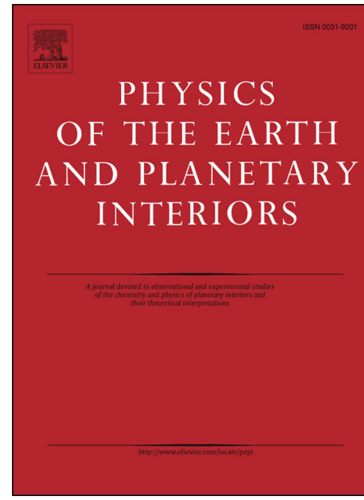


ISC-GEM: Global Instrumental Earthquake Catalogue (1900-2009), I. Data collection from early instrumental seismological bulletins

Domenico Di Giacomo, James Harris, Antonio Villaseñor, Dmitry A. Storchak, E. Robert Engdahl, William H.K. Leethe Data Entry Team

PII: S0031-9201(14)00148-4
DOI: http://dx.doi.org/10.1016/j.pepi.2014.06.003
Reference: PEPI 5751

To appear in: *Physics of the Earth and Planetary Interiors*

Received Date: 2 August 2013
Revised Date: 19 February 2014
Accepted Date: 2 June 2014

ISC-GEM: Global Instrumental Earthquake Catalogue (1900-2009), I. Data collection from early instrumental seismological bulletins

Domenico Di Giacomo[1], James Harris[1], Antonio Villaseñor[2], Dmitry A. Storchak[1], E. Robert Engdahl[3], William H. K. Lee[4], and the Data Entry Team[1]

1) *International Seismological Centre, Thatcham, UK*
2) *Institute of Earth Sciences Jaume Almera, ICTJA-CSIS, Barcelona, Spain*
3) *University of Colorado, Boulder, USA*
4) *862 Richardson Court, Palo Alto, CA 94303, USA*

ABSTRACT

In order to produce a new global reference earthquake catalogue based on instrumental data covering the last 100+ years of global earthquakes, we collected, digitized and processed an unprecedented amount of printed early instrumental seismological bulletins with fundamental parametric data for relocating and reassessing the magnitude of earthquakes that occurred in the period between 1904 and 1970. This effort was necessary in order to produce an earthquake catalogue with locations and magnitudes as homogeneous as possible. The parametric data obtained and processed during this work fills a large gap in electronic bulletin data availability. This new dataset complements the data publicly available in the International Seismological Centre (ISC) Bulletin starting in 1964. With respect to the amplitude-period data necessary to re-compute magnitude, we searched through the global collection of printed bulletins stored at the ISC and entered relevant station parametric data into the database. As a result, over 110,000 surface and body-wave amplitude-period pairs for re-computing standard magnitudes M_S and m_b were added to the ISC database. To facilitate earthquake relocation, different sources have been used to

retrieve body-wave arrival times. These were entered into the database using optical character recognition methods (International Seismological Summary, 1918-1959) or manually (e.g., British Association for the Advancement of Science, 1913-1917). In total, ~1,000,000 phase arrival times were added to the ISC database for large earthquakes that occurred in the time interval 1904-1970. The selection of earthquakes for which data was added depends on time period and magnitude: for the early years of last century (until 1917) only very large earthquakes were selected for processing ($M \geq 7.5$), whereas in the periods 1918-1959 and 1960-2009 the magnitude thresholds are 6.25 and 5.5, respectively. Such a selection was mainly dictated by limitations in time and funding. Although the newly available parametric data is only a subset of the station data available in the printed bulletins, its electronic availability will be important for any future study of earthquakes that occurred during the early instrumental period.

1. INTRODUCTION

The instrumental monitoring of earthquakes on a global scale began more than 100 years ago. Since then seismologists around the world store and exchange the results of standard observational seismological practice (e.g., picking of arrival times, amplitude-period measurements, etc.) or more complex waveform analyses (e.g., moment tensor inversion) in seismological bulletins/catalogues. Such products contain fundamental parametric data characterizing the seismicity of a region or the entire globe. However, instrumental parametric data (earthquake source related such as hypocentres, magnitudes, moment tensor solutions, or seismic station related such as phase arrival times, amplitudes and periods of different seismic waves) were available in computer-readable format, with a few exceptions, for only the last 50 years. A large volume of earthquake data starting with 1960 is publicly available from the International Seismological Centre (ISC, www.isc.ac.uk), where different types of parametric data are collected, processed and reviewed in order to produce the definitive summary of the

Earth's seismicity. In this respect, the ISC database can be considered the most comprehensive repository of instrumental earthquake data on a global scale thanks to the collaborations of the ISC members (www.isc.ac.uk/members/) and reporting agencies (www.isc.ac.uk/iscbulletin/agencies/). However, for the time period before 1964, only the basic earthquake parameters such as location and magnitude retrieved from different catalogues (e.g., Gutenberg and Richter, 1954; Rothé, 1969; Båth and Duda, 1979; Abe, 1981; Abe and Noguchi, 1983a,b) were available in the electronic format. A bulletin differs from a catalogue by containing the station data used in the computation of the earthquake location and, whenever possible, the magnitude. Therefore the lack of digitally available station parametric data (e.g., phase picks and amplitude-period measurements) represented a huge limitation for the task of re-computing hypocentres and magnitudes using modern techniques and standards for the vast majority of global earthquakes before 1960.

Before the era of modern computer-based bulletin production, individual observatories and research institutions around the world stored seismic station reports in printed or hand-written seismological bulletins. Many such bulletins have been collected by the ISS and ISC over the years and represent a fundamental source of seismological data for earthquakes that have occurred from the beginning of 20th century until the 1960s and to a lesser extent into the 1970s. This allowed us to digitize and process station parametric data from a multitude of printed bulletins covering approximately 70 years of instrumental seismology.

The availability of data in the ISC database at the start of the project is summarized in Table 1, where we distinguish between phase arrival times and amplitude-period data essential for relocation and magnitude re-computation, respectively. Although the ISC Bulletin starts in 1964, phase data covering the period 1960-1963 was already available in the ISC database based on the work of Villaseñor and Engdahl (2007). Both phase and amplitude-period data were completely missing in the ISC database for earthquakes before 1960, whereas only amplitude-period data needed to be retrieved from printed seismological bulletins for events in 1960 and partially also in the1970s.

Due to time and resource limitations however, we searched for station data relevant to selected earthquakes according to the following cut-off magnitudes for different time periods:

1. 1900-1917: $M_S \geq 7.5$ worldwide plus a selection of smaller shallow events in stable continental areas with M_S between 6.5 and 7.5;
2. 1918-1959: $M_S \geq 6¼$;
3. 1960-2009: $M_S \geq 5.5$.

For the initial selection of earthquakes before 1976 we used all types of magnitude estimates (many from individual stations), compiled from all available sources of location and magnitude, in an expanded version of the Centennial Catalogue (Engdahl and Villaseñor, 2002). M_S was our reference magnitude type before the introduction of the Global Centroid Moment Tensor database (GCMT, www.globalcmt.org), and when that was not available we used other magnitude estimates (including magnitudes of unknown type). From 1976 onwards we used an event selection cut-off moment magnitude of M_W 5.6 (M_S 5.5) from GCMT when available, or M_W proxy values from available M_S or m_b. In the latter case we used the Bormann *et al.* (2009) relationships, and lowered the cut-off magnitude down to M_W 5.5 during 1964-1977 in light of the larger uncertainty estimation in that period. Thus we avoided overlooking earthquakes with $M_W \geq 5.6$ in the modern period.

In Storchak *et al.* (2013) a brief overview of the data added in the early instrumental period was given, concerning both arrival time and amplitude-period measurements. Here we describe in detail the printed sources considered for the production of the ISC-GEM catalogue and the procedures used to digitize the data stored in the printed bulletins. This description will be useful to the users of the ISC-GEM catalogue in obtaining a better overview of the underlying data used for relocation (see Bondár *et al.*, 2014) and magnitude re-computation (see Di Giacomo *et al.*, 2014) in the early instrumental period. We also point out some limitations of the data collection, which, in some case, may have affected the quality of some earthquake parameter determinations.

The newly digitized data fills a significant time gap where no parametric data were previously available in the ISC database for relocating earthquakes (before 1960) and for re-computing standard magnitudes such as M_S and m_b (before 1971, except the WWSSN short-period body-wave amplitude data available from 1964). For the period 1960-2009, station data can be found in the ISC Bulletin via the ISC website (www.isc.ac.uk). We also give a summary of the newly available digital parametric data that effectively doubles the overall time length of the ISC Bulletin. Finally, we briefly discuss possible areas of improvement.

Table 1: Summary of the station parametric data availability in ISC database before the beginning of this work.

Period	1900-1917	1918-1959	1960-1963	1964-1970	1971-1977	1978-2009
Body-wave arrival times	Not available	Not available	Available	Available	Available	Available
Surface wave amplitudes and periods	Not available	Not available	Not available	Not available	Available, with the exception of certain important stations	Available
Short-period body-wave amplitudes and periods	Not available	Not available	Not available	WWSSN network available	Available	Available

2. PHASE DATA COLLECTION (1900-1959)

To re-compute locations of the ISC-GEM earthquakes that occurred before 1960, we collected arrival times of seismic phases from different sources that were only available in either printed or manuscript form. These have been converted to a digital form using one of two methods. For good quality printed bulletins with formats consistent over a long period of time we have used optical character recognition (OCR) techniques (see Engdahl and Villaseñor, 2002). For all other sources we entered the data manually. Once the data were available in a computer-readable format, they were inserted into the ISC database. Below we summarize various phase data sources digitized during this work in the order of their time coverage. We limit the description of the data collection to 1959, since from 1960 phase data was already available in the ISC database (Table 1). Appendix A summarizes the volume of data collected for each station. The station data comes from a variety of instruments (from early Milne seismographs to Wiechert, Galitzin, Bosh-Omori, Mainka, Milne-Shaw) and agencies/bulletins, which are summarized by Storchak *et al.* (2014).

2.1. Gutenberg notepads (1904-1912) and International Seismological Association (ISA) bulletins (1904-1907)

For the period 1904-1912 a fundamental source of seismic phase arrival times is the Gutenberg notepads. This collection is available on microfiche (Goodstein *et al.*, 1980) and also as a collection of scanned images kindly provided by Professor Katsuyuki Abe. Where available, we have used the scanned images because these are of slightly higher quality than the microfiches.

The Gutenberg notepads are hand-written (examples are shown in Figure 1), but both the scanned images and microfiches are of poor quality, making it impossible to use the OCR. Therefore we entered P and S wave arrival time data manually, for a total of ~1200 phases.

For earthquakes that occurred during the period 1904-1907, where Gutenberg notepads lacked station arrival times, we also manually entered phase picks from scans of the International Seismological Association (ISA) bulletins (see Schweitzer and Lee, 2003, and references therein). As a result, for 67 large earthquakes selected in the period 1904-1912 we added ~1,900 body-wave arrival times for ~100 seismic stations from around the world.

2.2. Seismological Bulletin of the British Association for the Advancement of Science (BAAS), 1913-1917

During the period 1913-1917, the most useful source of associated phase arrival times are the seismological bulletins of the British Association for the Advancement of Science (BAAS). These bulletins are the predecessors of the International Seismological Summary (ISS), and are available in good quality printed form (see Figure 2). However, the bulletin format changes sufficiently from year to year to make it difficult to use the OCR methods. Therefore, we opted to enter the phase arrival time data of P, S and supplementary phases manually. The number of earthquakes processed for this period is 45, and the number of body-wave arrival times added is approximately 3,800 for ~100 seismic stations from around the world.

2.3. International Seismological Summary (ISS), 1918-1959

The largest source of station arrival times for earthquakes that occurred during the period 1918-1963 is the International Seismological Summary (ISS), which was the predecessor of the ISC Bulletin. In the ISS bulletins the arrival times were not just compiled but also reviewed, and besides arrival times (and residuals) of P and S phases often included arrivals of others secondary phases such as depth phases, PP, SS etc.

The ISS bulletins are available in a fairly stable printed form (scans available at http://storing.ingv.it/ISS/index.html). Some of the data was already converted to digital form before the beginning of this work:

- All earthquakes in the ISS for the period 1960-1963 had already been relocated by Villaseñor and Engdahl (2007) and the phase data were also available in the ISC database;
- All earthquakes in the ISS bulletins with $M_S \geq 7.0$ were processed by Engdahl and Villaseñor (2002) to produce the Centennial Catalogue; the corresponding phase data were also available as a Fixed Format Bulletin (FFB) and are now included in the ISC database;
- A digital file containing hypocentre and phase data for most of the earthquakes during 1918-1942 was available thanks to the efforts of P. Willmore and E. Arnold (the first two Directors of the ISC), who had arranged for the ISS bulletins to be typed at a professional data preparation bureau based near Shannon airport in Ireland. Unfortunately, the ISC funds at the time were too short to allow this work to continue beyond data year 1942. This file (hereafter referred to as the "Shannon tape") was created by manually entering the observations from the ISS bulletins onto punch cards.

We nevertheless needed to enter the phase data for some earthquakes missing from the Shannon tape (1918-1942) as well as for the majority of earthquakes with $6.25 \leq M_S < 7.0$ in the interval 1943-1959. For phase data relevant to these earthquakes, we have therefore used OCR to facilitate the conversion of the ISS scans into digital form (text files). The application of the OCR technique was possible since the quality of the printed bulletins is good and the phase data are listed in a homogeneous tabular form. From the text files generated by the OCR technique, we proceeded to edit in a semi-automatic way a large number of ISS pages (for large earthquakes the ISS can list phase data from hundreds of stations in several pages) that, if manually processed, would require years and significant resources. The editing of the OCR output is done in an interactive way in order to fix errors/typos from the OCR and to check the validity of the station data (e.g., checks on station names, minutes of the arrival times of P and S waves, phase names, etc.) as well as the reported location parameters.

After including data for the earthquakes missing in the Shannon tape and the earthquakes with $6.25 \leq M_S < 7.0$ that were not relocated in the Centennial Catalogue, the total number of newly available ISS phase picks for the period 1918-1959 exceeds 620,000 (from ~850 seismic stations).

2.4. Japan Meteorological Agency (JMA) Bulletin, 1923-1970

The JMA has digitized bulletin data of seismic stations located in the Japanese region for the early instrumental period 1923-1970. These data were then kindly made available to the ISC by Professor Nobuo Hamada. This early JMA bulletin was in a computer-readable format and it was, therefore, possible to parse it automatically into the ISC database. After converting the JST time (+9h) to GMT, we observed identical arrival times for those stations also listed in the ISS bulletin.

The JMA network was quite dense for that period and a large volume of the data are related to the local/regional seismicity of Japan that could not be considered in the context of the ISC-GEM catalogue project. However, many phase arrival times picks (about 270,000 from about 230 stations) were used in the ISC-GEM relocation procedure.

2.5. Overview of the phase data collection (1900-1959)

An overview of the annual numbers of body-wave arrival times available for relocation and obtained from the data sources just described is shown in Figure 3. Clearly, there is a significant and steady increase in the data from 1918 (beginning of the ISS bulletins) until the late 1950s, with the exception of the Second World War years. Figure 3 also shows the additional phase data entered from seismological station/network bulletins as described in the next section.

Figure 4 illustrates the station distribution for which data were added from one of the sources described previously and for earthquakes that occurred before 1960. The stations symbols are colour-coded according to the number of arrivals added. For more details, Appendix A lists, for each station, the day of the first and last arrival time added, as well as the total number of phase picks and number of earthquakes for which the station is reported either in the ISS or the other phase data sources.

3. AMPLITUDE-PERIOD DATA COLLECTION (1900-1970)

A major drawback with the ISS and other data sources is the lack of the basic measurement data (e.g., amplitude, period and component information) for seismic phases useful for re-computing the standard magnitude scales such as M_S and m_b. In order to re-calculate magnitudes for the relocated hypocentres, we needed to retrieve all necessary information from the early instrumental seismological bulletins (either individual stations or network bulletins). This was the core of the work for this project and also the most time consuming part. Indeed, due to a large variability in formats, physical state and readability of station bulletins, the OCR technique that was used for the ISS bulletin could not have been applied. Therefore, the data for magnitude re-computation needed to be entered manually. To do this task up to five data entry officers worked at the ISC over about 1.5 years. Dedicated interactive web browser interfaces with underlying checks and database entry programs were developed to increase the speed and accuracy of data entry and limit the amount of data inserted manually. The next subsections summarize various steps we followed when collecting and digitizing the amplitude-period data for magnitude re-computation.

3.1. Organization of seismic bulletins

As mentioned earlier, the ISS and then the ISC has collected paper bulletins from various stations/institutions over many years and stored them in chronological order in cardboard boxes that moved with the ISS/ISC from one storage room to another. As a first step, we reorganized the bulletins by country and observatory in chronological order. Then we inserted into the database the essential information identifying each booklet (e.g., institution and publication names, station, year, town, country, etc.) to serve as an inventory of the early instrumental seismological bulletins available at the ISC (henceforth called Bulletin Registry). The parameters of each booklet were entered into the database in order to allow a unique link between the paper bulletins and the data entered into the database. The large ISC collection of early instrumental seismological bulletins was complemented by selected bulletins from USGS/Berkeley, scanned bulletins available from Schweitzer and Lee (2003), as well as scanned material provided by the Institute of Seismology in Bishkek, Kyrgyzstan and by the Geophysical Survey of Russian Academy of Sciences in Obninsk, Russia.

The Bulletin Registry (until the end of 1970) now includes over 15,000 individual booklets from ~290 institutions in 80 countries. The large number of these booklets is explained by the common practice of many institutions to publish monthly bulletins for long periods of time (e.g., for the early instrumental station at Riverview College in Australia, the inventory is composed of 345 booklets between 1909 and 1968).

After compiling the Bulletin Registry, we proceeded to select the more useful institutions/stations to fulfill the main aim of this task, which was to retrieve surface wave amplitude and period data for M_S re-computation of earthquakes covering the period 1900-1970. Therefore, priority was given to observatories providing reliable and systematic surface wave amplitude and period measurements for earthquakes recorded at teleseismic distances. Therefore, the bulletins were grouped according to their importance for M_S determination:

1. Bulletins of primary importance (red in Figure 5);

2. Bulletins that could be helpful but were to be used only if time and resources would permit (blue in Figure 5);

3. Bulletins that were not useful for earthquake magnitude determination (green in Figure 5). The

Bulletin Registry time coverage shown in Figure 5 does not guarantee that data for earthquakes in a given period were available in a specific bulletin, though the Bulletin Registry is nevertheless useful to indicate the data gaps for a station/institution in our collection. Usually such gaps have occurred due to station operational issues. At times, tough, this was because the bulletins were not made available to the ISC.

3.2. Entering parametric station records into the database

Given the goals of the ISC-GEM catalogue and the amount of time and resources available, not all the data contained in the station bulletins could be processed. Instead, we searched for data related to the earthquakes selected according to the cut-off magnitudes previously described. We followed the simple criteria that a *reading* (in the ISC jargon a *reading* groups all the parametric data from a single station associated to a specific earthquake and reported by the same agency) must include amplitude-period measurements for magnitude re-computation. An example of a *reading* selected is shown on Figure 6a for station Göttingen (GTT, Germany) for the well-known 1906 San Francisco earthquake. The *reading* is of good quality and with different measurements of surface wave amplitude-period pairs. The same data is shown in Figure 6b after entry into the database, with the amplitude related information (amplitude value and its period, component, amplitude unit and type) ready for magnitude re-computation. We aimed at replicating the information contained in the bulletin as much as possible, and therefore in addition to the amplitude related data we also included the phase arrival times (annual amount of the body-wave arrival time added manually from seismological bulletins was shown in Figure 3).

Searching by page for useful data, such a procedure was applied to thousands of booklets from observatories/institutions around the world. Over 110,000 phases in the period 1904-1970 with amplitude and period measurements for magnitude re-computation were added. Of those ~100,000 are surface wave measurements (~80,000 from horizontal components, ~20,000 from the vertical component), and the remaining P-wave measurements on the vertical component. Similarly to Appendix

A, Appendix B summarizes the amount of amplitude data added for each station to fill the gaps before 1971 of amplitude data availability in the ISC database. Featured in the earliest reports is station Göttingen, where the Geophysical Institute first began the operation of the Wiechert astatic horizontal seismographs (Wiechert, 1899). Starting with year 1904 the Göttingen Geophysical Institute produced a very detailed bulletin with amplitude and period data usable for re-computing M_S in a modern instrumental sense. Similar observational seismological practice slowly spread to other observatories so that an increasing number of stations also provided the necessary information to re-compute M_S for large earthquakes that occurred over 100 years ago. Based on the minimum and maximum day listed in Appendix B it is noticeable how few stations operated almost continuously during the early instrumental period. Such notable stations are Uppsala (UPP, Sweden), which is probably the best example of an excellent seismic observatory during the period 1906-1970, and two stations belonging to the Jesuit seismological network (Udías and Stauder, 1996), Riverview (RIV, Australia) and La Paz (LPZ, Observatorio San Calixto, Bolivia). The latter two are of particular importance considering the lack of stations in the southern hemisphere in the early instrumental period. Most good quality European stations as well as many stations in the former Soviet Union territory have gaps, mainly due to the impact of both World Wars I and II. Large gaps, both in space and time, are also present in the southern hemisphere, in North America and in Africa. Nevertheless, the time-space station distribution of available data usually provided satisfactory azimuthal coverage for the determination of magnitudes for over 4,600 relocated earthquakes between 1904 and 1970 (see also Di Giacomo *et al.*, 2014).

Figure 7 shows the annual distribution of manually entered amplitude data useful for magnitude re-computation. As in Figure 3, fluctuations in the distributions are observed over the years, and a clear decrease in the amount of data is related to the impact of World War II in the 1940s. Insufficient reliable parametric station data were found for the standard relocation and magnitude estimation during the period 1900-1903. Hence, the hypocentre parameters of earthquakes during this period were adopted from Abe and Noguchi (1983a,b).

Finally, additional surface wave measurements between 1971 and 1977 were added to the ISC database from high quality stations in the former Soviet Union, and from UPP and KIR (Kiruna, Sweden). These data have been added to remove gaps in surface wave measurements for long operating stations and to complement the surface wave measurements stored in the ISC database that were not previously used to compute M_S. In total, ~30,000 surface wave amplitude-period data from over 70 stations of the former Soviet Union were added by parsing electronic bulletins kindly provided by the Geophysical Survey of the Russian Academy of Sciences, Obninsk, Russia, and ~4000 surface wave amplitudes-periods were manually added for each of the UPP and KIR stations as a final task for the data entry team.

4. CONCLUSIONS

In order to deliver an updated catalogue based on instrumental data covering the last 100+ years of global earthquakes, we processed and digitized an unprecedented amount of phase and amplitude data from various collections of printed bulletins over about 1.5 years. This task was necessary for retrieving the basic data to run current techniques for event relocation and magnitude re-computation. Therefore, we processed individual and network station seismological bulletins to retrieve the necessary information for re-computing magnitude and various other sources of phase data (ISS, BAAS, Gutenberg notepads, etc.) for the relocation. Not all the data available in the printed bulletins was digitized but, within the time and funding capabilities, only the data relevant to earthquakes with magnitude $\geq$ 7.5 until 1917, $\geq$ 6.25 between 1918 and 1959, and > 5.5 from 1960.

Table 2 summarizes the amount of parametric data already available or added to the ISC database during this project to deliver the new global reference ISC-GEM earthquake catalogue. This project has contributed a large amount of phase arrival times (until the end of 1959) and amplitude-period data (until the end of 1977), which is now electronically available for periods where no digital data was available before (with a few exceptions).

Not all data sources available as paper bulletins could be processed. In this respect, an important example is represented by the bulletins of the Bureau Central International de Séismologie (BCIS, 1925-

1971) produced in Strasbourg, France. Similar to ISS, the BCIS bulletins contain arrival times from stations around the globe. The paper quality and format of the BCIS bulletins varies with time and the information appears less detailed than the ISS (e.g., the BCIS does not report time residuals and rarely reports additional arrivals beside P and S arrivals). Although the scanning of the BCIS has been made by other institutions, no basic parameters are available in digital form beyond the digital images of the bulletin pages. Based on our experience with the ISS, processing the BCIS scanned images would require a major effort and take a long period of time. However, from spot checks on a few issues of the BCIS we noticed that for earthquakes listed both in the ISS and BCIS the bulletins share a very similar set of stations with usually only a few stations missing in one or the other bulletin. Therefore, it is questionable whether the improvement in locations can be made by adding BCIS arrivals that are missing from the ISS. Future funding and resources may permit us to start digitizing the BCIS bulletins, or at least the arrivals missing from the ISS.

Finally, we remark that the ISC collection of station bulletins considered for digitizing amplitude-period measurements is not complete and that also within the set of printed bulletins available at the ISC we did not process many of them due to time limitations. In particular, many Central European stations were omitted (like Strasbourg, France) or only partially processed (like Moxa, Germany). Therefore, the number of stations contributing to magnitude (especially for M_S) could be larger, and for those earthquakes with only a limited number of station magnitudes (three to five) the event magnitude may not be well constrained. However, such cases may be of importance for the ISC-GEM catalogue only when the standard magnitudes (either M_S or m_b) were used as basis for M_W proxy computation (for more details see Di Giacomo *et al.*, 2014). Nevertheless, the data added in this project is a significant step forward towards improving any future study for earthquakes that occurred in the early instrumental period.

Table 2: Summary of the overall amount of station arrival (amplitude) data gathered from different sources that were used to produce the ISC-GEM catalogue, including the modern period (ISC Bulletin). The cells in grey

represent the data already available in the ISC database before this work.

<table>
<tr><th></th><th>1904-1912</th><th>1913-1917</th><th>1918-1922</th><th>1923-1942</th><th>1943-1959</th><th>1960-1963</th><th>1964-1970</th><th>1971-1977</th><th>1978-2009</th></tr>
<tr><td>Seismological bulletins</td><td colspan="7">~270,000 (~110,000 amplitudes)</td><td>~38,000 amplitudes</td><td></td></tr>
<tr><td>Gutenberg notepads and ISA bulletins</td><td>~1,900</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>BAAS</td><td></td><td>~3,800</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>ISS (including Shannon tape)</td><td></td><td></td><td colspan="3">~620,000</td><td></td><td></td><td></td><td></td></tr>
<tr><td>JMA Bulletin</td><td></td><td></td><td></td><td colspan="4">~270,000</td><td></td><td></td></tr>
<tr><td>ISS as from Villaseñor and Engdahl (2007)</td><td></td><td></td><td></td><td></td><td></td><td>~330,000</td><td></td><td></td><td></td></tr>
<tr><td>ISC Bulletin</td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="3">~11,000,000 (~2,500,000 amplitudes)</td></tr>
</table>

DATA AND RESOURCES

Most of the figures were drawn using Generic Mapping Tools (GMT, Wessel and Smith 1991) software.

ACKNOWLEDGEMENTS

Members of the Data Entry Team: Rebecca Verney, Natalia Safronova, Rosemary Wylie, Agne Baranauskaite, Jessica Wilson and Hepsi Simpson.

This work was supported by the GEM Foundation as one of the five GEM Global Hazard Components.

We wish to recognize the initiative of the late Edouard Arnold and Pat Willmore, former Directors of the ISC, for their efforts in organizing first the transfer of a large fraction of the ISS paper bulletins onto punch-cards and preparing the Shannon tape. We also wish to acknowledge the effort of Tom Boyd, Colorado School of Mines, USA, for his contribution in correcting the data on the Shannon tape.

We thank Robin Adams (ISC Honorary Seismologist) and David McGregor (Former ISC Deputy Director) for providing historical information on the operations in the early years of the ISC.

We acknowledge Nobuo Hamada and the Japan Meteorological Agency (JMA) for providing previously unavailable early instrumental (1923-1970) bulletins of earthquakes in Japan. We are grateful to Katsuyuki Abe for providing a digital copy of the Abe and Noguchi Catalogue and the scanned images of the Gutenberg's notepads. We are also grateful to Anna Berezina (Institute of Seismology, National

Academy of Sciences, Kyrgyzstan) for organizing the scanning of the original operator's notepads of the seismic station Frunze. We thank Irina Gabsatarova, Olga Kamenskaya, Vera Babkina and Raisa Mihailova (Geophysical Survey of the Russian Academy of Sciences, Obninsk, Russia) for providing the digital archive and scanning copies of the seismological bulletins of the Unified Network of Backbone Seismic Stations of the Former Soviet Union in 1971-1977.

We thank Siegfried Wendt (Universitaet Leipzig, Germany) for sending missing bulletins of the Collm Observatory, and Jina Gachechiladze (Ilia State University, Tbilisi, Georgia) and Arkady Aronov and Vladislav Aronov (Centre of Geophysical Monitoring of the National Academy of Sciences, Belarus) for providing information on positions of long-serving seismic stations. We thank Peggy Hellweg, Berkeley Seismological Laboratory, University of California, for arranging the transfer of their collection of seismic station bulletins to the U.S. Geological Survey.

REFERENCES

Abe, K., 1981. Magnitudes of large shallow earthquakes from 1904 to 1980, *Phys. Earth Planet. Int.*, 27, 72-92.

Abe, K., & Noguchi S., 1983a. Determination of magnitudes for large shallow earthquakes, 1898-1917, *Phys. Earth Planet. Inter.*, 32, 45-59.

Abe, K., & Noguchi S., 1983b. Revision of magnitudes of large shallow earthquakes, 1897-1912, *Phys. Earth Planet. Inter.*, 33, 1-11.

BAAS (1913-1917). British Association for the Advancement of Science, Seismological Committee, quarterly issues.

Båth, M., & Duda, S.J., 1979. Some aspects of global seismicity, *Tectonophysics*, 54, Tl-T8.

BCIS (1930-1971). Bureau Central International de Séismologie, monthly issues.

Bondár, I., Engdahl, E.R., Villaseñor, A., Harris, J. & Storchak, D.A., 2014. ISC-GEM: Global Instrumental Earthquake Catalogue (1900-2009), II. Location and Seismicity Patterns, *Phys. Earth Planet. Inter.*, under review.

Bormann, P., Liu, R., Xu, Z., Ren, K., Zhang, L., & Wendt, S., 2009. First application of the new IASPEI teleseismic magnitude standards to data of the China National Seismographic Network, *Bull. Seis. Soc. Am.*, 99, 1868–1891.

Di Giacomo, D., Bondár, I., Storchak, D.A., Engdahl, E.R., Bormann, P. & Harris, J., 2014. ISC-GEM: Global Instrumental Earthquake Catalogue (1900-2009), III. Re-computed M_S and m_b, proxy M_W, final magnitude composition and completeness assessment, *Phys. Earth Planet. Inter.*, under review.

Engdahl, E.R., & Villaseñor, A., 2002. Global seismicity: 1900-1999. in *International Handbook of Earthquake and Engineering Seismology*, Part A, 665-690, ed. Lee, W.H.K., Kanamori, H., Jennings, J. C., & Kisslinger, C., Academic Press, San Diego.

Goodstein, J.R., Kanamori, H., & Lee, W.H.K. (Editors), 1980. Seismology microfiche publications from the Caltech archives, *Bull. Seismol. Soc. Am.*, 70, 657-658.

Gutenberg, B. & Richter, C.F., 1954. *Seismicity of the Earth and Associated Phenomena*, 310 pp, Princeton Univ. Press, Princeton, N.J.

ISA (1904-1907). International Seismological Association, annual volumes.

ISS (1918-1963). International Seismological Summary, annual volumes.

Rothé, J.P., 1969. The seismicity of the Earth, 1953-1965, UNESCO, Paris.

Schweitzer, J., & Lee, W.H.K., 2003. Old seismic bulletins to 1920: A collective heritage from early seismologists, in *International Handbook of Earthquake and Engineering Seismology, Part B*, 1665-1723, ed. Lee, W.H.K., Kanamori, H., Jennings, P.C., & Kisslinger, C., Academic Press, Amsterdam.

Storchak, D.A., Di Giacomo, D., Bondár, I., Engdahl, E.R., Harris, J., Lee, W.H.K., Villaseñor, A., & Bormann, P., (2013). Public release of the ISC-GEM Global Instrumental Earthquake Catalogue (190-2009), *Seismol. Res. Lett.*, 84(5), 810-815, doi: 10.1785/0220130034.

Storchak, D.A., Di Giacomo, D., Engdahl, E.R., Harris, J., Bondár, I., Lee, W.H.K., Bormann, P., & Villaseñor, A., (2014). The ISC-GEM Global Instrumental Earthquake Catalogue (190-2009): Introduction, *Phys. Earth Planet. Inter.*, under review.

Udías, A., & Stauder, W. (1996). The Jesuit contribution to seismology, *Seismol. Res. Lett.*, 67, 10-19.

Villaseñor, A., & Engdahl, E.R., 2007. Systematic Relocation of Early Instrumental Seismicity: Earthquakes in the International Seismological Summary for 1960-1963, *Bull. Seismol. Soc. Am.*, 97(6), 1820-1832.

Wessel, P. & Smith, W. H. F., 1991. Free software helps map and display data, *EOS Trans. AGU*, 72, 441, 445-446.

Wiechert, E., 1899. Seismometrische Beobachtungen im Göttingen Geophysikalischen Institut, *Nachrichten von der Königl. Gesellschaft der Wissenschaften zu Göttingen*, Mathematisch-physikalische Klasse, 195-208.

Figure Captions

Figure 1: Example scans of the hand-written Gutenberg notepads: a) station data for an earthquake occurred the 1904-04-04 in Bulgaria and b) for two earthquakes that occurred in the Hindu-Kush region in October 1908. Digitizing such data is a problematic and time consuming effort.

Figure 2: Example scan of the BAAS bulletin: station data for an earthquake that occurred the 1913-05-30.

Figure 3: Annual number of arrival times of P-wave (red) and S-wave (blue) types before 1960 collected from different. The annual number of P-wave (yellow) and S-wave type (brown) manually added as part of the amplitude data entry task (described later) is also shown.

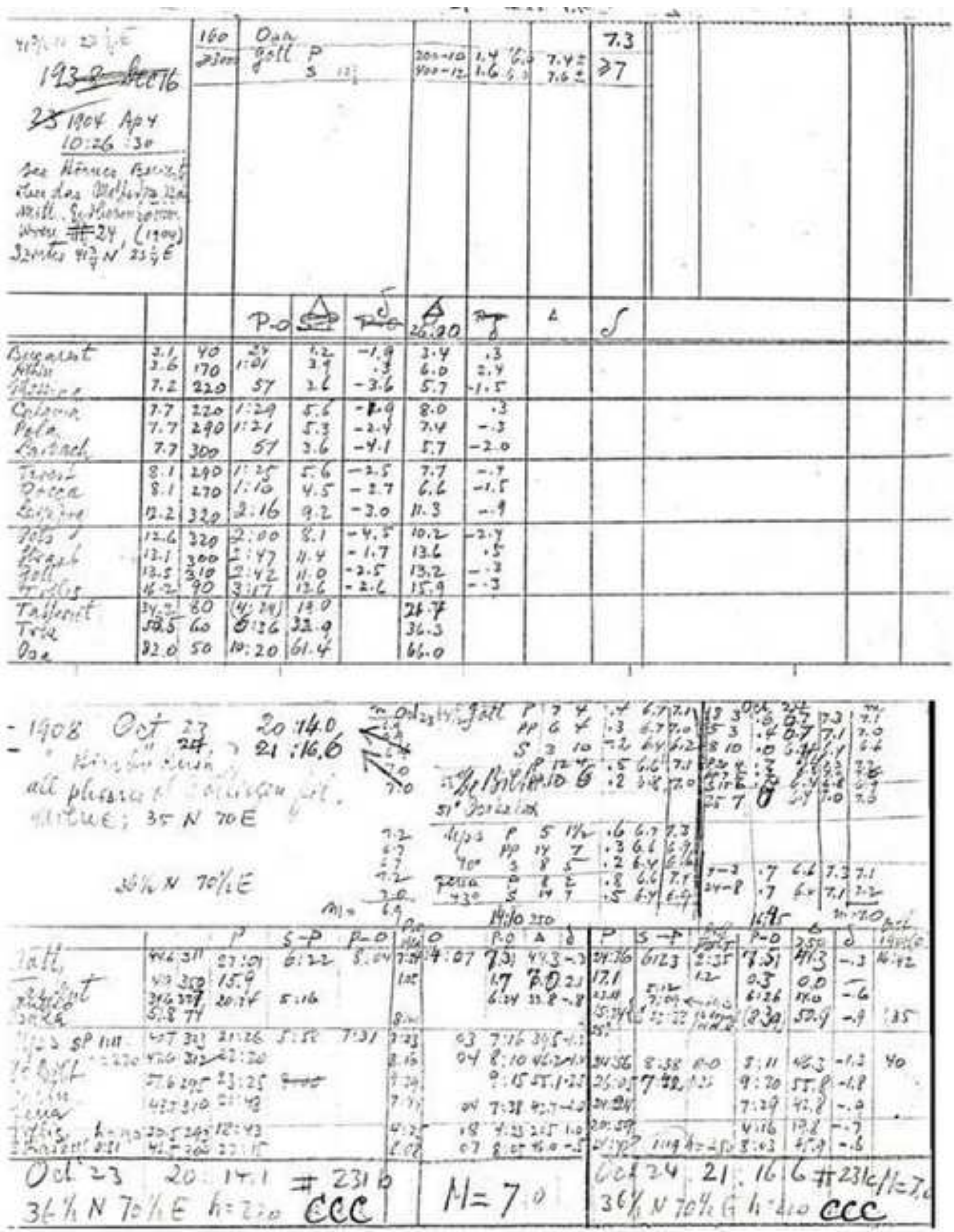
Figure 4: Distribution of the stations for which phase data were added before 1960 from the sources previously described. Each station is colour-coded by total number of arrivals added before 1960.

Figure 5: Bulletin Registry time coverage: important bulletins for magnitude re-computation (red), bulletins helpful if time and resources permit (blue) and not useful for magnitude re-computation (green). See text for details.

Figure 6: a) An example of the Göttingen station (GTT) bulletin with the *reading* of the 1906 San Francisco earthquake. The surface wave maxima and related periods after the body-wave arrivals are essential for M_S re-computation; b) the same *reading* after entering the data in the ISC database.

Figure 7: Annual number of surface wave (black) and vertical component P-wave (gray) amplitude-period measurements manually entered from individual/network seismological bulletins.

Figure(s)



a)

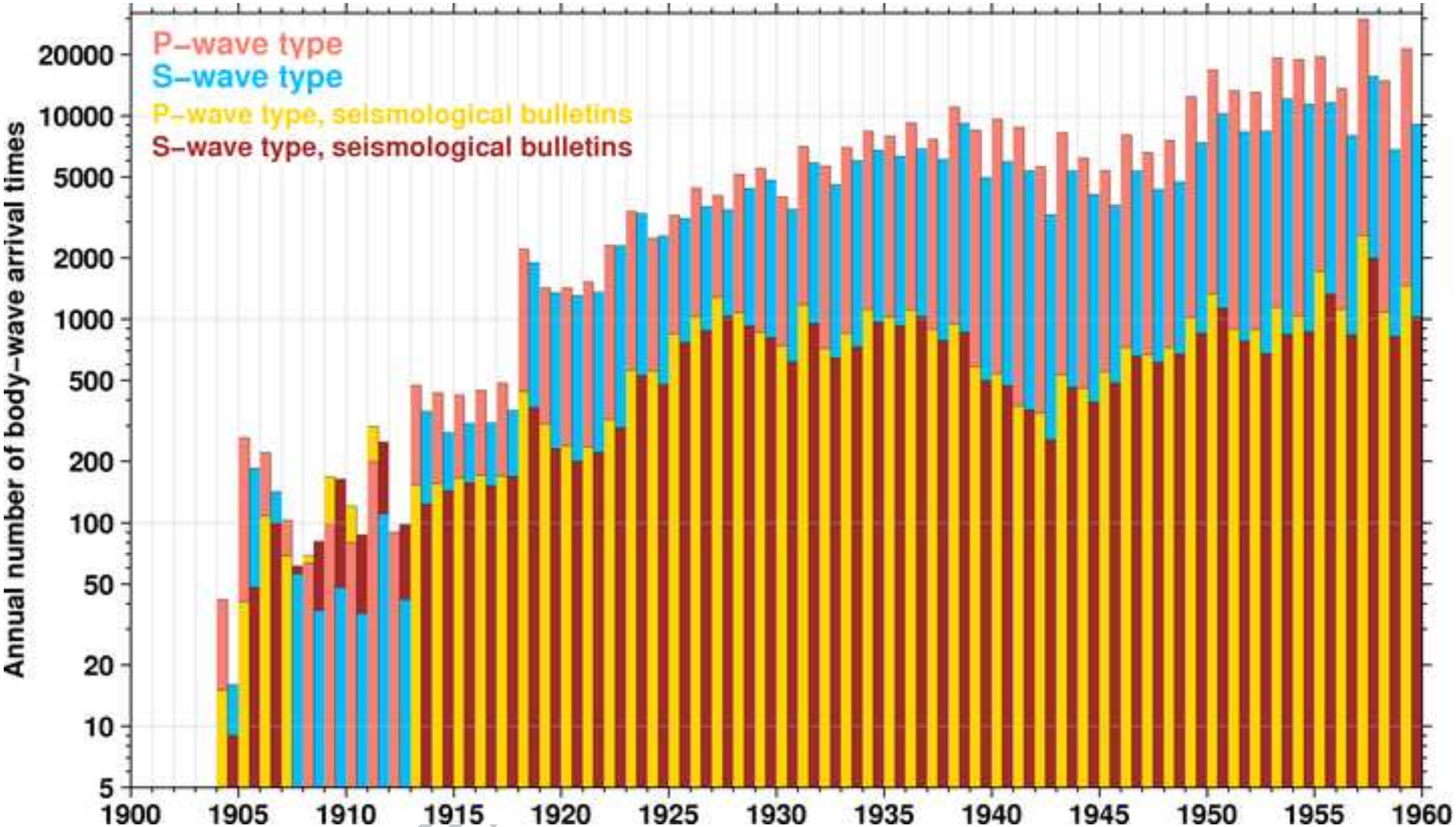
b)

1913. May 30d. 11h. 46m. 46s.

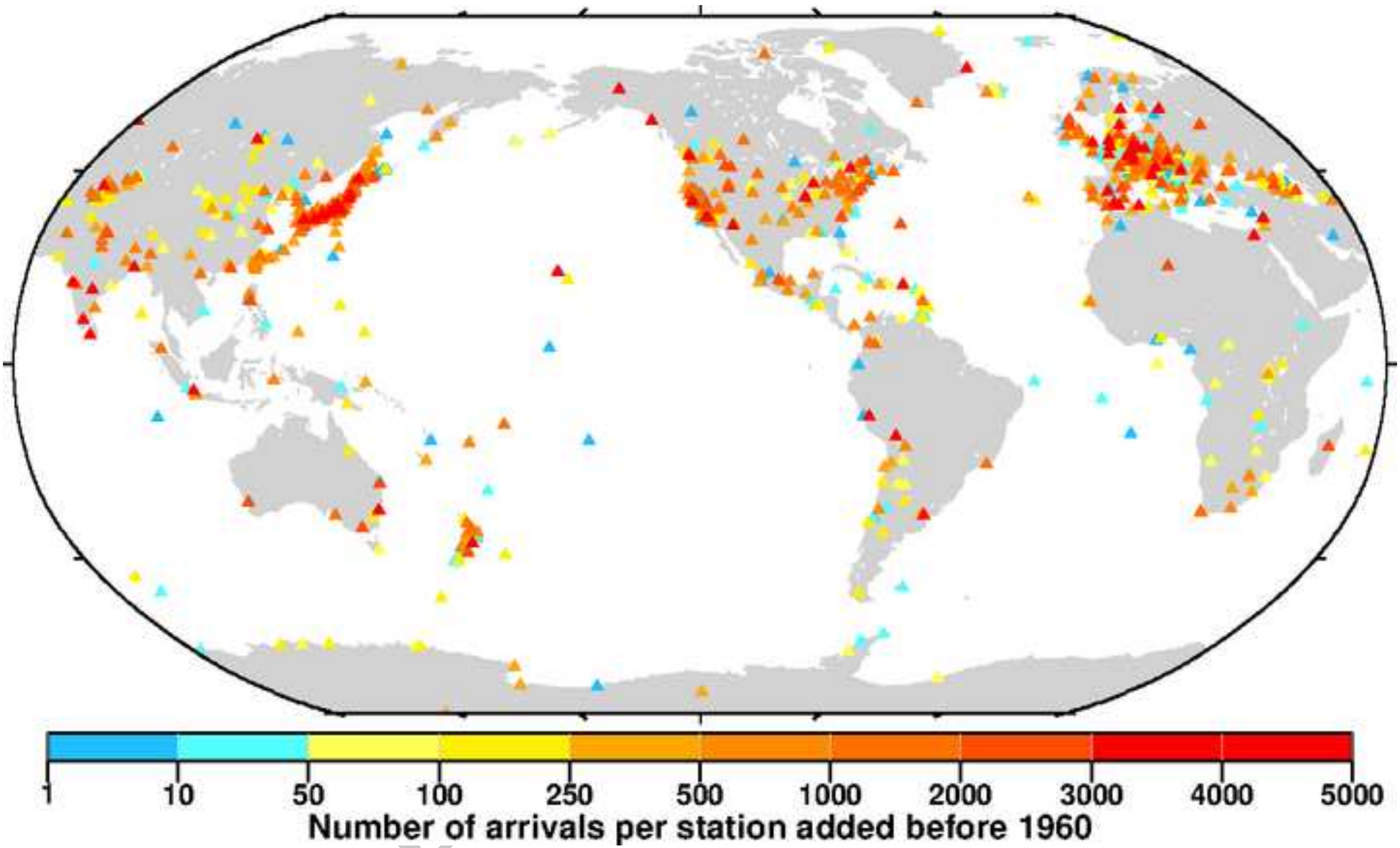
Adopted Origin: 5°·0s. 154°·0E.; Irkutsk, 1°·0N. 159°·6E.; Pulkovo, 6°·0S. 154°·0E.

Station and Component.		Machine.	Distance from Origin. °	Azimuth. °	P. H.	P. M.	P. S.	O–C. S.	S. H.	S. M.	S. S.	O–C. S.
Sydney	E.	M.	29·0	185	11	53	54	+50	11	58	18	+15
Riverview		—	29·0	185	11	53	12	+ 8	11	58	2	− 1
Adelaide	E.	M.	33·2	203	11	53	48	+ 4	11	58	48	−25
Manila		W.	38·1	301	11	54	30	+ 5	12	0	32	+ 7
Osaka		O.	43·3	338	11	55	15	+ 9	12	1	23	−16
Batavia		W.	46·9	267	11	55	40	+ 8	? 11	58	20	? P.R.
Zi-ka-wei		W.	47·7	321	11	55	38*	0	12	1	26	−70
		O.	47·7	321	11	55	35	− 3	12	1	42	−54
Tsingtau		W.	51·8	325	11	55	9	−56	12	2	28	−59
Irkutsk		G.	71·0	330	11	58	13	+ 4	12	7	24	0
Colombo	E.	M.	75·0	277	11	59	48	+73	12	8	30	+18
Kodiakanal	E.	M.	77·8	282	11	59	42	+50		—		...
Simla		O.E.	81·3	303	11	59	54	+41	12	9	18	− 6
Bombay	E.	M.	83·3	290	11	56	18	+186		—		...
Victoria	E.	M.	88·9	41	12	6	48	? P.R.		—		...
Tashkent		G.	89·3	312	11	59	59	+ 1	? 12	10	33	−19
Mauritius	N.	M.	93·9	249	12	11	54	? S.	12	11	54	+13
Seychelles	E.	M.	98·4	265	12	12	0	? S.	12	12	0	−26
Baku		G.	103·9	310	12	1	10	− 6	12	14	18	+60
Pulkovo		G.	110·7	333	12	1	47	0	12	14	22	+ 2
Makeyevka		G.	110·9	329	12	1	36	−12	12	15	40	+78
St. Louis	E.	W.	113·3	50	? 12	2	35	+36		—		...
	N.	W.	113·3	50	? 12	2	42	+43		—		...
Ksara		Ma.	116·0	305	12	6	40	? P.R.		—		...
Beirut	E.	M.	116·3	306	12	5	0	? P.R.		—		...
Königsberg		W.	117·7	332	12	5	48	? P.R.	12	14	20	−58
Czernowitz		Ma.	118·2	321	12	6	30	? P.R.	12	16	49	+87
Lemberg		B.O.	118·8	325	12	7	8	? P.R.		—		...
Toronto	E.	M.	119·4	42	12	9	42	? P.R.		—		...
Helwan	E.	M.	120·6	302	12	1	12	−79	12	7	18	? P.R.
Ottawa		B.	121·0	39	? 12	8	0	? P.R.	? 12	17	38	+115
Breslau		W.	121·7	330	12	6	15	+218	12	17	38	+110
Budapest		W.	122·7	325	12	7	21	? P.R.		—		...
Cape Town	E.	M.	122·8	223	12	17	30	? S.	12	17	30	+94
Hamburg		—	123·3	335	12	6	10	+206	? 12	17	39	+99
Vienna		W.	123·7	327	12	6	4	+198	12	17	58	+115
Göttingen		—	124·7	334	12	7	58	? P.R.	? 12	16	49	+39
Graz		W.	124·9	327	12	6	9	+197	12	17	58	+107
Sarajevo	E.	W.	125·0	322	12	6	4	+192	12	17	28	+72
Harvard University		B.O.	125·3	40	12	9	32	? P.R.	12	18	12	+118
Zagreb		W.	125·5	325	12	6	8	+194	12	17	52	+96
Edinburgh	E.	M.	125·9	344	12	8	18	? P.R.		—		...
Laibach		G.	126·0	326	12	6	8	+192	12	18	11	+112
Paisley	E.	M.	126·3	345	12	12	0	...		—		...
Eskdalemuir		G.	126·5	344	? 12	8	18	? P.R.	? 12	21	8	? S.R.
Triest		W.	126·7	326	12	8	18	? P.R.	12	18	13	+109
Innsbruck		Ma.	127·0	329	12	6	15	+195	12	14	0	−146
Heidelberg		—	127·0	333	12	6	42	+222	? 12	19	48	+202
Pola		W.	127·2	325	12	9	33	? P.R.		—		...
Stonyhurst	E.	M.	127·6	343	12	13	30	? S.	12	13	30	−180
Padova		V.	127·9	327	12	6	9	+185	12	15	17	−75
Aachen		W.	127·9	335	12	8	12	? P.R.	? 12	15	24	−68
Bidston	E.	M.	128·1	343	12	0	0	? P.R.	12	14	12	−142
West Bromwich		M.	128·5	342	12	8	0	? P.R.		—		...
Pilar	E.	M.	128·6	138	12	9	48	? P.R.		—		...
Kew	E.	M.	128·9	340	12	8	42	? P.R.		—		...
Pompeii		O.A.	129·1	319	12	6	50	+220		—		...
Guildford		M.	129·1	340	12	1	20	−110	12	14	12	−149
Florence		A.	129·2	326	12	8	27	? P.R.		—		...
Haslemere	E.	M.	129·3	340	? 12	10	24	? P.R.		—		...
Shide	E.	M.	129·8	340	12	3	0	−13		—		...
Paris		—	130·0	336	12	8	31	? P.R.	? 12	15	37	−70
Limerick		—	130·3	346	12	9	6	? P.R.		—		...
Moncalieri		S.	130·4	329	12	6	29	+193	12	17	18	+29
La Paz		—	133·2	118	12	6	30	+191		—		...
Granada		Bif.	141·9	331	? 12	4	57	+52	12	16	31	−89
Rio Tinto	E.	M.	142·8	334	12	17	0	? S.	12	17	0	−66
San Fernando		M.	143·7	333	12	9	12	? P.R.		—		...
Azores	E.	M.	147·3	360	12	7	42	+196		—		...

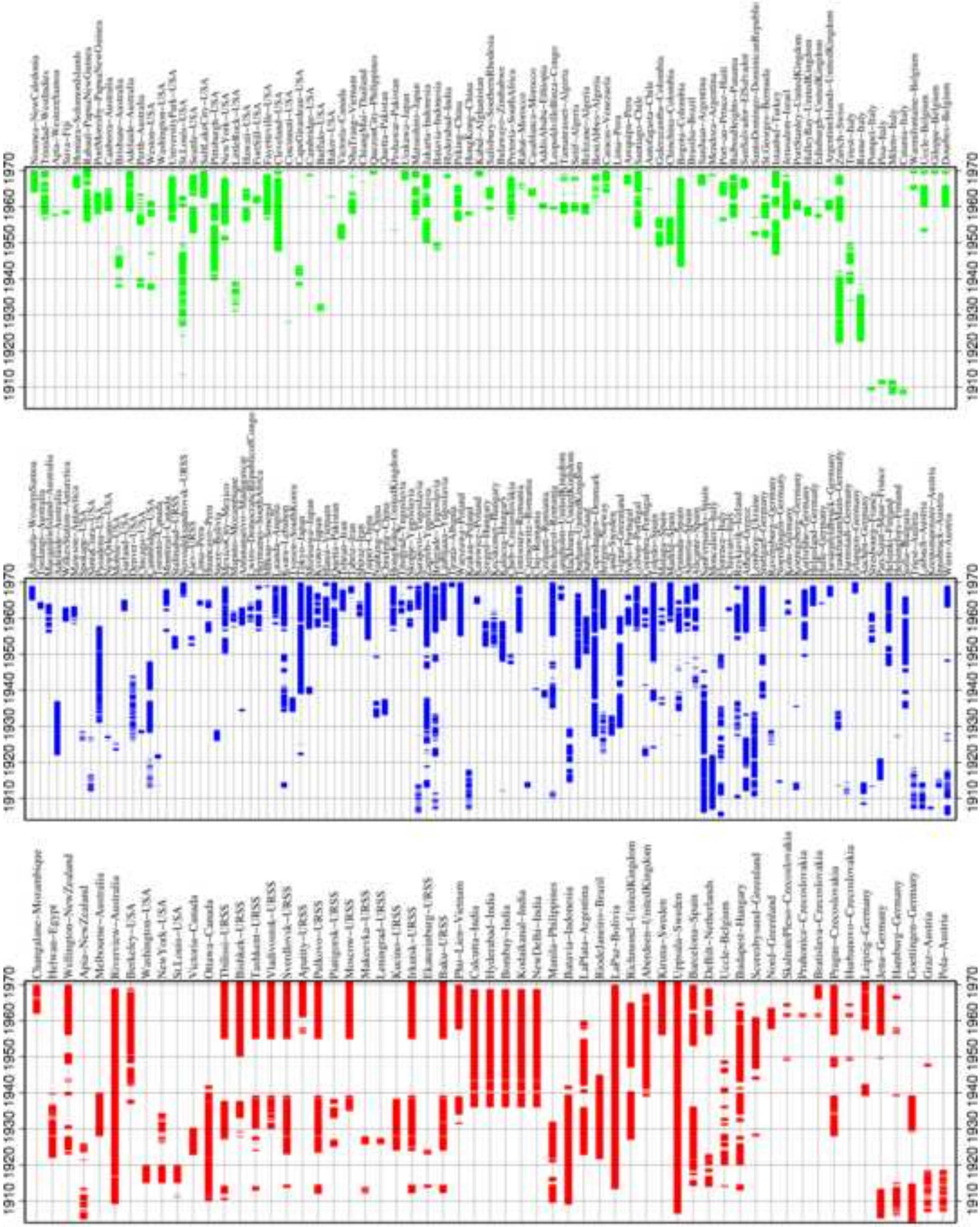
Figure(s)


P–wave type
S–wave type
P–wave type, seismological bulletins
S–wave type, seismological bulletins
Annual number of body–wave arrival times
20000
10000
5000
2000
1000
500
200
100
50
20
10
5
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960

Figure(s)


1
10
50
100
250
500
1000
2000
3000
4000
5000
Number of arrivals per station added before 1960

Figure(s)



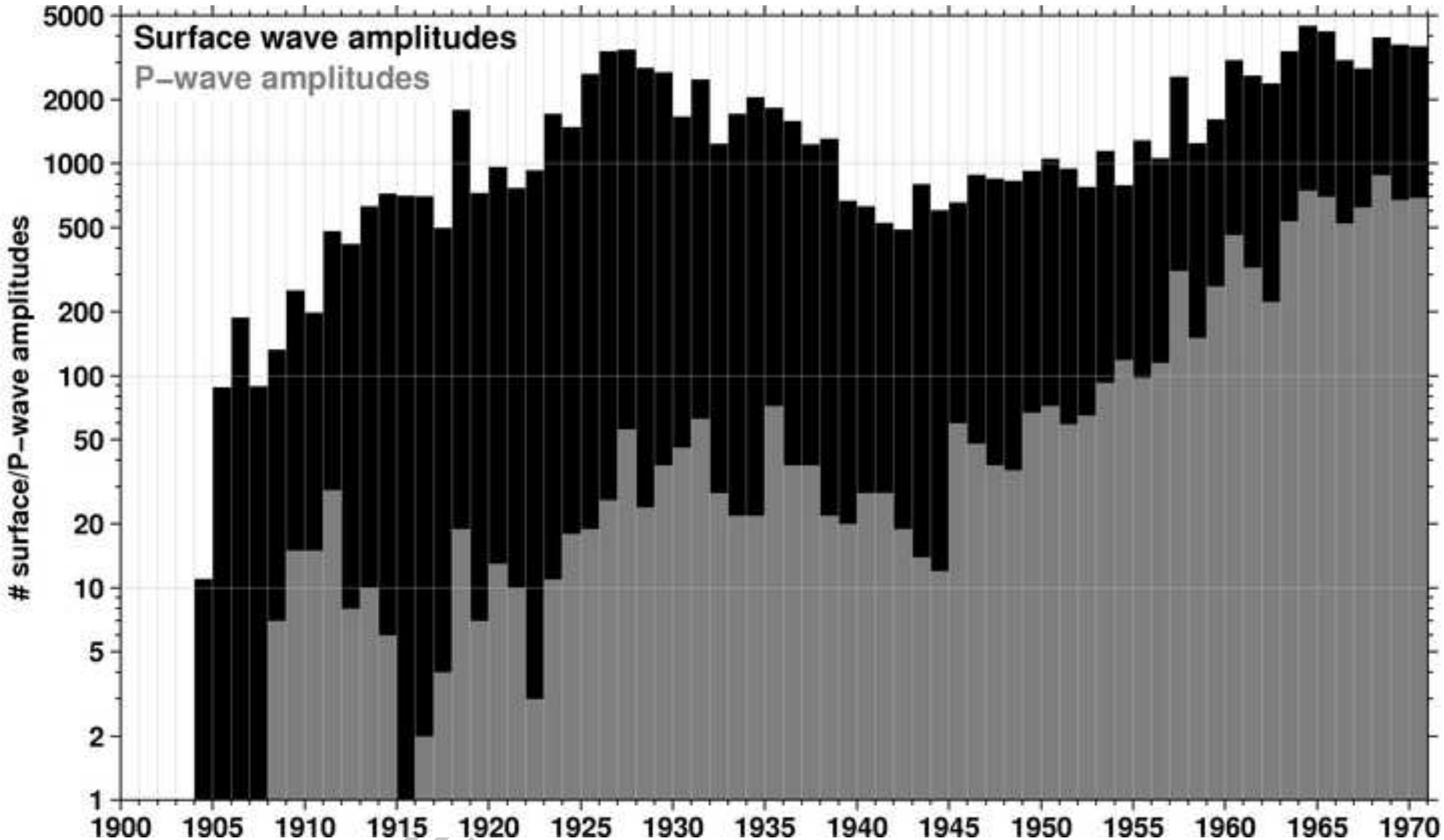
Figure(s)

Datum	Charakter	Phasen	Zeiten (Greenwich)	Perioden Sekunden	Amplituden A_E	Amplituden A_N	Bemerkungen
			h m s		μ	μ	
April 18	III u	e P	13 24 30	3, 6, 9			Kalifornien (San Francisco zerstört).
				18	3	4	
		i	43				Unregelmäßige Wellen.
		P R_1	ca. 28,2				
		P R_2	ca. 30,1				
		S	34 29	17	8		
				20			
		i	35 34	8	25		
				17		80	
		e L	50,6	60			
				35	950	1500	
		M_1	53,1	30		800	Die Ost-Westkomponente des 1200 kg Pendels hat an die Hemmung geschlagen.
		M_2	57	22		750	
		M_N	59,4	20		1600	
		F	19				

a)

sta	day	onset	phase	chan	per	amp	ampunits	amptype
goettingen	1906-04-18 13:24:30	e	P	??E	3;6;9;18	3	micro	0-to-p
goettingen	1906-04-18 13:24:30	e	P	??N	3;6;9;18	4	micro	0-to-p
goettingen	1906-04-18 13:24:43	i						
goettingen	1906-04-18 13:28:12		PR1					
goettingen	1906-04-18 13:30:06		PR2					
goettingen	1906-04-18 13:34:29		S	??E	17	8	micro	0-to-p
goettingen	1906-04-18 13:34:29		S	???	20		micro	0-to-p
goettingen	1906-04-18 13:35:34	i		??E	8	25	micro	0-to-p
goettingen	1906-04-18 13:35:34	i		??N	17	80	micro	0-to-p
goettingen	1906-04-18 13:50:36	e	L	???	60		micro	0-to-p
goettingen	1906-04-18 13:50:36	e	L	??E	35	950	micro	0-to-p
goettingen	1906-04-18 13:50:36	e	L	??N	35	1500	micro	0-to-p
goettingen	1906-04-18 13:53:06		M1	??N	30	800	micro	0-to-p
goettingen	1906-04-18 13:57:00		M2	??N	22	750	micro	0-to-p
goettingen	1906-04-18 13:59:24		MN	??N	20	1600	micro	0-to-p
b) goettingen	1906-04-18 19:00:00		F					

Figure(s)


Surface wave amplitudes
P−wave amplitudes
surface/P−wave amplitudes
5000
2000
1000
500
200
100
50
20
10
5
2
1
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970

Highlights for the paper "ISC-GEM: Global Instrumental Catalogue (1900-1903), I. Data collection from early instrumental bulletins"

- A multitude of printed seismological bulletins between 1904 and 1971 are digitized;
- About 1,000,000 arrivals and 110,000 amplitudes are now part of the ISC database;
- The newly available data relates to global large earthquakes during 1904-1971;
- The digitized data allowed us to re-asses earthquake locations and magnitudes.