of America*, 58(4): 1261–1271, 1968. doi: 10.1785/BSSA0580041261.

Geiger, L. Herdbestimmung bei Erdbebenaus den Ankunftzeite. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1910:331–349, 1910. http://eudml.org/doc/58769.

International Seismological Centre. On-line Bulletin, 2023a. https://doi.org/10.31905/D808B830.

International Seismological Centre. ISC-GEM Earthquake Catalogue, 2023b. https://doi.org/10.31905/D808B825.

International Seismological Centre. IASPEI Reference Event List, 2023c. https://doi.org/10.31905/32NSJF7V.

ISS. International Seismological Summary. Quarterly issues, 1918–1963. Available at http://www.isc.ac.uk/printedStnBulletins/townpublist.php?townid=544.

Jeffreys, H. The Revision of Seismological Tables. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 2(7):329–348, 1931. doi: 10.1111/j.1365-246X.1931.tb05419.x.

Jordan, T. H. and Sverdrup, K. A. Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific. *Bulletin of the Seismological Society of America*, 71(4): 1105–1130, 1981. doi: 10.1785/BSSA0710041105.

Kanamori, H. The importance of historical seismograms for geophysical research. In Lee, W., editor, *Historical Seismograms and Earthquakes of the World*, pages 16–33. Academic Press, New York, 1988.

Kanamori, H., Rivera, L., and Lee, W. H. K. Historical seismograms for unravelling a mysterious earthquake: The 1907 Sumatra Earthquake. *Geophysical Journal International*, 183(1):358–374, 2010. doi: 10.1111/j.1365-246x.2010.04731.x.

Kennett, B. L. N. and Engdahl, E. R. Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, 105(2):429–465, 1991. doi: 10.1111/j.1365-246x.1991.tb06724.x.

Kennett, B. L. N., Engdahl, E. R., and Buland, R. Constraints on seismic velocities in the Earth from traveltimes. *Geophysical Journal International*, 122(1):108–124, 1995. doi: 10.1111/j.1365-246x.1995.tb03540.x.

Lomax, A., Virieux, J., Volant, P., and Berge-Thierry, C. Probabilistic Earthquake Location in 3D and Layered Models. In Thurber C.H., R. N., editor, *Modern Approaches in Geophysics*, pages 101–134. Springer Netherlands, 2000.

Martin, S. S., Li, L., Okal, E. A., Morin, J., Tetteroo, A. E. G., Switzer, A. D., and Sieh, K. E. Reassessment of the 1907 Sumatra “Tsunami Earthquake” Based on Macroseismic, Seismological, and Tsunami Observations, and Modeling. *Pure and Applied Geophysics*, 176(7):2831–2868, 2019. doi: 10.1007/s00024-019-02134-2.

Morozov, A. N., Vaganova, N. V., and Konechnaya, Y. V. The October 14, 1908 MW 6.6 earthquake in the Barents and Kara sea region of the Arctic: Relocation based on instrumental data. *Polar Science*, 20:160–166, 2019a. doi: 10.1016/j.polar.2019.05.001.

Morozov, A. N., Vaganova, N. V., Shakhova, E. V., Konechnaya, Y. V., Asming, V. E., Antonovskaya, G. N., and Evtyugina, Z. A. Seismicity of the Arctic in the Early Twentieth Century: Relocation of the 1904–1920 Earthquakes. *Bulletin of the Seismological Society of America*, 109(5):2000–2008, 2019b. doi: 10.1785/0120190018.

Morozov, A. N., Vaganova, N. V., Dulentsova, L. G., Asming, V. E., and Evtyugina, Z. A. The 1927 earthquakes and aftershocks in the Crimea: relocation based on instrumental data. *Journal of Seismology*, 25(2):393–417, 2020. doi: 10.1007/s10950-020-09972-x.

Myers, S. C., Begnaud, M. L., Ballard, S., Pasyanos, M. E., Phillips, W. S., Ramirez, A. L., Antolik, M. S., Hutchenson, K. D., Dwyer, J. J., Rowe, C. A., and Wagner, G. S. A Crust and Upper-Mantle Model of Eurasia and North Africa for Pn Travel-Time Calculation. *Bulletin of the Seismological Society of America*, 100(2): 640–656, 2010. doi: 10.1785/0120090198.

Niemz, P. and Amorèse, D. Relocalizing a historical earthquake using recent methods: The 10 November 1935 Earthquake near Montserrat, Lesser Antilles. *Journal of South American Earth Sciences*, 66:166–179, 2016. doi: 10.1016/j.jsames.2015.12.010.

Nishenko, S. P. and Singh, S. K. The Acapulco-Ometepec, Mexico, earthquakes of 1907-1982: Evidence for a variable recurrence history. *Bulletin of the Seismological Society of America*, 77(4): 1359–1367, 1987. doi: 10.1785/BSSA0770041359.

NOAA National Geophysical Data Center. ETOPO1 1 Arc-Minute Global Relief Model, 2009. doi: 10.7289/V5C8276M.

Okal, E. A. A re-evaluation of the great Aleutian and Chilean earthquakes of 1906 August 17. *Geophysical Journal International*, 161(2):268–282, 2005. doi: 10.1111/j.1365-246x.2005.02582.x.

Okal, E. A. and Borrero, J. C. The ‘tsunami earthquake’ of 1932 June 22 in Manzanillo, Mexico: seismological study and tsunami simulations: The 1932 June Mexican tsunami earthquake. *Geophysical Journal International*, 187(3):1443–1459, 2011. doi: 10.1111/j.1365-246x.2011.05199.x.

Okal, E. A. and Hartnady, C. J. The South Sandwich Islands Earthquake of 27 June 1929: Seismological Study and Inference on Tsunami Risk for the South Atlantic. *South African Journal of Geology*, 112(3–4):359–370, 2009. doi: 10.2113/gssajg.112.3-4.359.

Okal, E. A., Synolakis, C. E., and Kalligeris, N. Tsunami Simulations for Regional Sources in the South China and Adjoining Seas.

Pure and Applied Geophysics, 168(6–7):1153–1173, 2010. doi: 10.1007/s00024-010-0230-x.

Okal, E. A., Borrero, J. C., and Chagué-Goff, C. Tsunamigenic predecessors to the 2009 Samoa earthquake. *Earth-Science Reviews*, 107(1–2):128–140, 2011. doi: 10.1016/j.earscirev.2010.12.007.

Oliver, J. and Murphy, L. WWNSS: Seismology's Global Network of Observing Stations. *Science*, 174(4006):254–261, 1971. doi: 10.1126/science.174.4006.254.

Peterson, J. R. and Hutt, C. R. World-Wide Standardized Seismograph Network: a data users guide. U.S. Geological Survey Open-File Report 1218, 2014. doi: 10.3133/ofr20141218.

Petroy, D. E. and Wiens, D. A. Historical seismicity and implications for diffuse plate convergence in the northeast Indian Ocean. *Journal of Geophysical Research: Solid Earth*, 94(B9): 12301–12319, 1989. doi: 10.1029/JB094iB09p12301.

Poupinet, G. On the relation between P-wave travel time residuals and the age of continental plates. *Earth and Planetary Science Letters*, 43(1):149–161, 1979. doi: 10.1016/0012-821X(79)90163-8.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024. https://www.R-project.org/.

Sambridge, M. Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space. *Geophysical Journal International*, 138(2):479—494, 1999. doi: 10.1046/j.1365-246x.1999.00876.x.

Schweitzer, J. HYPOSAT – An Enhanced Routine to Locate Seismic Events. *Pure and Applied Geophysics*, 158(1):277–289, 2001. doi: 10.1007/PL00001160.

Storchak, D. A., Di Giacomo, D., Engdahl, E. R., Harris, J., Bondár, I., Lee, W. H. K., Bormann, P., and Villaseñor, A. The ISC-GEM Global Instrumental Earthquake Catalogue (1900–2009): Introduction. *Physics of the Earth and Planetary Interiors*, 239:48–63, 2015. doi: 10.1016/j.pepi.2014.06.009.

Storchak, D. A., Harris, J., Brown, L., Lieser, K., Shumba, B., Verney, R., Di Giacomo, D., and Korger, E. I. M. Rebuild of the Bulletin of the International Seismological Centre (ISC), part 1: 1964–1979. *Geoscience Letters*, 4:32, 2017. doi: 10.1186/s40562-017-0098-z.

Storchak, D. A., Harris, J., Brown, L., Lieser, K., Shumba, B. T., and Di Giacomo, D. Rebuild of the Bulletin of the International Seismological Centre (ISC)—part 2: 1980–2010. *Geoscience Letters*, 7(1):18, 2020. doi: 10.1186/s40562-020-00164-6.

Suleiman, A. S. and Doser, D. I. The seismicity, seismotectonics and earthquake hazards of Libya, with detailed analysis of the 1935 April 19, M = 7.1 earthquake sequence. *Geophysical Journal International*, 120(2):312–322, 1995. doi: 10.1111/j.1365-246x.1995.tb01820.x.

Tape, C. and Lomax, A. Aftershock Regions of Aleutian-Alaska Megathrust Earthquakes, 1938–2021. *Journal of Geophysical Research: Solid Earth*, 127(7), 2022. doi: 10.1029/2022jb024336.

Tape, C., Lomax, A., Silwal, V., Agnew, J. D., and Brettschneider, B. The 1904 Ms 7.3 Earthquake in Central Alaska. *Bulletin of the Seismological Society of America*, 107(3):1147–1174, 2017. doi: 10.1785/0120160178.

Udias, A. and Stauder, W. The Jesuit Contribution to Seismology. *Seismological Research Letters*, 67(3):10–19, 1996. doi: 10.1785/gssrl.67.3.10.

Wessel, P., Smith, W. H. F., Scharroo, R., Luis, J., and Wobbe, F. Generic Mapping Tools: Improved Version Released. *Eos, Transactions American Geophysical Union*, 94(45):409–410, 2013. doi: 10.1002/2013eo450001.

Wiechert, E. Ein astatisches Pendel hoher Empfindlichkeit zur mechanischen Registrierung von Erdbeben. *Phys. Zeit.*, 4: 821–829, 1903.

Wielandt, E. Seismometry. In Lee, W., Kanamori, H., Jennings, J., and Kisslinger, C., editors, *International Handbook of Earthquake and Engineering Seismology*, volume A, chapter 18, pages 283–304. Academic Press, San Diego, 2002.

Wysession, M. E., Okal, E. A., and Miller, K. L. Intraplate seismicity of the Pacific Basin, 1913–1988. *Pure and Applied Geophysics*, 135(2):261–359, 1991. doi: 10.1007/BF00880241.

Zeiler, C. and Velasco, A. A. Seismogram Picking Error from Analyst Review (SPEAR): Single-Analyst and Institution Analysis. *Bulletin of the Seismological Society of America*, 99(5):2759–2770, 2009. doi: 10.1785/0120080131.