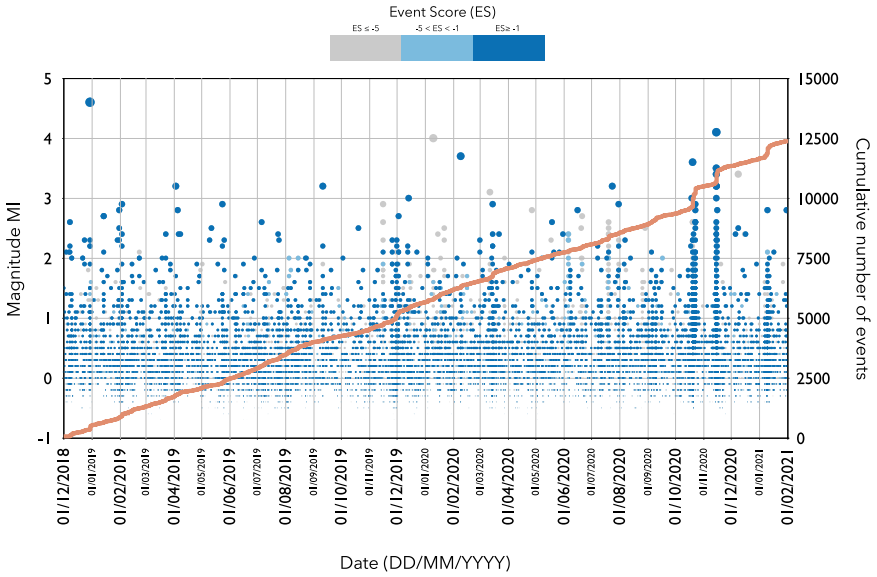


Fig. 5 Temporal evolution of seismicity for the **high**, **medium** and **low quality** catalogues. Each event is indicated by a dot. The score associated to each event is color coded, in grey events with an event score $S \leq -5$, in light blue the events with events score $-5 < S < -1$, and in dark blue the events with events score $S \geq -1$. The cumulative number of events is indicated by the red line.

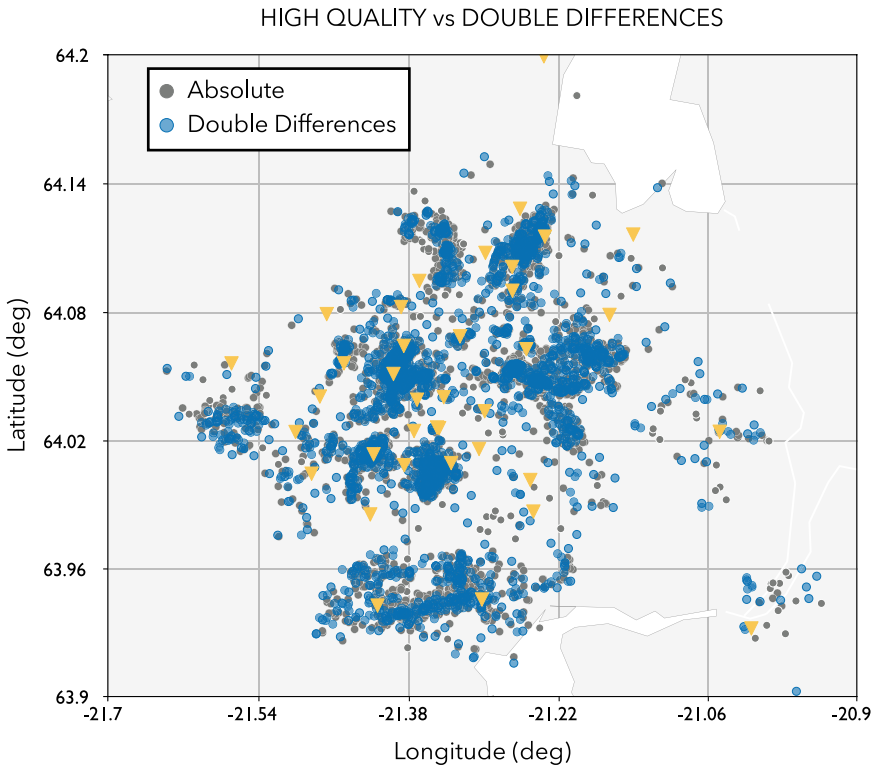


Fig. 6 Comparison of locations from the absolute (**high quality**) and double difference catalogues.The location of the seismic stations are indicated by yellow triangles.

such a background catalogue, these high-quality events can be relocated using a multi-event double-difference relative relocation procedure (i.e. using rtDD in offline mode). We create a double-difference catalogue using the multi-event procedure restricted only to events in the **high quality** catalogue that have been relocated by using rtDD in offline mode (Fig. 6). Note the significantly enhanced clustering and emergence of lineaments for the double difference catalogue.

fdsnws-station	
address	http://eida-federator.ethz.ch/fdsnws/station/1/query?
network	net = 2C,OR,VI,4Q
format	&format = text
level	&level = station&nodata = 404
query	address+network+format+level
fdsnws-dataselect	
address	http://eida-federator.ethz.ch/fdsnws/dataselect/1/query?
network	net = 2C
station	&station = BIT06
time-period	&starttime = 2019-01-01T21:41:14&endtime = 2019-01-01T21:41:34&nodata = 404
query	address+network+station+time-period
fdsnws-event	
address	http://coseismiq.ethz.ch:8080/fdsnws/event/1/query?
time-period	starttime = 2019-01-01T00:00:00&endtime = 2019-01-01T23:59:59
contributor	&contributor = SED_auto_HQ
format	&format = text&nodata = 404
query	address+time-period+contributor+format

Table 4. Fdsnws query examples to retrieve: (top) stations metadata using the **fdsnws-station** service, (middle) waverform data using **fdsnws-dataselect** service, and (3) the available catalogues using the **fdsnws-event** service.

Data Records

The datasets are provided in formats and through services following seismological community standards defined by the International Federation of Digital Seismograph Networks (FDSN, https://www.fdsn.org). Data can be accessed through the the following FDSN web services:

fdsnws-station service to access the station metadata in text and XML format
fdsnws-dataselect service to access the waveform data in miniSEED format
fdsnws-event service to access the event parameters in text and QuakeML format

The continuous raw seismic waveforms are avaialble as binary files in miniSEED format, which is derived from the SEED (Standard for the Exchange of Earthquake Data) data format. While a SEED file consists in both time series values and metadata, the miniSEED format contains only the time series values (binary) and a very limited metadata (identification information). The complete metadata (i.e. station and instrument response information) is stored in a separate file called DATALESS. The metadata describing the stations is available in ascii (i.e. text) and stationXML format (https://stationxml-doc.readthedocs.io/en/release-1.1.0/). The catalogues are available in ascii and quakeMl format (https://quake.ethz.ch/quakeml/). Waveforms, station metadata and seismicity catalogues are available using standard FDSN webservices (https://www.fdsn.org/webservices/). The majority of the temporary COSEISMIQ stations are assigned to a temporary FDSN network code (https://www.fdsn.org/networks) 2 C[23]. For the small aperture array managed by GFZ, the network code is 4Q[24]. The existing stations operated by ISOR use network code OR[25], and those operated by IMO use network code VI.

Waveform data and its associated metadata from 2 C are permanently hosted at the ETH node of the European Integrated Data Archive (EIDA, https://www.orfeus-eu.org/data/eida/). Data from OR and VI are temporarily hosted at the ETH node, and will be moved to a Icelandic node once it is created. Waveform data and station information can be transparently accessed using the EIDA Federator, which provides direct access to the data irrespective of the actual location of the data. Data from the 4Q network are archived at the GFZ EIDA node. Data at the ETH and GFZ EIDA nodes are stored using the SeisComP Data Structure (SDS, https://www.seiscomp3.org/doc/applications/slarchive/SDS.html), where folders are hierarchically organized by year, network code, station names, and channels. Each miniSEED file is 1-day long is named to uniquely identify the time series. The name of each file includes the network code; the station name; the channel; and the Julian date. The catalogues are available using a persistent ETH endpoint. The Table 4 we show few examples on how to access to the data using the different services. More specifically, the query in Table 4 associated with the **fdsnws-station** can be used to provide a list of all the COSEISMIQ stations. This query returns a text file as the format parameter is set to *text*. The location of the station and the temporal duration of available data is indicated. For the permanent networks OR and VI, only data recorded during the COSEISMIQ project is available, the entire dataset will be made available once an Icelandic EIDA node is created. Information at the network and the channel level can be obtained by setting the parameter *level* equal to network or channel, respectively. Custom requests can be performed by adding or modifying query parameters (more detail in the FDSN webservice site).

The second query in Table 4 associated with the **fdsnws-dataselect** service describes, with a simple example, the access to waveform data. This request will return the waveform plotted in Fig. 3.

Finally, the last query of Table 4 and associated with **fdsnws-event** service explains how to access the different seismicity catalogues. With this example we retrieve information about the 3 events included in the **high quality**

Contributor	Description	Number of events
SED_auto_LQ	Low Quality	about 12000
SED_auto_MQ	Medium Quality	about 9900
SED_auto_HQ	High Quality	about 8500
SED_auto_HQ_MEDD	High Quality Double Difference	about 8500
ISOR_manual	Manually Reviewed	about 15000

Table 5. Catalogues available for download sorted by contributor.

catalogue on the date 1.1.2019 and in text format. By changing the contributor parameter events from the other available catalogues can be retrieved. There are 5 different catalogues that can be requested - SED_auto_LQ, SED_auto_MQ, SED_auto_HQ, SED_auto_HQ_MEDD, ISOR_manual as summarized in the Table 5. These catalogues are also accessible through the figshare repository associated with this paper[26]. The figshare repository also contains shell scripts containing pre-compiled FDSN queries allowing to download both continuous waveforms (full dataset) and event waveforms for each seismic catalogue previously described.

Technical Validation

Quality checking of the recorded waveforms has been performed by looking at data completeness and noise analysis. We analyse catalogue quality and completeness by comparing with the manual ISOR catalogue from the same period. The data completeness for each station of the network (within the entire time-frame of the project) is presented in Figure e.1 (in the electronic supplement) that presents the data availability for each station and the percentage of data completeness. In addition, we calculated the Power Spectral Density (PSD) of the noise at each station of the network. These PSDs are accessible at http://www.seismo.ethz.ch/en/research-and-teaching/products-software/station-information/noise-coseismiq/). We observed that high noise levels affect broadband waveforms within the band frequency 0.1–1.0 Hz (mainly related to the primary and secondary microseisms), hence to correctly determine the magnitude of the seismic events we filtered the waveforms with a bandpass filter in the frequency range 2–50 Hz. To evaluate the overall performance of our automatically generated catalogues, we compare them with the manually reviewed catalogue provided by ISOR. In order to match automatically and manually located events we selected the following matching parameters: 1) origin time difference between two events less than 30 seconds and 2) latitude and longitude difference less than 0.1 degrees. If multiple events satisfy this condition we chose the event pair with the smallest origin-time difference. Figure 7 compares the locations of matching events between each of the **low**, **medium** and **high** quality automated catalogues and the manual catalogue.

From Fig. 7 it is clear that the **low** quality catalogue includes events with significant location errors, while the **medium** and **high** quality catalogues are more consistent with the manual catalogue provided by ISOR. The average location error (i.e. average difference between the manual and automated locations) for the **low**, **medium** and **high** quality catalogues are 2.6, 1.2 and 0.7 km respectively. Due to large errors on hypocentral coordinates and origin time of the **low** quality catalogue, we were not able to find a match with all the manually inspected earthquakes. An overview of the location errors for each automatic catalogue with respect to matched locations from the manual ISOR catalogue is shown in Fig. 8, showing that for about 80% of the events the hypocentral location difference between the automated (any quality) and matching manual locations is within 1 km. It is important to note that the the **low**, **medium** and **high** quality catalogues are obtained using fully automated procedures and the quality based classification has been performed by filtering the raw catalogue using the quality score and the number of phases as described in the previous section.

Usage Notes

The Hengill region is characterized by an intense seismic activity, and using the dense seismic network that operated across the 26 months analysed in this manuscript, more than 10,000 events have been detected. The COSEISMIQ seismic network, comprising about 40 stations deployed with an average inter-station distance of about 2 km, is a unique dataset for its genre. The massive number of earthquakes that have occurred in the area, combined with the presence of many seismic sequences characterized by very short inter-event times (about 10 s) makes the analysis of this dataset particularly challenging, and hence is a perfect playground for data intensive techniques such as full-waveform or machine learning based analysis methods[27]. The seismic catalogues (both manual and automated) accompanying this paper can be used as reference to evaluate the performance of newly tested methods. In addition, due to the complex geology of this region, the dataset presented within this paper can be a valuable asset to better studying the natural and induced seismicity of the area. In publishing this dataset (consisting of both continuous raw waveforms and seismicity catalogues) one of our main aim is to provide a baseline for the comparison of fully automated methods for the analysis of seismicity, hence our automatic catalogues have been only sorted by quality score and not manually inspected after their generation. For this reason if not used for the benchmark of newly developed methods, these catalogues should be handled with caution, this is particularly true for the low quality catalogue that includes events with large location errors and false events. The medium and the high quality catalogue (and, of course, the double difference catalogue), on the other hand, are better suited to be used as a starting point for additional seismological analyses (e.g. focal mechanism determination, b-value analysis etc.) or interpretation.

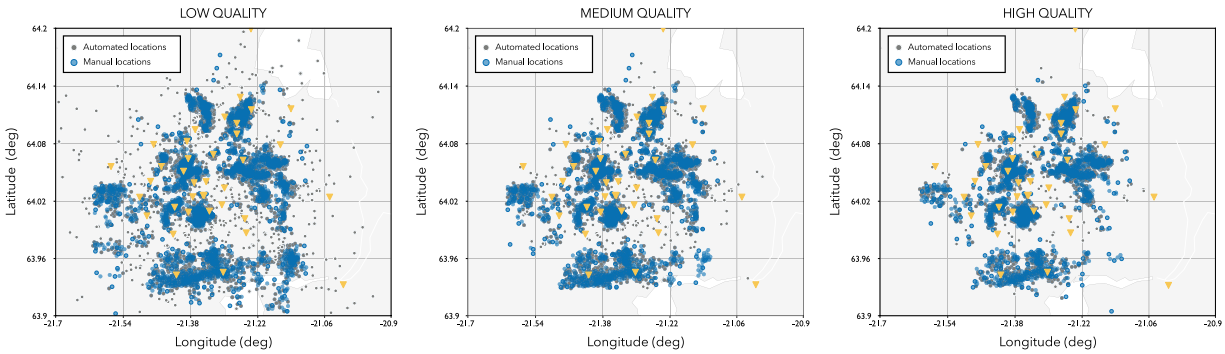


Fig. 7 Comparison between the automatic (blue) and ISOR manual (red) seismic events. The comparison is performed for **low** (left panel), **medium** (central panel) and **high** (right panel) quality catalogues.

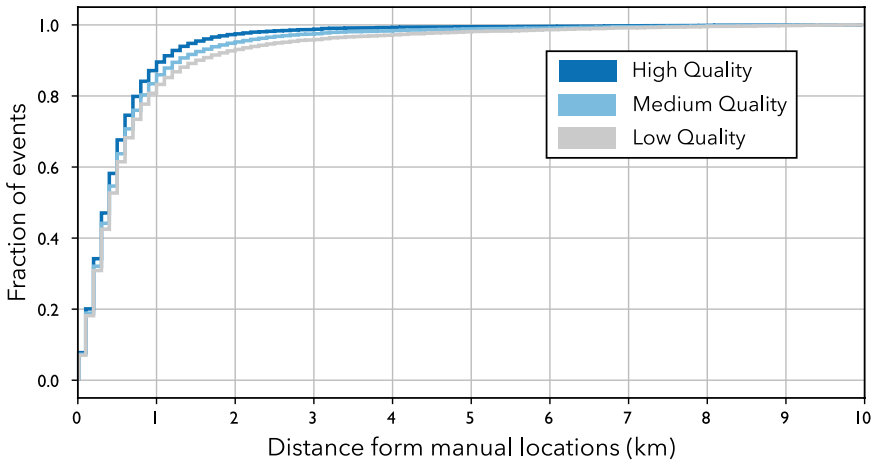


Fig. 8 Cumulative distribution of the distances between the automated and manual locations (hypocentral distances) for each matched event in the 3 catalogues. Results for **low** (grey color), **medium** (light blue) and **high** (dark blue) quality catalogues. Note that we plot the cumulative only up to 10 km distance in order to aid visualization, a small fraction of events have location distances in excess of 10 km.

Code availability

Both the ISOR manual and the automated catalogues were produced using the SeisComP software package. While the core of SeisComP is open source and freely available at http://seiscomp.de, in our analysis we also used the scanloc module which is provided under license by Gempa GmbH http://gempa.de. The evscore and scrtDD modules are open-source and available at https://github.com/swiss-seismological-service.

Received: 1 December 2021; Accepted: 19 April 2022;
Published online: 19 May 2022

References

1. Ellsworth, W. L. Injection-induced earthquakes. *Science* **341**(6142), 1225942 (2013).
2. Grigoli, F. *et al.* The November 2017 Mw 5.5 Pohang earthquake: A possible case of induced seismicity in South Korea. *Science* **360**(6392), 1003–1006 (2018).
3. Woo, J. U. *et al.* An in-depth seismological analysis revealing a causal link between the 2017 MW 5.5 Pohang earthquake and EGS project. *Journal of Geophysical Research: Solid Earth* **124**(12), 13060–13078 (2019).
4. Lee, K. K. *et al.* Managing injection-induced seismic risks. *Science* **364**(6442), 730–732 (2019).
5. Broccardo, M. *et al.* Induced seismicity risk analysis of the hydraulic stimulation of a geothermal well on Geldinganes, Iceland. *Natural Hazards and Earth System Sciences* **20**(6), 1573–1593 (2020).
6. Schultz, R., Beroza, G. C. & Ellsworth, W. L. A risk-based approach for managing hydraulic fracturing–induced seismicity. *Science* **372**(6541), 504–507 (2021).
7. Gunnarsson, A., Steingrimsson, B. S., Gunnlaugsson, E., Magnusson, J. & Maack, R. Nesjavellir geothermal co-generation power plant. *Geothermics* **21**(4), 559–583 (1992).
8. Flovenz, O. G., Agustsson, K., Gudnason, E. A. & Kristjánsdóttir, S. Reinjection and induced seismicity in geothermal fields in Iceland. In *Proceedings world geothermal congress* (pp. 1–15) (2015).
9. Gunnlaugsson, E. Environmental management and monitoring in Iceland: Reinjection and gas sequestration at the Hellisheidi power plant. SDG Short Course I on Sustainability and Environmental Management of Geothermal Resource Utilization and the Role of Geothermal in Combating Climate Change", organized by UNU-GTP and LaGeo, in Santa Tecla, El Salvador, September, 4–10 (2016).
10. Agustsson, K., Kristjansdottir, S., Flóvenz, O. G. & Gudmundsson, O. Induced Seismic Activity during Drilling of Injection Wells at the Hellisheidi Power Plant, SW Iceland. *Proceedings World Geothermal Congress 2015* Melbourne, Australia, 19–25 April 2015.

11. Halldorsson, B. *et al*. On the effects of induced earthquakes due to fluid injection at hellisheidi geothermal power plant, Iceland. Proc. 15th WCEE, Lissabon, Portugal (2012).
12. Rossi, C. *et al*. Full-Waveform based methods for Microseismic Monitoring Operations: an Application to Natural and Induced Seismicity in the Hengill Geothermal Area, Iceland. *Advances in Geosciences* **54**, 129–136 (2020).
13. Olivieri, M. & Clinton, J. An almost fair comparison between Earthworm and SeisComp3. *Seismological Research Letters* **83**(4), 720–727 (2012).
14. Grigoli, F. *et al*. Pick-and waveform-based techniques for real-time detection of induced seismicity. *Geophysical Journal International* **213**(2), 868–884 (2018).
15. Zhang, H., Thurber, C. & Rowe, C. Automatic P-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings. *Bulletin of the Seismological Society of America* **93**(5), 1904–1912 (2003).
16. Lomax, A., Virieux, J., Volant, P. & Berge-Thierry, C. Probabilistic earthquake location in 3D and layered models. In Advances in seismic event location (pp. 101–134). Springer, Dordrecht (2000).
17. Kissling, E. Geotomography with local earthquake data. *Reviews of Geophysics* **26**(4), 659–698 (1988).
18. Kissling, E., Ellsworth, W. L., Eberhart-Phillips, D. & Kradolfer, U. Initial reference models in local earthquake tomography. *Journal of Geophysical Research: Solid Earth* **99**(B10), 19635–19646 (1994).
19. Duran, A. Numerical sensitivity kernels in elastic media for imaging purposes and seismic tomography in the Hengill geothermal field.Doctoral dissertation. *ETH-Zurich* https://doi.org/10.3929/ethz-b-000518037 (2021).
20. Waldhauser, F. & Ellsworth, W. L. A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California. *Bulletin of the Seismological Society of America* **90**(6), 1353–1368 (2000).
21. Waldhauser, F. Near-Real-Time Double-Difference Event Location Using Long-Term Seismic Archives, with Application to Northern California. *Bulletin of the Seismological Society of America* **99**(5), 2736–2748, https://doi.org/10.1785/0120080294 (2009).
22. Scarabello, L. swiss-seismological-service/scrtdd https://github.com/swiss-seismological-service/scrtdd (2021).
23. Swiss Seismological Service (SED) at ETH Zurich. COSEISMIQ - COntrol SEISmicity and Manage Induced earthQuakes. *ETH Zurich* https://doi.org/10.12686/sed/networks/2c (2018).
24. Dahm, T., Cesca, S., Heimann, S., Mikulla, S. & Milkereit, C. GFZ contribution to the COSEISMIQ seismic network. *GFZ Data Services* https://doi.org/10.14470/9d7564946595 (2021).
25. Reykjavik Energy (Iceland) OR - Reykjavik Energy. *International Federation of Digital Seismograph Networks*. https://doi.org/10.7914/SN/OR (2016).
26. Grigoli, F. *et al*. Monitoring microseismicity in the Hengill Geothermal Field, Iceland. *figshare* https://doi.org/10.6084/m9.figshare.c.5725592 (2022).
27. Nooshiri, N. *et al*. A Multi-Branch, Multi-Target Neural Network for Rapid Point-Source Inversion in a Micro-Seismic. *Environment: Examples from the Hengill Geothermal Field, Iceland* **229**(2), 999–1016, https://doi.org/10.1093/gji/ggab511 (2022).

Acknowledgements

We thank the editor Dr. Guy Jones and three anonymous reviewers that helped us to improve the quality of the manuscript. This work was funded by the project COSEISMIQ of the EU Geothermica program, Project no: 170167-44011. FG was supported by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska Curie Grant agreement (790900).

Author contributions

F.G. and J.C. wrote and generated the figures of the manuscript. F.G., L.S. and J.C. designed the data analysis procedure. F.G., J.C. and P.K. designed and organized the infrastructure for data release. F.G., L.S. and T.D. performed data analysis. T.A., S.K. and R.M. curated the ISOR manual catalogue and managed the station metadata. All the other authors contributed to the manuscript in terms of activities related to data collection, quality control, data management, data storage or project management and coordination. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-022-01339-w.

Correspondence and requests for materials should be addressed to F.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022