REPORT
doi:10.26443/seismica.v5i1.1815

A quasi-real-time system for automatic local event monitoring in Germany

Catalina Ramos * [1], **Stefanie Donner** [1,2], **Klaus Stammler**[1]

[1]Federal Institute for Geosciences and Natural Resources (BGR), Hanover, Germany, [2]Institute of Geophysics, Center for Earth System Research and Sustainability (CEN), University of Hamburg, Hamburg, Germany

Author contributions: *Conceptualization*: C.R., S.D. *Software*: C.R., K.S. *Validation*: C.R. *Formal Analysis*: C.R. *Investigation*: C.R. *Writing - Original draft*: C.R. *Writing - Review & Editing*: C.R., S.D., K.S. *Visualization*: C.R. *Supervision*: S.D. *Funding acquisition*: S.D.

Abstract We present TieBeNN, a wrapper that integrates open-source, state-of-the-art seismic monitoring tools, including advanced machine learning–based approaches, to enhance the German Federal Seismological Survey's (EdB) automatic real-time earthquake monitoring system. TieBeNN extends the existing workflow by adding automatic, probabilistic focal depth estimation using NonLinLoc and introduces a Location Quality Score (LQS) to quantify location reliability with a single metric. In testing, TieBeNN's automated locations approach the accuracy of human analyst solutions, demonstrating comparable performance in well-instrumented regions. By automating depth determination and providing immediate quality assessment, the system reduces analysts' daily workload, allowing them to focus on events flagged as low-quality or complex. The LQS effectively distinguishes well-constrained event locations from those with large uncertainties or poor network geometry, enabling rapid identification of high-quality automatic results versus those requiring review. However, events below the Moho depth (i.e., deeper than approximately 30 km), which are rare in Germany, remain challenging: their uncertainties are larger, and LQS values tend to be lower, indicating limitations in the current calibration. Overall, these enhancements significantly advance real-time local seismic event monitoring in Germany by increasing both the speed and reliability of automatic event characterization.

Production Editor:
Andrea Llenos
Handling Editor:
Hongyu Sun
Copy & Layout Editor:
Sarah Jaye Oliva

Received:
July 7, 2025
Accepted:
January 14, 2026
Published:
February 10, 2026

1 Introduction

The rapid growth of vast seismic data repositories in recent years reflects the necessity for robust automatic real-time processing workflows in earthquake monitoring, capable of performing at a level comparable to human analysts (e.g. Quinteros et al., 2021; Stammler et al., 2021; Strollo et al., 2021; Arrowsmith et al., 2022; Mousavi and Beroza, 2022). This demand is driven by the expansion of seismic networks, increasing computational capabilities, and continuous improvements in seismic equipment. Collectively, these developments have facilitated the detection of smaller magnitude earthquakes, which occur far more frequently than larger events. These advances have transformed seismology into a data-rich science, creating numerous opportunities to apply machine learning to a broad range of seismology-related problems, including earthquake monitoring, subsurface characterization, and seismic image interpretation (Arrowsmith et al., 2022; Mousavi and Beroza, 2022, 2023). Specifically, deep learning, a data-driven technique, has been widely employed in big data analysis and has been applied to tasks including event detection and seismic phase picking (Ross et al., 2018a; Zhu and Beroza, 2019; Mousavi et al., 2020; Soto and Schurr, 2021; Münchmeyer et al., 2022; Li et al., 2024), event classification (Renouard et al., 2021; Wang et al., 2023), seismic moment tensor estimation (Steinberg et al., 2021), magnitude estimation (Mousavi and Beroza, 2020; Quinteros-Cartaya et al., 2024) and polarity estimation (Ross et al., 2018b; Hara et al., 2019).

Automated event monitoring workflows that at least partially leverage advances in machine learning have been successfully applied in various contexts. For instance, Shi et al. (2022) generated continuous seismic phase probabilities, which were subsequently backprojected and migrated for seismic source location. This approach was used to monitor induced earthquakes around geothermal production sites in Iceland. Zhang et al. (2022) developed an earthquake location workflow that integrates multiple routine functions for local and regional events. They applied it to the 28 September 2004 M_w 6.0 Parkfield sequence, locating over three times as many events as the Northern California Earthquake Data Center. Likewise, Retailleau et al. (2022) created a wrapper to perform high-quality seismicity analysis in real time, enhancing the reaction time of observatories monitoring active volcanic systems and their associated tectonic settings.

In this report, we introduce an automatic, near-real-time local event detector developed for Germany. Our system leverages community-standard and state-of-the-art tools in seismic monitoring, including machine learning–based models for seismic event detection, phase picking, and waveform denoising (Münch-

*Corresponding author: catalina.ramosdomke@bgr.de

meyer et al., 2022; Zhu et al., 2019). It also incorporates a novel phase associator inspired by the oct-tree data structure (Münchmeyer, 2024) and employs the traditional probabilistic earthquake location method for absolute hypocenter estimation (Lomax et al., 2000).

2 Automatic local event detections at BGR

2.1 The German earthquake catalog

In Germany, the Federal Seismological Survey (EdB, *Erdbebendienst des Bundes*), part of the Federal Institute for Geosciences and Natural Resources (BGR, *Bundesanstalt für Geowissenschaften und Rohstoffe*), addresses seismological and geophysical issues at the national level and sustains the German Regional Seismic Network (GRSN), along with several additional seismic stations (Stammler et al., 2021). The EdB collaborates closely with the seismological surveys of Germany's federal states, as well as affiliated universities and research institutes, to monitor seismicity across the country. It operates a data center that serves as an official node of the European Integrated Data Archive (EIDA; https://eida.bgr.de/). Waveform data and station information are freely accessible via the FDSN/EIDA web services. An FDSN event service is currently in preparation and is already operational internally. The internal catalog includes both preliminary events detected automatically and those reviewed by the analyst on duty using the SeismicHandler software package (Stammler, 1993) as part of routine verification procedures.

The German earthquake catalog has a magnitude of completeness of 2 (Stammler et al., 2021). In addition to tectonic earthquakes, which account for approximately 30% of the analyzed events, 56% are quarry blasts and explosions. Moreover, induced seismicity represents a significant portion of the catalog in Germany, a country with substantial geothermal resources and growing investment in fossil fuel–free energy sources. Consequently, 14% of seismic events are labeled as induced, such as those related to natural gas production or geothermal energy utilization.

2.2 Detection algorithms at BGR

A range of algorithms has been implemented at BGR to analyze the full waveform dataset for automatic event detection. These include Short Time Average over Long Time Average (STA/LTA) detectors on single traces; an array detector applied to Gräfenberg data in southern Germany (primarily for detecting teleseismic events arriving with nearly plane wavefronts); template matching; a SeisComP instance; and an amplitude decay (AD) detector (Stammler et al., 2021, and references therein). The outputs of these detection algorithms are compiled into a summary detection list, which is made available on an internal website and serves as the foundation for daily manual analysis.

The primary automatic detector for local and regional events at BGR is the AD-Detector. Unlike the other systems, it does not rely on seismic phase picks. Instead, it performs a search over an irregularly dense grid covering Germany and neighboring regions, progressing in one-second time steps using waveform segment files of continuous data. Each grid point represents a hypothetical epicenter. Grid-point density is derived from seismicity maps, with grid-point spacings in Germany ranging between 20 and 40 km, denser in areas of higher seismicity. For each point, signal-to-noise ratios are calculated in time windows based on expected arrival times of the P_g, P_n, S_g, and S_n phases, derived using the Regional Seismic Travel Time (RSTT) model (Myers et al., 2010). These signal-to-noise values are evaluated in a number of ways, including amplitude decay with epicentral distance, to compute a quality-function value. This quality is taken as the maximum across all grid points at a given time step. A detection is triggered when the quality function exceeds a dynamic threshold derived from an elevated smoothed mean of the quality values. Each detection includes the latitude and longitude of the grid point with the highest quality value, along with the event time. Since the AD-Detector does not estimate depth, theoretical travel times are calculated for a source at 0 km depth. Nonetheless, the algorithm is highly sensitive and reliably detects events as small as approximately $\mathrm{M}_L \sim 1.4$, depending on local station density and ground properties (see Figure 1). Because grid density scales with regional seismicity, the AD-Detector's preliminary epicenters are typically within < 10 km of the final source locations. The AD-Detector also estimates a magnitude for each detection based on amplitudes in the signal windows. While the input traces are not Wood–Anderson simulations—and thus not fully compliant with standard local magnitude definitions—the estimates are adequate for a first approximation of event strength.

The AD-Detector currently has a delay of approximately five to ten minutes in generating detection messages after the event time. This is due to both its internal setup, which processes data in five minute segments, and its computational intensity, as the algorithm performs millions of complex operations per second during grid searches. However, advances in multicore processor technology have substantially reduced the system's runtime in recent years.

The latest version of the AD-Detector operates in near real time using a set of software containers that perform a series of tasks, including waveform reading from SeedLink servers (https://docs.fdsn.org/projects/seedlink/en/latest/protocol.html) or disk; waveform preparation (windowing, filtering) using ObsPy (Beyreuther et al., 2010; Megies et al., 2011; Krischer et al., 2015); executing the core detection algorithm; result message generation; and web-accessible visualization for operators and analysts. Statistics are also compiled. Communication between containers and external consumers is handled via detection messages using the RabbitMQ message broker (https://www.rabbitmq.com/).

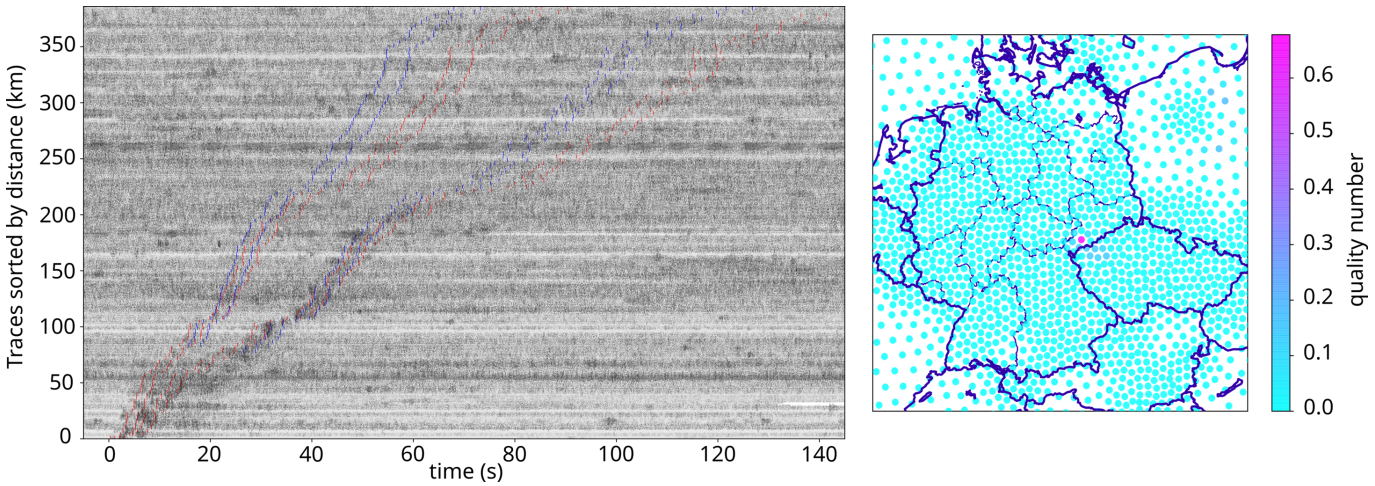

Figure 1 Quick-look graphics automatically generated by the AD-Detector for analyst orientation, showing an M_L 1.3 event detected by the AD-Detector. Left: Waveform traces sorted by epicentral distance from the grid point with the highest quality value, with colored lines indicating time windows for the expected P_g and S_g phases (red) and P_n and S_n phases (blue). Right: The irregular grid on which the quality function is calculated at one-second intervals, showing the event epicenter in the West Bohemia/Vogtland region, represented by the grid point with the highest quality value. This example illustrates the AD-Detector's sensitivity to small-magnitude events.

3 From automatic local event detection to hypocenter estimation

The automatic local event monitoring system developed in this work, called TieBeNN (*TIEfenBEstimmung mittels Neuronaler Netze*, Depth Estimation using Neural Networks), was designed to provide more precise automatic event locations and to integrate focal depth estimation into the daily real-time monitoring routine at EdB. The system is triggered whenever a quasi-real-time detection message is generated by the AD-Detector. The TieBeNN wrapper performs three key tasks, detailed below.

3.1 Phase picking

In the first step, after receiving a detection message from the AD-Detector—which includes the event time and epicenter coordinates—an ObsPy client accesses waveform data stored locally in the SeisComP Data Structure (SDS) directory at the Federal Seismological Survey. Waveform data are collected from both broadband and short-period channels, where first arrivals of local events are usually well detected. The network selection is automatic and depends on the epicenter coordinates. Typically, this includes the GRSN (about 50 stations), some stations of the GEOFON network maintained by the German Research Center for Geosciences (GFZ, *Deutsches GeoForschungsZentrum*), as well as local and regional networks operated by German states and universities, complemented by stations in neighboring countries (France, Belgium, Luxembourg, Netherlands, Switzerland, Austria, Czech Republic, and Poland; Stammler et al., 2021 and references therein), in total about 350 stations in and around Germany. Some federal network data are under temporary embargo and used only internally. Any waveform data not available locally are requested via FDSN web servers, ensuring broader station coverage, especially near borders.

Sixty-second waveform windows are retrieved from stations within a fixed epicentral distance, which is user-defined. Our tests show that optimal retrieval distances depend strongly on station density: in well-instrumented regions, a well-distributed station set can already be obtained within 60 km. However, in poorly covered areas, even extending the epicentral distance to 150 km may still result in insufficient station coverage. We focus exclusively on P_g and S_g phases, as refracted waves can introduce significant depth errors without an accurate model of the refractor (Diehl et al., 2021).

The collected waveforms undergo preprocessing, including denoising with the machine learning–based DeepDenoiser model (Zhu and Beroza, 2019), but only for stations within 100 km. For more distant stations, denoising alters the P-wave coda in a way that often results in S-waves incorrectly classified as P arrivals. For phase picking, we use models integrated into SeisBench—a framework for training, evaluating, and comparing machine learning–based models on benchmark datasets (Münchmeyer et al., 2022; Woollam et al., 2022). Specifically, PhaseNet (Zhu et al., 2019) and EQTransformer (Mousavi et al., 2020) proved robust on our dataset of local events in Germany, showing strong translation invariance and consistent phase detection, in agreement with the findings of Münchmeyer et al. (2022). Both models support picking on single-channel and three-component stations. Based on their accurate predictions on our dataset—with median absolute errors of 0.08 s (P) and 0.10 s (S), and mean absolute errors of 0.16 s (P) and 0.22 s (S)—we use the pre-trained versions provided by SeisBench rather than training on our own data.

3.2 Phase association

False picks can degrade event location and especially depth estimation. Their impact is most severe for low signal-to-noise ratios, sparse pick sets, and small-

magnitude events. Therefore, reliably associating phase picks is critical.

TieBeNN supports two phase associators. The first, GaMMA (Zhu et al., 2022), treats phase association as an unsupervised clustering problem using arrival times and amplitude features. The second, PyOcto (Münchmeyer, 2024), is specifically designed for addressing sensitivity and long runtimes when dealing with dense seismic sequences by using 4D space-time partitioning (mimicking a grid search), while supporting both homogeneous and 1D velocity models. Both models were tested for real-time feasibility.

In our tests, PyOcto performed slightly better—correctly associating more picks and missing fewer—so it was set as the default associator for TieBeNN's real-time workflow. GaMMA remains available as an alternative. The 1D velocity model used for PyOcto is the same one later employed for probabilistic hypocenter estimation (see below).

3.3 Probabilistic hypocenter estimation

Once picks are associated, TieBeNN exports the location information, along with station metadata and control files, for hypocenter estimation using NonLinLoc (Lomax et al., 2000, 2014). We employ its suite of programs for velocity model construction, travel-time calculation, and global probabilistic location.

An advantage of our automatic location workflow is the availability of the AD-Detector's preliminary, yet relatively accurate, epicenter estimate. This information enables the use of local 1D velocity models—when available—for a more refined hypocenter search within a finer local grid using NonLinLoc. These dedicated models improve epicentral accuracy over the coarser fixed-grid estimate from the AD-Detector.

Local 1D velocity models integrated into TieBeNN for hypocenter estimation cover various regions in Germany and neighboring countries. In southwestern Germany, the system employs the vertical seismic-profile–derived model for Insheim (Küperkoch et al., 2018). For western Germany, models include the BENS velocity model, developed from well-recorded local events in the Northern Rhine area (Reamer and Hinzen, 2004), and the more recent KIT6 velocity model for the East Eifel Volcanic Field (Ritter et al., 2024). In southern Germany and along the German–Austrian and German–Swiss borders, the system uses the ASZmod1 velocity model for the Albstadt Shear Zone (Mader et al., 2021), which is characterized by simple and stable layering; the AlpsLocPS model (Braszus et al., 2024), a machine learning–based velocity structure for the Greater Alpine Region; and a model for the Central Alps derived from the inversion of high-quality P- and S-phase data (Diehl et al., 2021). In West Bohemia/Vogtland—one of the most seismically active regions surrounding Germany—available 1D models include the isotropic upper crustal model WB2005 (Málek et al., 2005), a 1D approximation of the 3D velocity model WB2012 (Růžek and Horálek, 2013), and the P_{O_2} velocity structure derived via isometric inversion for the Nový Kostel earthquake swarm (Málek et al., 2023). For all other regions, TieBeNN uses a standard 1D layer-over-half-space model applied routinely at EdB for local event analysis (Schlittenhardt, 1999; Stammler et al., 2021). This model is based on average travel-time curves from seismic phases recorded by GRSN stations over the past 25 years.

Although TieBeNN supports 3D velocity models for hypocenter estimation, no robust nationwide model exists. Currently, local, dedicated 1D models outperform regional or global 3D alternatives (Stammler et al., 2021).

3.4 Limitations in automatic local event location and some mitigation strategies

Despite their strong performance, deep learning–based phase pickers are vulnerable to generalization errors, sensitivity to noise, and false detections—all of which can impair downstream location accuracy. To mitigate this, several strategies have been proposed. For example, superimposing local seismic noise—readily obtained from continuous seismic recordings—onto training data improves model robustness (e.g., Zhu et al., 2020; Yin et al., 2022). Yuan et al. (2023) introduced a multi-model, multi-frequency ensemble approach using SeisBench models to generate multiple predictions per seismogram. Picks are then combined to increase reliability. However, the application of this approach—which is not implemented in our workflow—in quasi-real-time requires evaluating its computational cost. Multicore processing may help offset delays.

Park et al. (2023) highlight that predictions made by machine learning–based phase pickers can vary—sometimes substantially—depending on the alignment of waveform data within the processing window. Since our automatic depth estimation workflow is triggered upon receiving detection messages from the AD-Detector, which provide event times accurate to within a few seconds, careful definition of waveform windowing is essential to mitigate prediction inconsistency. In our dataset, we repeatedly observed several manifestations of this problem: for instance, predicted arrival times—represented as timestamps with maximum predicted probability near the actual phase—can vary depending on the waveform start time; alternatively, predicted timestamps might remain stable across different time-window alignments but exhibit sufficiently low predicted probabilities in some alignments to fall below the fixed detection threshold. Additionally, the most accurate P-phase predictions sometimes occur under a different waveform alignment than that yielding the most accurate S-phase picks. To address these issues, we implemented a multiple-window approach, varying the start time for phase-picking windows between 0 and 10 s prior to the event's origin time. Specifically, discrete window start times at 0, 2, 5, and 10 s before the event were used, each with a total duration of 60 s. For each phase and station, we retain the timestamp with the maximum predicted probability across all window alignments. This approach significantly reduced prediction inconsistencies while remaining practical for quasi-real-time applications.

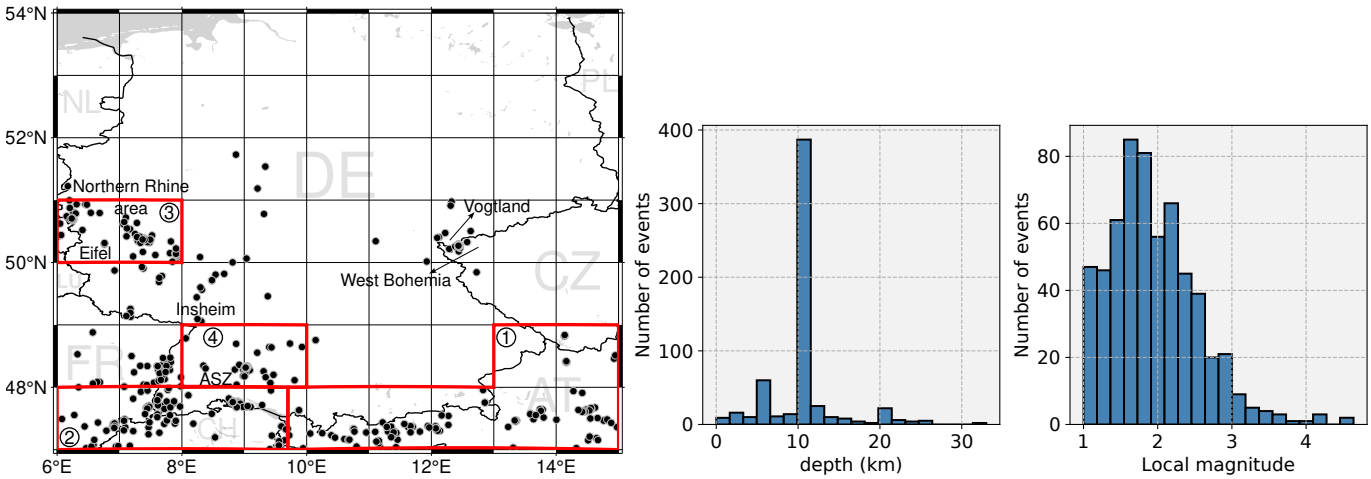

Figure 2 The tectonic events in the BGR manual catalog used for testing TieBeNN. Red boxes indicate regions where local velocity models were applied for hypocenter estimation with NonLinLoc: box 1 (AlpsLocPS model; Braszus et al., 2024), box 2 (Central Alps model; Diehl et al., 2021), box 3 (BENS model; Reamer and Hinzen, 2004), and box 4 (ASZmod1 model; Mader et al., 2021). ASZ: Albstadt Shear Zone. Histograms show the local magnitude and depth distribution.

4 Testing the automatic local event location system

4.1 An event catalog for the test

To evaluate TieBeNN's performance, we produced a catalog of manually located events at BGR to be reprocessed using our automatic system. This catalog comprises tectonic events in Germany and neighboring regions from 2021 through 2023, constrained spatially within latitude 47–54°N and longitude 6–15°E, and with local magnitudes greater than 1, totaling 594 events. Around 60% of the events have a fixed manual depth, typically due to limited station coverage and/or insufficient S-phase constraints. Figure 2 displays the spatial distribution of the events along with histograms illustrating their depth and local magnitude distributions. The histograms show that the majority of cataloged events have local magnitudes less than 2 and that often depths in the catalog are manually fixed at 10 km.

4.2 TieBeNN setup

TieBeNN requires the event time and preliminary epicentral coordinates as inputs, which, in our test configuration, were taken from the BGR manual catalog; in real-time operation these will be supplied by AD-Detector detection messages. We set a maximum epicentral distance of 150 km for waveform retrieval, limiting to the 70 closest stations with phase picks. Based on our experience, this ensures high-quality locations and adequate azimuthal coverage in densely instrumented regions. PhaseNet was employed for phase picking, although EQTransformer yielded comparable results on our dataset.

Phase association using PyOcto reduced false detections and selected the correct phases when multiple events occurred close in time. In such cases, the phase picks associated with the PyOcto solution whose origin time was closest to that in the AD-Detector message were retained, and phases associated with other solutions were discarded. However, for events only a few seconds apart (e.g. during seismic swarms), the phase pickers struggled to detect correct phases for the later events, particularly at greater distances where the later P-wave arrivals were obscured by the earlier event's coda waves.

For hypocenter estimation, we employed EdB's standard 1D layer-over-half-space velocity model for events outside the dedicated regions outlined by red boxes in Figure 2. In the present tests, the manual-catalog epicenters were used to select the appropriate local velocity model automatically; in real-time operation the same selection will be driven by the AD-Detector's preliminary epicenter (see Figure 2 for more details). Interestingly, automatic locations in the West Bohemia/Vogtland region resulted in locations with smaller root-mean-square (RMS) values and uncertainties with the simplified model for Germany than with the local WB2005 and WB2012 models (Málek et al., 2005; Růžek and Horálek, 2013).

4.3 Location Quality Score

Given the heterogeneous quality of automatic locations throughout our catalog region, we define the Location Quality Score (LQS) to quantitatively evaluate automatic location reliability using a single comprehensive metric. The LQS combines eight normalized parameters, each closely related to location quality:

1. **(Primary) azimuthal gap (°):** Largest angle separating two adjacent stations, measured from the epicenter.
2. **Distance to nearest station (km):** Epicentral distance to closest station with phase picks.
3. **Station density (stations/km^2):** Number of stations per area (πR^2, where R is the epicentral distance that contains 80% of the stations).
4. **Secondary azimuthal gap (°):** Largest gap filled by one station, measured from the epicenter.

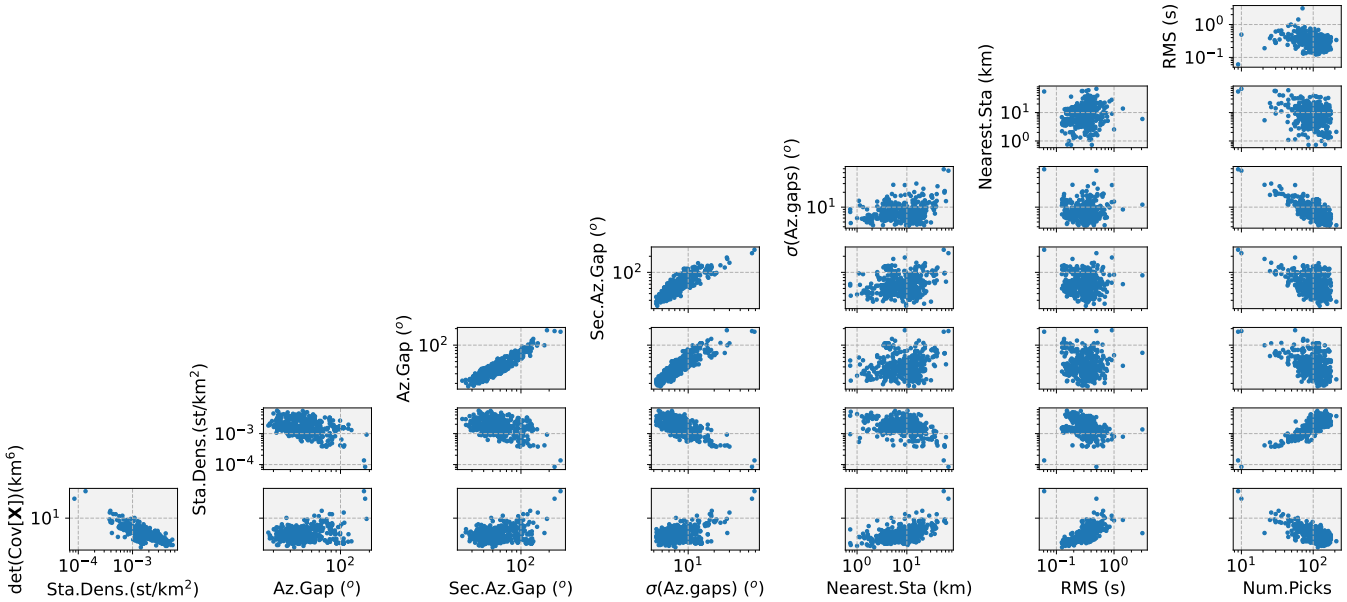

Figure 3 Pairwise dependency of the eight parameters used to define the Location Quality Score.

5. **Azimuthal uniformity (°):** Standard deviation of azimuthal gaps between consecutive stations.

6. $\det(Cov[\mathbf{X}])$ **(km^6):** Determinant of hypocentral location covariance matrix.

7. **RMS (s):** Root-mean-square of travel-time residuals at the maximum-likelihood hypocenter.

8. **Number of picks:** Total number of P- and S-phase picks used.

A well-distributed station network is key for accurate event locations (e.g., Bondár et al., 2004; Bondár and McLaughlin, 2009; Havskov et al., 2012; Tiira et al., 2016; Theunissen et al., 2018). However, traditional measures such as the primary azimuthal gap and distance to the nearest station alone are insufficient to comprehensively evaluate station geometry. Thus, we include three additional network-related parameters. The first is the station density, which penalizes e.g., configurations where, despite a very close station, the next closest station is considerably distant. The secondary azimuthal gap further enhances robustness, providing a more reliable measure of network geometry (Bondár et al., 2004). Additionally, azimuthal uniformity quantifies the deviation of the station distribution from an ideal, evenly spaced arrangement with gaps of $360°/N$, where N is the number of stations. Larger deviations from this optimal azimuthal geometry increase susceptibility to location biases (Bondár and McLaughlin, 2009).

The covariance matrix determinant combines all uncertainties of the hypocentral parameters. The diagonal elements represent the location variance, while the off-diagonal elements describe the orientation and shape of the error ellipse (Havskov et al., 2012). The RMS captures phase pick accuracy and the overall travel-time fit, assessing simultaneously the suitability of the velocity model. Finally, although a higher number of picks typically improves the solutions, their spatial distribution is critical—uneven distributions may weaken the contribution of the number of picks to location quality.

Figure 3 illustrates pairwise relationships among the eight LQS parameters, revealing strong interdependencies (e.g., between station density and number of picks, or azimuthal uniformity and azimuthal gaps). Notably, parameters like nearest station and azimuthal gaps alone show limited influence on the covariance determinant (location uncertainties). The figure also reveals outliers, which we accounted for during parameter normalization (see Supplemental Figure S1 for parameter histograms). Strongly skewed parameters (covariance determinant, station density, nearest station, RMS, number of picks) were logarithmically transformed before normalization to manage their skewness. Such transformation compresses the extreme high values and spreads out the lower range. Normalization followed robust statistics (Equation 1):

$$X_{\log,\mathrm{Norm}} = \frac{X_{\log} - X_{\log,5th}}{X_{\log,95th} - X_{\log,5th}}, \tag{1}$$

where $X_{\log} = \log(X + \epsilon)$ and ϵ is a small constant that avoids taking the logarithm of very small values of X. Outliers were clipped at the 5th and 95th percentiles. This ensured that each normalized parameter X_{Norm} ranges between 0 and 1, thus contributing equally to the LQS. A variation was applied during the normalization of the RMS parameter. In our dataset, we observed anomalously small RMS values that paradoxically corresponded to poor-quality locations, which had a very small number of picks (see Figure 3). This apparent contradiction arises because RMS primarily reflects the travel-time residual fit rather than overall location quality. Consequently, a small number of picks can artificially produce extremely low RMS values without indicating a better location solution (Husen and Hardebeck, 2010). Such anomalously small RMS values overlapped

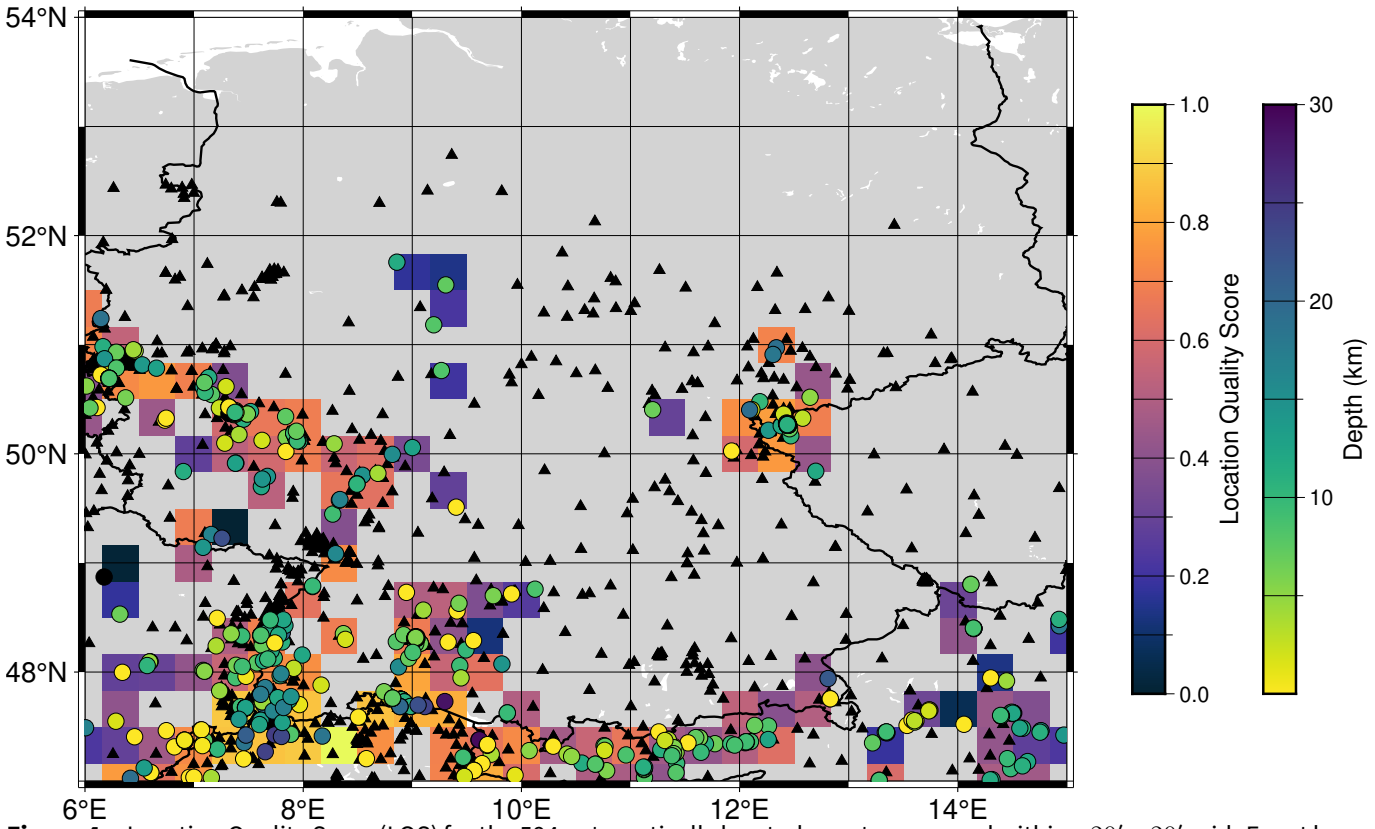

Figure 4 Location Quality Score (LQS) for the 594 automatically located events, averaged within a $20' \times 20'$ grid. Event locations are color-coded by automatically estimated depths (values clipped at 30 km). Black triangles denote seismic stations, including temporary deployments, used for phase picking during automatic location.

with genuinely good low-RMS solutions but tended to cluster around very low station densities, which allowed us to identify and exclude them prior to normalization. After normalizing the RMS, these identified anomalous outliers were clipped to a value of 1. This same approach against spuriously low RMS is applied in the near-real-time system. Guardrail thresholds, derived from the event dataset used in this study, are presently fixed but will be periodically re-evaluated—and, if warranted, regionally adapted—to maintain conservative behavior as more automatic locations accumulate. Another exception is the nearest station parameter, for which we explicitly set the lower and upper normalization bounds at 10 and 30 km, respectively, guided by the Ground Truth (GT) criteria in Bondár et al. (2004); Bondár and McLaughlin (2009). These criteria define standards commonly used in global and regional bulletins to classify the reliability of event locations.

Primary and secondary azimuthal gaps and azimuthal uniformity exhibited less skewed distributions. The parameter spread was not extreme for these three parameters and their scales remained manageable. For this reason, we normalized these parameters directly without logarithmic transformation. For the azimuthal gap, meaningful values existed above the 95th percentile (approximately $92°$). For instance, events in the West Bohemia/Vogtland region often exhibit relatively large azimuthal gaps of about $130°$ due to sparser station coverage on the Czech side, yet these locations maintain high quality. The epicenter accuracy criteria for local networks established by Bondár et al. (2004) recommend primary azimuthal gaps less than $110°$ and secondary azimuthal gaps less than $160°$. Such strict criteria aim to prevent contamination of their high-quality location sets (Bondár and McLaughlin, 2009). Based on our practical monitoring experience and well-documented practices, however, we heuristically set a more lenient upper boundary of $140°$ for the primary azimuthal gap. For the secondary azimuthal gap, we adhered to the recommended $160°$ upper limit.

We computed the LQS as a weighted sum of our eight normalized parameters (Equation 2):

$$\begin{aligned}\mathrm{LQS} = {} & 0.35(1 - \mathrm{Det.Cov.X_{Norm}}) + 0.125\ \mathrm{Sta.Dens_{Norm}} \\ & + 0.125(1 - \mathrm{Az.Unif_{Norm}}) + 0.05(1 - \mathrm{Az.Gap_{Norm}}) \\ & + 0.1(1 - \mathrm{Sec.Az.Gap_{Norm}}) + 0.1(1 - \mathrm{RMS_{Norm}}) \\ & + 0.1(1 - \mathrm{Near.Sta_{Norm}}) + 0.05\ \mathrm{N.Picks_{Norm}} \end{aligned} \quad (2)$$

By construction, LQS ranges from 0 to 1; higher values indicate better location quality. Weights assigned to each parameter reflect empirical considerations: the highest weight (0.35) is assigned to the covariance determinant, as low uncertainty implicitly indicates good station geometry and phase-pick quality. Station distribution received the next highest weight (0.25 total), equally split between station density and azimuthal uniformity (0.125 each). Secondary azimuthal gap received more weight (0.1) than primary gap (0.05) due to its robustness. RMS and nearest station were weighted moderately (0.1 each), while the number of picks received the lowest weight (0.05), being indirectly represented by other parameters.

4.4 Results

Figure 4 illustrates the automatic event locations computed with TieBeNN and their corresponding LQS values, averaged within a $20' \times 20'$ grid. Areas characterized by dense seismic station networks, such as northern Switzerland, southwestern Germany, the West Bohemia/Vogtland region, western Germany, and the Upper Rhine Graben, clearly exhibit high LQS values. These high scores highlight regions where automatic detections, including focal depth estimations, are expected to be highly reliable due to favorable network geometry. Conversely, sparse station coverage—notably in northern Germany and in Rhineland-Palatinate (western Germany, along the German–French border)—adversely affects automatic location quality.

The West Bohemia/Vogtland region is known for recurrent seismic swarms of relatively small magnitudes. During such swarms, events occurring only a few seconds apart pose a particular difficulty for automatic detection workflows. The coda waves of preceding events decrease the signal-to-noise ratio for subsequent events, complicating P-wave detections at greater distances. This frequently leads to missed phase picks or incorrect phase associations, where phases belonging to an earlier event are incorrectly associated with a later event. Consequently, even with an otherwise excellent station distribution and waveform data quality, automatic locations in this region occasionally exhibit higher uncertainties and increased RMS values, which significantly reduces the LQS.

Waveform data from the temporary AlpArray seismic network (AlpArray Seismic Network, 2015; Hetényi et al., 2018), publicly available via EIDA since April 2022, notably improved station coverage for events located in southern Germany and Austria, thereby positively impacting their respective LQS values.

4.5 Visualizing automatic location quality using the LQS metric

Figure 5 shows two illustrative examples of automatic event locations extracted from our results. Each row corresponds to one event, clearly contrasting in terms of LQS and location quality. The upper row illustrates a high-quality automatic location for an event in the West Bohemia/Vogtland region. The event attains an LQS of 0.83 because of its overall favorable station distribution and correspondingly low location uncertainties. In contrast, the lower row shows an event from the Eifel Volcanic Field in western Germany, which receives a significantly lower LQS of 0.19. This low score is mainly attributed to a combination of suboptimal station geometry, higher RMS values, and higher uncertainties compared to the upper-row example in the hypocenter estimation. It is noteworthy that the event's automatically determined focal depth (35.5 km) is atypically deep compared to most seismic events in Germany. However, this deeper depth is corroborated by manual locations reported by local seismic agencies (e.g., Department of Geosciences, Bensberg Observatory, University of Cologne, 2016; Erdbebendienst Südwest Baden-Württemberg and Rheinland-Pfalz, 2009; Federal Institute for Geosciences and Natural Resources, 1976). A detailed discussion regarding the relationship between automatically estimated depth uncertainties and actual event depths is provided in subsequent sections.

5 Cross-validating automatic focal depth estimation and the LQS metric

5.1 Comparison of automatic vs. manual focal depths

To cross-validate the automatic depth estimations obtained with TieBeNN, we first compared them to manually determined depths for events not included in our primary test catalog of 594 events (Section 4.1). The events used previously served to define the LQS metric and normalization boundaries for the eight parameters described in Section 4.3. The primary objective of this validation was to assess whether the LQS effectively reflects the quality of automatic locations beyond the initial dataset by comparing them with manually derived solutions.

For this validation, we retrieved event catalogs from three German seismic agencies: the Southwestern German Earthquake Service (EDSW, *Erdbebendienst Südwest*), the Seismological Network of Thuringia (TSN, *Thüringer Seismologisches Netz*), and the Bensberg Observatory (BENS). Additionally, events from the Swiss Seismological Service (SED, *Schweizerischer Erdbebendienst*) were also included. From each catalog, we randomly selected 50 manually located local tectonic events from the year 2024, all with magnitudes $\geq$1 and without fixed focal depths. Figures 6 and 7 show the epicenters of the events in the left panels and the automatic versus manual focal depths in the right panels.

For the EDSW catalog (Figure 6a), the automatic depth estimates generally align well with the manual depths. Events in areas with poorer station coverage (e.g., near the German–French border) received below-average LQS values of approximately 0.4. Some events along the Upper Rhine Graben, despite excellent station coverage, displayed elevated RMS values around 0.4 s, resulting in increased depth uncertainties. This could indicate that a simple 1D velocity model may inadequately explain observed travel-times in that area. Notably, two deeper events in Baden-Württemberg, manually located at 30 km depth, were automatically located at approximately 37 km depth, each with uncertainties around 3 km. Due to the uncertainties and limited station coverage, these events received low LQS values ($<$0.2), despite being in reasonable agreement with manual locations. Thus, our results suggest that LQS provides a more reliable quality assessment for shallower crustal events. Additionally, some events in the EDSW catalog overlapped spatially with the BENS catalog region (Figure 6b). In these cases (not explicitly shown), the EDSW manual depths better matched our automatic estimates compared to the depths of the same events from the BENS catalog.

Most automatic depth estimates for events in the BENS catalog (Figure 6b) were concentrated within

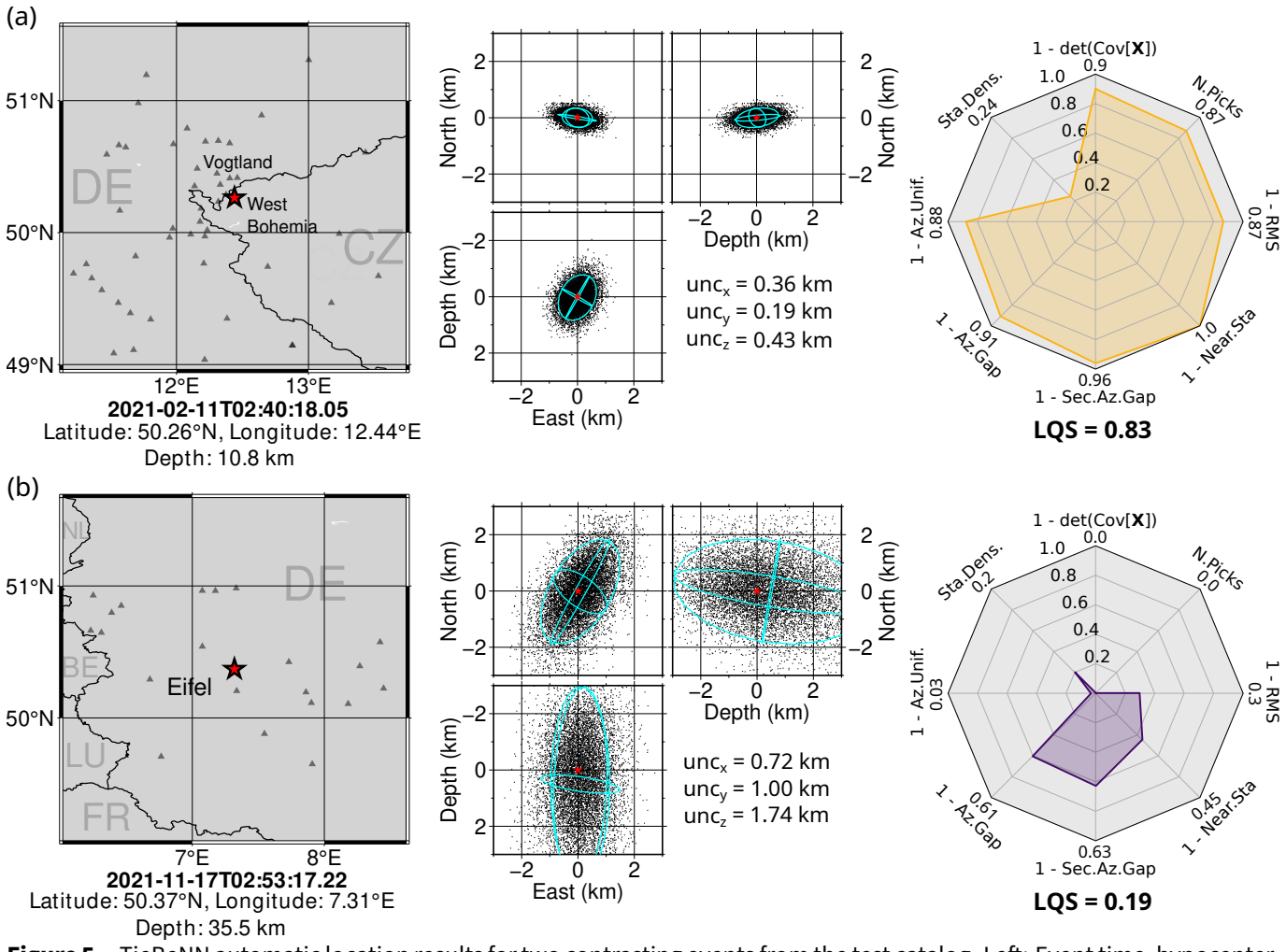

Figure 5 TieBeNN automatic location results for two contrasting events from the test catalog. Left: Event time, hypocenter, and associated station distribution. Center: Probabilistic location uncertainties from NonLinLoc, centered at each event's hypocenter. Black dots indicate grid samples with non-negligible posterior probability. Uncertainty components follow x = East, y = North, z = Depth. Right: Individual contributions of the eight normalized parameters toward the computed LQS.

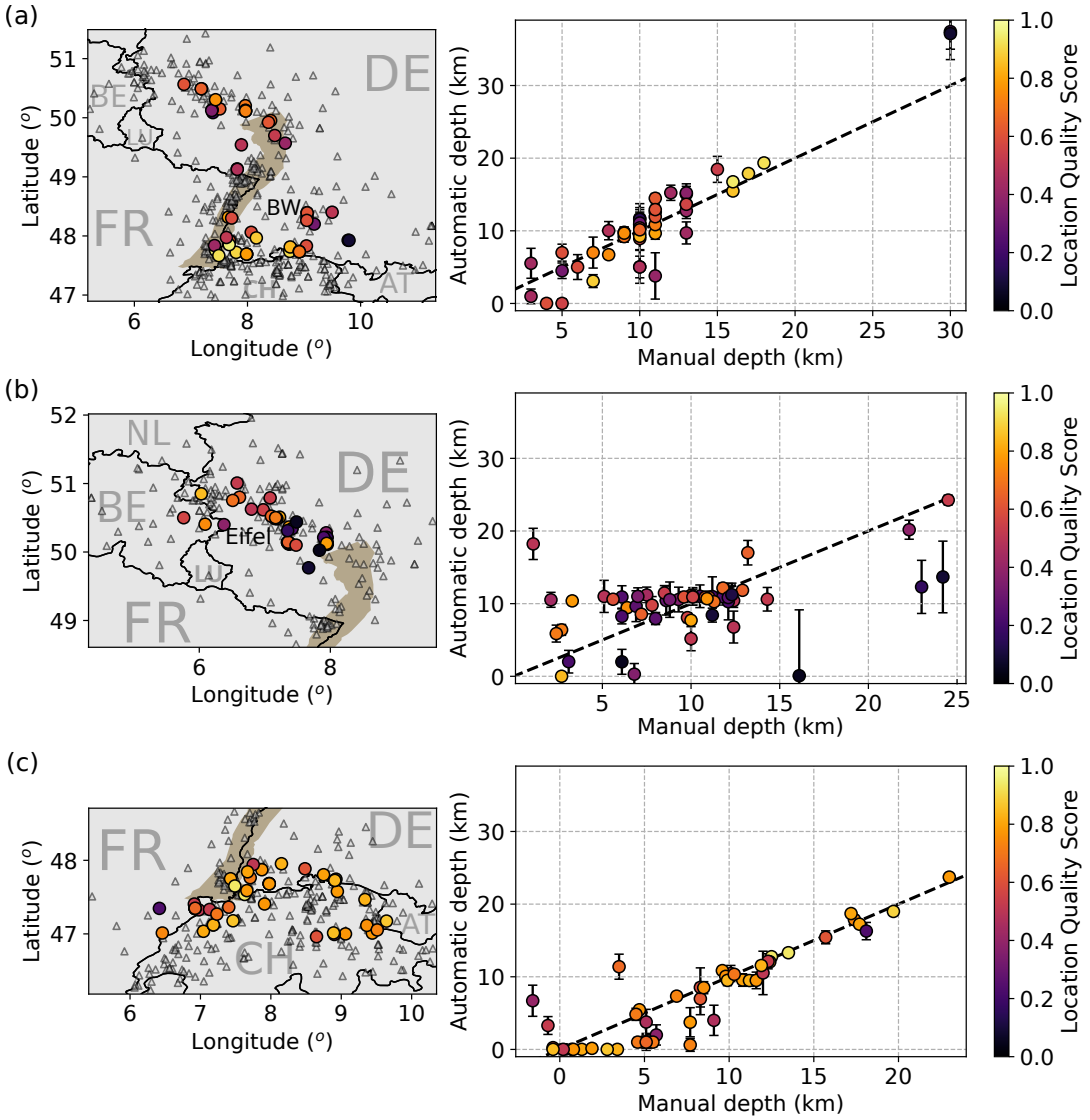

Figure 6 Focal depth comparison between automatic and manual localizations. Left: event epicenters (circles) and seismic stations used in automatic localization (black triangles). A shaded outline delineates the Upper Rhine Graben. Right: automatic versus manual focal depths, including uncertainty error bars. All events are color-coded by LQS. (a) Events in the EDSW catalog. BW: the German southwestern state of Baden-Württemberg, (b) events in the BENS catalog, and (c) events in the SED catalog. For BENS, a total of 55 events were automatically located, while 50 events were used for EDSW and SED (see text for details).

a narrow depth interval (8–12 km), whereas manual depths spread more uniformly across 2–14 km. To broaden the depth range for validation, five deeper events with depth >15 km were included. Automatic locations with high-quality picks and good station coverage achieved LQS values around 0.7. However, due to discrepancies in depth ranges between automatic and manual results, good LQS values did not always imply close depth agreement. Events with poor station geometry and higher uncertainties received LQS values <0.3, and events with extremely sparse coverage and very large uncertainties (> 3 km) were correctly assigned very low LQS values (<0.1). A notable temporal variation in station coverage affected the BENS catalog region, causing older events to achieve higher LQS values compared to the more recent events in 2024. This appropriately reflects decreased automatic location quality as station coverage deteriorated over time.

The automatic and manual focal depths showed notably good agreement for events in the SED catalog (Figure 6c). The dense seismic network in this region, coupled with the local velocity model by Diehl et al. (2021), resulted in highly reliable automatic depth estimates with correspondingly low uncertainties and consistently high LQS values (>0.7). A recurring observation, also noted in other catalogs, is that events manually located at shallow depths (<5 km) were often automatically located close to the surface (near 0 km depth). The LQS correctly identified events with sparser station networks (such as in parts of France and near the German–Swiss border) or those with higher RMS values due to significant discrepancies between observed and theoretical travel times, assigning them lower-quality scores.

Automatic locations for the TSN catalog events, despite sparse station coverage on the Czech side, generally showed very low depth uncertainties (Figure 7). However, the larger azimuthal gaps and lower station densities negatively influenced LQS values. Events from this catalog were located automatically using two velocity models: the standard layer-over-half-space model utilized at EdB (Figure 7a) and the 1D approximation of the 3D velocity model WB2012 from Růžek and Horálek (2013) (Figure 7b). The EdB model consistently produced automatic localizations with lower RMS values (average 0.12 s versus 0.33 s) and lower depth uncertainties compared to WB2012 (average 0.49 km vs. 0.65 km), resulting in higher LQS values. However, for an unusual seismic swarm near Nový Kostel—unusual in both its duration and northern location, partially extending into German territory—automatic locations using the WB2012 model better matched TSN manual depth estimates. The EdB model systematically overestimated depths for these swarm events, maintaining a consistent "depth excess" across all swarm events. Comparisons with the West Bohemia Local Seismic Network (WEBNET) catalog (Institute of Geophysics, Academy of Sciences of the Czech Republic, 1991) showed that their manually determined depths generally fell between the automatic results from the EdB model and the WB2012 model (Supplemental Figure S2). For events outside the swarm region, the EdB model provided better overall depth agreement with the manual catalog, coupled with lower uncertainties, compared to WB2012.

5.2 Robustness of LQS metric

In this section, we quantitatively assess how reliably the LQS metric represents the quality of automatically located events by evaluating the relative importance of each of the eight parameters comprising the score. For this purpose, we plotted the LQS values against each parameter, using the full set of 594 events described in Section 4.1. We compute Pearson and Spearman correlation coefficients. Both coefficients range from -1 to $+1$ (0 indicates no association; ± 1 indicates perfect negative/positive association). Pearson's coefficient quantifies the strength of a linear relationship for near-normally distributed variables, whereas Spearman's coefficient is rank-based, less sensitive to non-normality, and detects monotonic (not necessarily linear) trends. We report both because LQS–parameter relations can be skewed and non-linear, so the two measures are complementary (e.g., Schober et al., 2018). Scatterplots of the LQS versus each contributing parameter, along with correlation coefficients, are shown in Figure 8.

These analyses confirm the expected behavior of the LQS metric. As demonstrated repeatedly in Sections 4.5 and 5.1, events manually classified as well-located generally received high LQS values, whereas events with greater uncertainties and poorer station coverage received lower scores. Other parameters directly associated with improved depth constraints—such as a greater number of picks, low RMS residuals, and higher station density—show strong correlations with the LQS. However, we observed some spread within each scatterplot, indicating that no single parameter fully captures the complexity of event location quality.

The determinant of the location covariance matrix displayed an interesting relationship with LQS: a very low Pearson correlation coefficient (-0.13) contrasted with a high negative Spearman coefficient (-0.89). This indicates a strong nonlinear relationship, confirming that events with larger uncertainties are consistently ranked among those with lower LQS values. Thus, the covariance determinant is confirmed as a critical parameter within the LQS formulation.

A robust positive correlation (Pearson: 0.73) was observed between the LQS and the number of picks. Since this parameter is closely related to station density (Pearson: 0.67), these two parameters collectively reinforce the conclusion that higher-quality locations generally involve a greater number of reliable phase picks. Nevertheless, the observed scatter partly reflects scenarios with station networks that have many stations concentrated in small areas, such as dense arrays, leading to high pick counts without necessarily improving overall azimuthal coverage or location accuracy.

Both primary and secondary azimuthal gaps demonstrated moderate correlations with the LQS. Generally, smaller azimuthal gaps correlate with improved location quality, but several exceptions were apparent. Notably, a subset of events with moderate-to-high LQS values (around 0.5 or greater) occurs at large primary az-

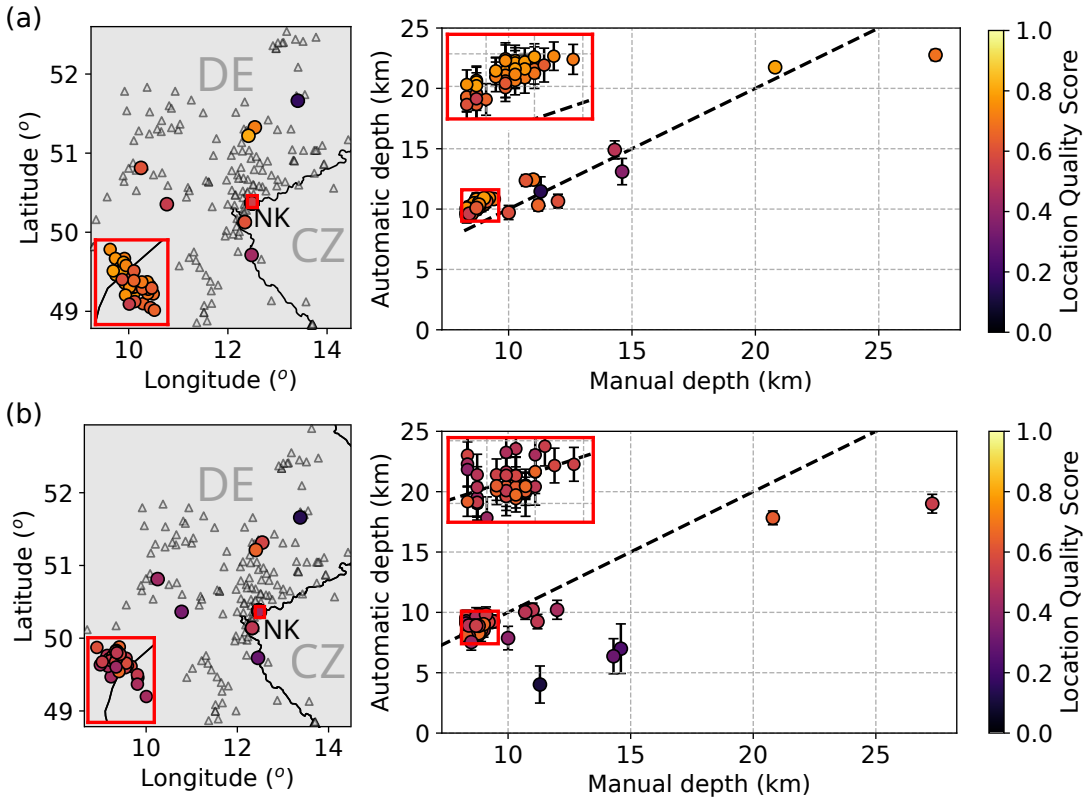

Figure 7 Same as Figure 6, but for 50 events in the TSN catalog, automatically located using (a) EdB's standard velocity model and (b) the WB2012 velocity model (Růžek and Horálek, 2013). The red box indicates a zoom-in on a dense seismic swarm in the region. NK: Nový Kostel region, West Bohemia.

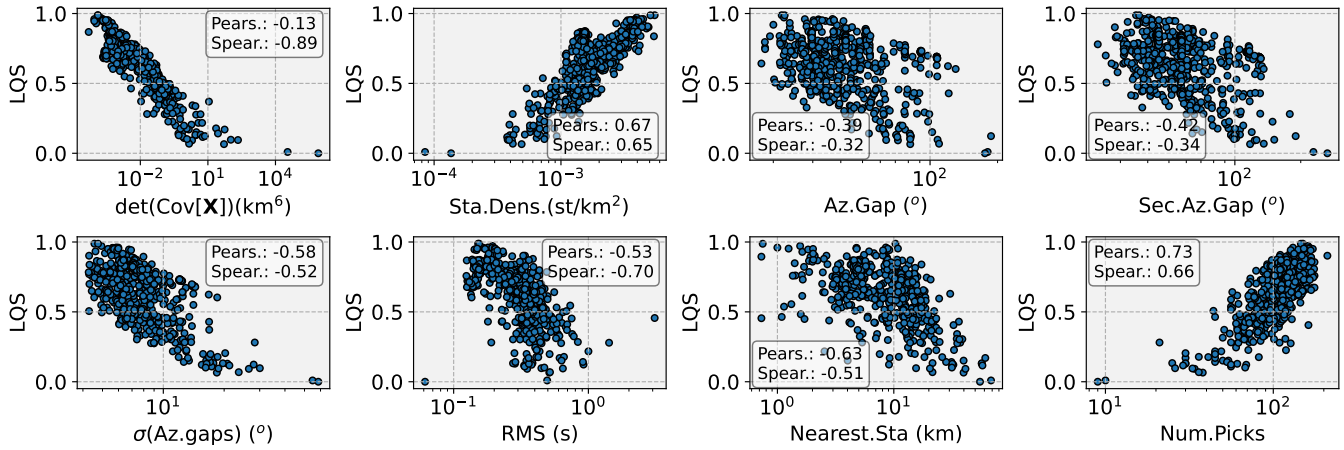

Figure 8 Correlation between the LQS metric and each of its eight contributing parameters (described in Section 4.3). Pearson and Spearman correlation coefficients are provided in each panel.

imuthal gaps of approximately 100–130°, many of them in the West Bohemia/Vogtland region. Despite relatively large azimuthal gaps due to sparser station coverage on the Czech side, these events achieved high-quality locations due to very low RMS residuals and small uncertainties. Compared to both azimuthal gaps, the azimuthal uniformity index showed a stronger negative correlation (Pearson: -0.58) and less spread, correctly reflecting its significant contribution to location accuracy, since uniformly distributed stations are essential for minimizing location biases.

The distance to the nearest station negatively correlated strongly with LQS (Pearson: -0.63), confirming that greater distances typically result in poorer locations. Nevertheless, some outliers illustrate that a single close station (less than 10 km away), though favorable according to criteria by Bondár and McLaughlin (2009), does not always guarantee high-quality locations if the next station is substantially farther away.

The RMS residual parameter negatively correlated strongly with LQS, emphasizing that smaller RMS values generally indicate better location quality. High RMS values are associated with poor automatic phase picks or inadequate velocity models that fail to explain travel times in complex geological settings. However, several noteworthy outliers appeared, primarily related to ex-

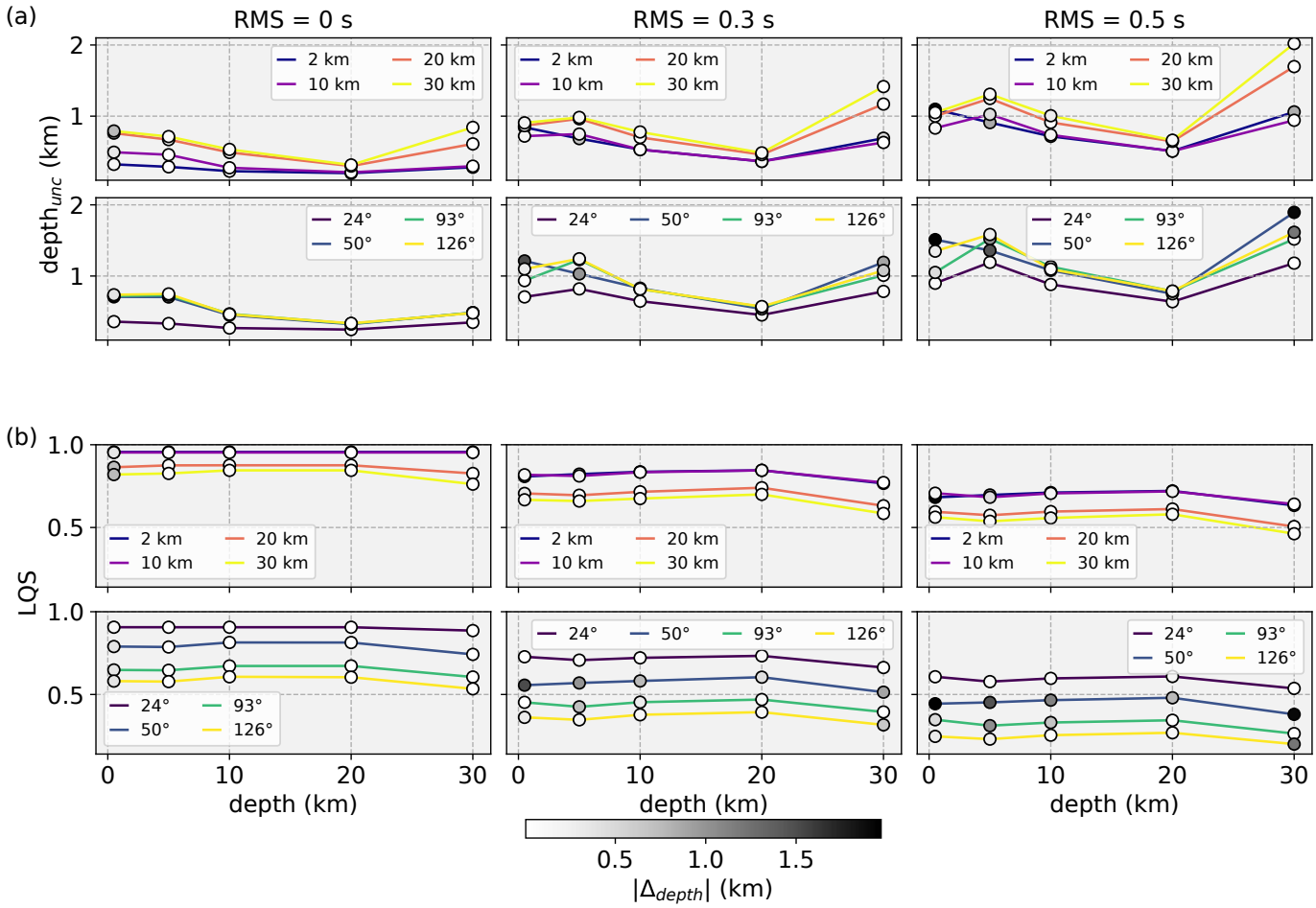

Figure 9 (a) Depth uncertainty of synthetic event locations under varying distances to the nearest station (top row) and primary azimuthal gaps (bottom row) for five synthetic focal depths (0.5, 5, 10, 20, and 30 km). Left panels show results without synthetic phase-picking errors, middle panels introduce a 0.3 s picking error, and right panels introduce a 0.5 s picking error. Scatter points are color-coded according to the absolute difference between the synthetic (true) and automatically estimated focal depths. (b) Effect of varying the distance to the nearest station (top) and primary azimuthal gap (bottom) on the LQS metric for the same five synthetic events. The layout and coloring follow the same scheme as described in (a).

tremely low RMS values. As discussed in Section 4.3, events with very few picks (approximately 25 or fewer) can misleadingly yield extremely low RMS values due to insufficient constraints. In these cases, identified by low station densities, we corrected the RMS parameter by assigning its maximum normalized value, preventing artificial inflation of the LQS. Conversely, two events with RMS values greater than 1 s corresponded to seismic swarm events occurring a few seconds after a previous event. Here, automatic phase pickers erroneously selected arrivals from the preceding event due to overlapping waveforms. These high RMS values reflect a recognized challenge in machine learning–based automatic picking rather than a flaw in the LQS metric. In such instances, the automatic locations remained accurate, resulting in relatively high LQS scores despite elevated RMS values.

5.3 Cross-validation of TieBeNN hypocenters and LQS with synthetic data

Since ground-truth tectonic events are unavailable within the German catalog, we tested TieBeNN's focal depth estimation and evaluated the behavior of the LQS metric using synthetic events, where the true focal depth is precisely known. Our main objectives were to determine how accurately TieBeNN could recover known depths under varying station coverage conditions and travel-time picking errors, and to examine whether our LQS metric sensitively reflects these variations. To achieve this, we created five synthetic events located in western Germany at 50.35°N, 7.38°E, with predefined focal depths of 0.5, 5, 10, 20, and 30 km. This region was selected due to its dense seismic station network, enabling us to systematically test different station configurations and coverage scenarios.

We used NonLinLoc to compute synthetic travel times, simulating seismic wave propagation from each synthetic hypocenter through the crust, modeled using the 1D seismic velocity profile from Reamer and Hinzen (2004). These travel times were calculated to real station locations, supplemented by additional synthetic stations randomly positioned to generate various realistic station coverage scenarios. Supplemental Figure S3 shows the epicenter and spatial distribution of real and synthetic stations used in this analysis.

Figure 9a presents the depth uncertainties for the automatically located synthetic events under various sta-

tion configurations and phase-picking errors. Our analysis demonstrates that the accuracy and uncertainty of focal depth estimations depend not only on station coverage and picking errors but also significantly on the true focal depth itself. Very shallow events typically show higher sensitivity to travel-time variations at similar epicentral distances compared to deeper events (Havskov et al., 2012). Our tests generally reveal larger uncertainties and greater differences between synthetic (true) and estimated focal depths for the shallowest synthetic event (0.5 km), while uncertainties are lowest for events between approximately 5 and 20 km depth. As expected, scenarios characterized by dense station coverage, proximity of stations to the epicenter, and low phase-picking errors produce depth estimations with minimal uncertainty and depth discrepancies across all tested focal depths. Although absolute depth differences between the synthetic and automatically estimated depths typically remained below 2 km—partially because our simulation retained a relatively dense station network despite selectively removing nearby stations—there is a consistent and gradual increase in depth uncertainty as the distance to the nearest station and primary azimuthal gap increase. Events at 30 km depth appear particularly sensitive to both poor station coverage and elevated picking errors, likely due to seismic wave raypaths becoming increasingly horizontal at greater depths, thus reducing depth sensitivity.

Regarding the behavior of the LQS metric, Figure 9b consistently captures the impact of the various tested station configurations and travel-time errors on automatic location quality. Scenarios characterized by robust station coverage and accurate phase picks consistently yield higher-quality automatic locations, independent of focal depth. Conversely, increased picking errors, greater distances to the nearest station, and larger azimuthal gaps systematically degrade location quality, resulting in reduced LQS values. Notably, our synthetic test results provide insight into why event locations in the West Bohemia/Vogtland region, despite having relatively large azimuthal gaps around 120°, maintain good-quality scores (LQS around 0.6). This is attributed primarily to excellent automatic phase picks—enabled by high signal-to-noise ratios—and dense station coverage on the German side of the border, factors that compensate effectively for the suboptimal azimuthal gap.

Furthermore, depth constraint inherently depends on ray coverage and directions at the hypocenter; thus, automatically located events at focal depths near or beyond 30 km inherently exhibit lower LQS values due to unfavorable ray geometry. Given the scarcity of seismic events at depths greater than approximately 30 km in Germany compared to shallower seismicity, our LQS metric and normalization parameters were primarily calibrated on shallower crustal earthquakes, reflecting the majority of the German seismic catalog. Consequently, for events deeper than around 30 km, our results suggest that either the interpretation of the LQS must be reconsidered—acknowledging it is optimized for shallower crustal events—or that deeper automatic locations should be flagged as inherently less reliable, recognizing the primary calibration of the LQS for typical shallow crustal seismicity.

6 Discussion

By leveraging community-standard, community-supported, and state-of-the-art seismic monitoring tools, including advanced machine learning–based models, we enhanced the capabilities of the automatic real-time local earthquake monitoring system at EdB. Previously, the system provided sensitive detections, but locations were limited to fixed grid-point coordinates without depth estimations. The enhanced system presented here delivers more precise probabilistic hypocenter estimations, allows the use of dedicated local velocity models during the location process, and introduces the Location Quality Score (LQS) as a visually intuitive and quantitative metric. This metric provides valuable support for analysts in rapidly assessing the reliability of automatic locations.

6.1 Automatic probabilistic depth estimation

To evaluate the automatic hypocenter estimation system, we computed focal depths for 594 local tectonic events. Wherever available, we utilized local 1D velocity models during event location to reduce discrepancies between observed and theoretical travel times. Several high-resolution 3D velocity models exist for specific regions of Germany and neighboring countries, and TieBeNN readily supports their use in automatic localization with NonLinLoc. Examples include the 3D velocity model by Diehl et al. (2021) for events in the Central Alps region, the WB2012 model by Růžek and Horálek (2013) for the seismically active West Bohemia/Vogtland region, and the model by Lengliné et al. (2023) for the Upper Rhine Graben. While robustly constructed 3D velocity models typically result in more accurate depth estimations due to properly resolved lateral velocity variations, their applicability to real-time monitoring must be carefully evaluated. Travel-time calculations in NonLinLoc significantly increase computational requirements when employing 3D models, potentially extending computation time from a few seconds (using 1D models) to several minutes, depending on the number of phase picks involved. Given this practical limitation, our results suggest that adequately dense station coverage coupled with suitable 1D velocity models usually suffices to achieve accurate hypocenter estimations for most crustal seismicity monitored by EdB. Nonetheless, dedicated 3D velocity models have demonstrated substantial benefits for monitoring shallow induced seismicity associated with geothermal reservoirs or hydrocarbon extraction sites, significantly reducing hypocentral uncertainties compared to simpler 1D or regional velocity models, as reported, for example, by Uta (2017) in northern Germany.

In this study, pre-trained phase-picking models provided by SeisBench performed well in detecting local seismic phases within our dataset; thus, we did not

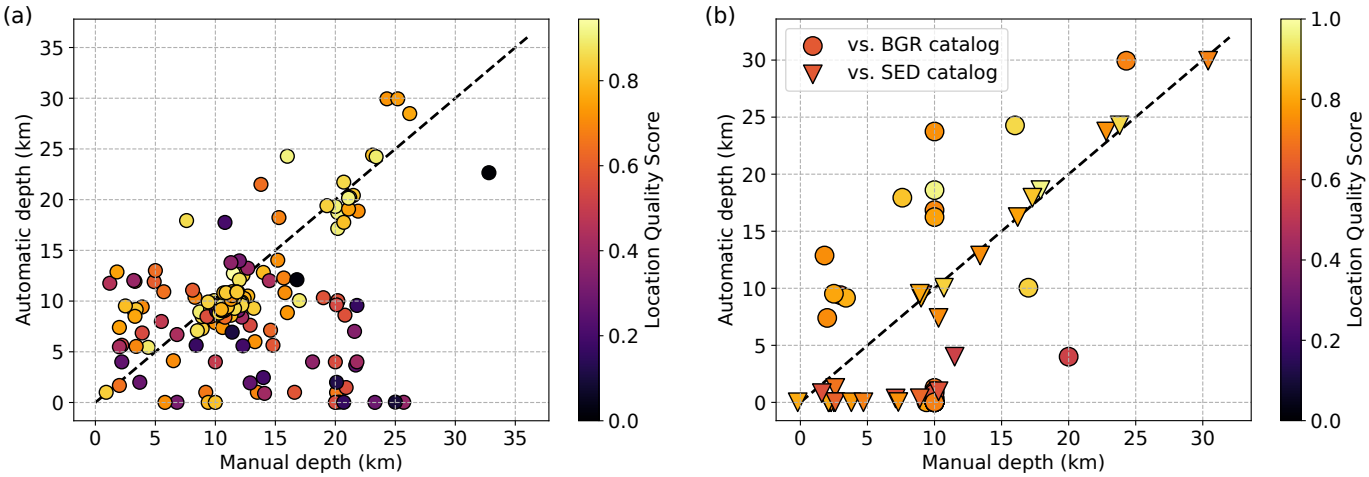

Figure 10 (a) Comparison of manually estimated (unfixed) focal depths at EdB versus automatically estimated depths for events in our 594-event catalog, with points colored according to their respective LQS values. (b) Comparison of automatic versus manual depth estimations for events from the same catalog located within Switzerland. Circles indicate depths from EdB manual locations, including fixed-depth estimations visible as round-number depths (e.g., 10 or 20 km). A minimum absolute difference of 5 km exists between manual and automatic estimations. Inverted triangles show depths for the same events extracted from the SED catalog. Pairs of EdB and SED locations for the same event can be identified by their identical automatic depth (y-axis) and LQS color.

re-train PhaseNet specifically for our region. However, one significant challenge encountered during automatic phase picking was handling closely spaced overlapping events, a scenario frequently observed during seismic swarms, particularly in the West Bohemia/Vogtland region. Recent internal tests at EdB indicate that re-training PhaseNet on local phase picks from events recorded by the German Regional Seismic Network (GRSN) significantly enhances the number of correct detections, especially for S-wave picks (A. Steinberg, pers. comm., 2025). Future improvements to a specialized phase picker for German seismicity could potentially employ data augmentation methods similar to those described by Armstrong et al. (2023). Their approach synthesizes waveforms with nearly coincident P-wave arrivals to improve automatic phase-picking performance during periods of intense seismic activity in the Yellowstone volcanic plateau. While these enhancements represent promising directions for further research, they extend beyond the immediate scope of the current study.

A comparison between manually and automatically estimated focal depths for the subset of events in our test catalog with non-fixed manual depths is presented in Figure 10a. To complement the scatter, a histogram of signed depth differences is provided in Supplementary Figure S4. Overall, a clear agreement is observed between the two sets of depths, but several noteworthy exceptions exist. Events with very low LQS values occasionally still show good depth agreement. To examine whether epicentral uncertainty contributes to reduced LQS in such cases, we analyzed LQS as a function of the absolute depth misfit, $\Delta z = |z_{\text{auto}} - z_{\text{manual}}|$, and the vertical uncertainty (Supplementary Figure S5). The few small-Δz cases with low LQS exhibit elevated horizontal and/or vertical uncertainties, consistent with weaker epicentral constraint. However, because sparse station coverage usually inflates horizontal and vertical uncertainties jointly, LQS—through the covariance determinant—primarily reflects this combined effect rather than cleanly separating horizontal from vertical uncertainties. In some instances, a weakly peaked posterior probability distribution can also yield a depth close to the reference by chance while the overall uncertainty (and thus LQS) remains large. Additionally, our automatic depth estimations sometimes show reduced accuracy for events manually located at shallow depths (up to ∼10 km), despite relatively good LQS scores (>0.5). Moreover, events manually located at depths greater than 20 km generally exhibit higher discrepancies relative to automatic depths. These events are typically assigned significantly shallower depths by our automatic system and receive consistently low LQS scores (<0.4). Such discrepancies likely occur due to limited station coverage around these event epicenters. If the true focal depth indeed exceeds 20 km, greater uncertainties—and correspondingly lower LQS values—are expected, as demonstrated by our synthetic tests in Section 5.3. It is important to emphasize that our 594-event catalog lacks ground-truth events. Consequently, event depths often differ across local seismic catalogs for the same events. An illustrative example is shown in Figure 10b, comparing automatic depths with manually determined depths from EdB and the Swiss Seismological Service (SED) for the same set of events located within Switzerland. For completeness, a comparison that includes events with fixed manual depths is provided in Supplementary Figure S6. As expected, most fixed-depth cases cluster at the cataloged fixed value (commonly 10 km) and correspond to automatic locations tending toward lower LQS, although a subset attains high LQS where network coverage was improved in the automatic location (e.g., by adding stations such as those from the Raspberry Shake Seismic Network; Raspberry Shake, S.A., 2016). In a few cases, strong regional prior information can also yield good agreement

between automatic and fixed depths (e.g., the event located at >30 km depth noted in Figure 5).

6.2 Location quality assessment using the LQS metric

The Location Quality Score (LQS), defined as a weighted sum of eight normalized parameters, is conceptually similar to the Ground Truth (GT) criteria established by Bondár et al. (2004); Bondár and McLaughlin (2009). It was specifically designed to facilitate rapid evaluation of location quality during real-time monitoring of local tectonic seismicity by the EdB. We established the normalization of the individual parameters using robust statistics derived from 594 automatically located events spanning a period of three years, with only a few exceptions based on established literature values. To visually convey the automatic event localization quality to the analysts on duty, we implemented a radar plot as illustrated in Figure 5. This intuitive visualization immediately identifies the parameters driving either high or low scores, such as small uncertainties, suboptimal station coverage, or poor travel-time fits between observed and theoretical phases.

Despite our parameter weighting being heuristic and grounded in empirical experience, the LQS clearly captures the critical role of station distribution in achieving accurate automatic event locations (see Figure 4). Areas with dense and well-distributed seismic networks typically yield accurate hypocenter estimations with low uncertainties and correspondingly high LQS values ($\geq$0.7). Under these conditions, automatic localizations considerably reduce the daily monitoring load for analysts. A practical example of this benefit is evident in the automatic localization of seismic swarms in the West Bohemia/Vogtland region, provided the events do not significantly overlap in time.

Some regions with apparently good station coverage and reliable automatic phase picks occasionally produce nevertheless slightly lower LQS scores (between approximately 0.6 and 0.7). In these instances, elevated RMS values (typically around 0.3–0.4 s) indicate poor agreement between observed and theoretical travel times. Such cases were frequently identified in events located along the Upper Rhine Graben. Here, the LQS accurately reflects and clearly communicates the specific factor that adversely affects the quality of automatic localizations.

Regions characterized by sparse station coverage, such as northern Germany, pose a different scenario. Automatic localizations in these areas might yield focal depths similar to manually derived estimations, but with great depth uncertainties. The LQS correctly identifies these cases as inherently unreliable by assigning correspondingly low scores, often as low as approximately 0.2.

A recent revision of the GT event list criteria (Gallacher et al., 2025) relaxes the requirement of a nearest station within 10 km when sufficient P–S pairs are available. It also replaces the azimuthal-gap proxy for azimuthal coverage with the Cyclic Polygon Quotient (CPQ), a normalized measure computed from event–station azimuths on the unit circle. Our current LQS already downweights the nearest-station distance relative to station density and azimuthal uniformity, so it is consistent in spirit with the relaxed near-station rule. Because the azimuthal-gap parameter showed comparatively weak correlation with LQS in our tests (see Figure 8), CPQ appears to be a promising compatible enhancement to characterize azimuthal coverage more comprehensively. In the same spirit, and given the role of S phases for depth resolution, we will evaluate replacing the "number of picks" parameter with the "number of stations carrying both P- and S-picks." Total pick count already shows strong correlation with LQS (see Figure 8), but a P–S-pair measure may better capture depth-resolving power and reduce redundancy with station density. We plan to assess this variant together with CPQ in future work.

Most local seismicity in Germany occurs at crustal depths ranging roughly from 2 to 20 km, with an average Moho depth beneath Germany lying between 27 and 30 km (Grad and Tiira, 2009; Michailos et al., 2023). Despite this, some automatic localizations in our study produced depths exceeding 30 km (e.g., Figures 5b and 6a). Even deeper seismic events, reaching depths around 40 km and thus clearly below the Moho discontinuity, are reported in relocated seismicity catalogs for the Northern Rhine area (Hinzen et al., 2021). In our dataset of 594 events, only one event was located below 30 km depth, occurring in the Eifel region, a well-documented location for deep, low-frequency earthquakes (Hensch et al., 2019). While the automatic depth estimation of this specific event aligned well with manual depth estimations at both EdB and other local seismic agencies, such as the Seismological Station Bensberg, it received a low LQS value of only 0.19. Given that the parameter normalization boundaries defining our metric primarily represent shallow crustal events due to their prevalence in the German catalog, our results suggest that the LQS performs best when describing the quality of shallower seismic events.

Our synthetic tests presented in Section 5.3 further illustrate this point, showing how depth uncertainties and the resulting LQS for events deeper than approximately 30 km are particularly sensitive to inadequate station coverage—much more so than for shallower events under identical network configurations. Poor phase-pick accuracy further exacerbates this effect, resulting in large depth uncertainties and correspondingly reduced LQS values. Fundamentally, these results reflect the inherent decrease in depth resolution for seismic events located near or below the Moho boundary. In Germany, where deep events represent statistical outliers and reliable ground-truth depth data are scarce, this limitation is particularly relevant. While developing an alternative metric specifically for deep seismic events might seem desirable, the absence of robust reference data for calibration at such depths complicates this approach. Practically, we therefore recommend flagging automatic locations deeper than approximately 30 km as inherently less reliable, explicitly acknowledging that the current LQS calibration targets predominantly shallow crustal seismicity.

Ideally, defining clear LQS thresholds to distinguish reliably versus poorly located events would allow analysts at EdB to rapidly prioritize events requiring manual verification, while confidently relying on automatic reports for high-quality locations in real-time monitoring. However, establishing fixed thresholds for the LQS metric is problematic for several reasons. Firstly, extreme LQS values—approximately above 0.8 or below 0.2—clearly indicate good-quality locations or events requiring careful manual scrutiny, respectively. Yet, statistically, even locations with low LQS values may occasionally yield focal depths coincidentally close to manually derived estimates, albeit with large uncertainties. Conversely, events with seemingly good LQS scores can sometimes significantly differ from manual depths. Secondly, as demonstrated in synthetic tests, the LQS's sensitivity to depth varies significantly depending on the specific network geometry and depth of the event. Any fixed threshold would likely require adjustments for deeper events, or these deeper events could instead be automatically flagged as requiring additional scrutiny. Finally, any potential thresholds should be viewed as provisional guidelines rather than rigid boundaries. Analysts might consider adaptive thresholds, dynamically adjusted according to real-time network conditions or region-specific calibrations, especially if significant variations occur in station geometry or ambient seismic noise. Such adaptive strategies would enhance operational flexibility and accuracy over time, allowing the automatic monitoring system to respond appropriately to evolving conditions and data availability.

7 Conclusions

We expanded the capabilities of the German Federal Seismological Survey's automatic local earthquake monitoring by integrating TieBeNN into the real-time workflow. TieBeNN leverages community-supported, state-of-the-art machine-learning tools to provide automatic probabilistic depth estimation for local seismic events, introducing the Location Quality Score (LQS) metric for quick assessment of location reliability. This integration augments the existing system, enabling near-manual performance while operating continuously and rapidly.

Tests on a catalog of 594 tectonic events show that TieBeNN's automatic locations closely match those determined by human analysts in well-instrumented areas with dense, uniform station coverage. In such regions, most hypocenters were located within a small distance of their manually derived locations, while notable discrepancies appeared in regions with sparse station networks or unmodeled velocity complexities. These cases are readily identified by low LQS values, which reflect the higher travel-time residuals and uncertainties, alerting analysts to interpret such results with caution.

Cross-validation against independent events and external catalogs further confirms TieBeNN's reliability. Overall, automatic focal depths generally differ from manual solutions by only a few kilometers for typical crustal earthquakes. The LQS correlates strongly with key quality indicators—such as the number of phase picks, station density, location uncertainties and travel-time residuals—consistently highlighting well-constrained solutions versus those with large uncertainties or poor station geometry.

Despite these positive results, certain scenarios remain challenging. Earthquake swarms with events only seconds apart can confuse phase-picking and association algorithms, occasionally producing mislocated events and higher residuals despite otherwise good coverage. Additionally, the location quality of events below the Moho depth (deeper than approximately 30 km) tends to be underestimated, yielding low LQS values, despite a good agreement with diverse independent manual locations. These outcomes align with synthetic tests that show how deeper hypocenters are more sensitive to errors in velocity models or station geometry.

Integrating TieBeNN into routine operations offers significant practical benefits. Beyond epicenter and origin time, the system automatically generates a depth estimate and quality score, substantially reducing the daily workload of on-duty seismologists. Events with high LQS can be trusted and reported automatically, allowing analysts to concentrate on lower-scoring (and likely lower-quality) solutions. In effect, the LQS provides rapid triage of automatic results, guiding which events need human intervention.

Looking ahead, targeted improvements could further boost accuracy and robustness at EdB. Although current 1D velocity models generally suffice for most local events, incorporating high-resolution 3D models in complex regions (e.g., the Alps or the Upper Rhine Graben) may reduce depth biases, although with higher computational cost. Similarly, retraining or fine-tuning the phase-picking models on region-specific data could enhance performance during earthquake swarms or overlapping sequences. Finally, the LQS itself may be refined as more events—particularly deeper earthquakes—are recorded, potentially adopting depth-dependent scoring or region-specific calibration to better evaluate atypical scenarios.

In summary, the upgraded monitoring system with TieBeNN markedly improves near-real-time detection and analysis of local seismic events in Germany. It narrows the gap between automated processing and expert analysis, providing faster, high-quality earthquake information to decision-makers while maintaining a robust standard of accuracy and reliability.

Acknowledgements

The authors thank the Editor Hongyu Sun and two anonymous reviewers for their constructive comments, which helped improve the clarity and quality of this manuscript. The authors gratefully acknowledge the waveform data retrieved from seismic networks accessible via FDSN and EIDA web services. The complete list of networks used in this study is provided below. Figures were created using Matplotlib (Hunter, 2007), Seaborn (Waskom, 2021) and the Generic Mapping Tools (GMT; Wessel et al., 2019; Tian et al., 2024).

Data and code availability

TieBeNN is available at https://doi.org/10.5281/zenodo.15825093 for reproducible citation, and at https://github.com/Cthuulhaa/tiebenn.git for installation instructions and issue reporting. The catalog of 594 events automatically located by TieBeNN between 2021–2023, as well as the automatic locations shown in Figures 6 and 7, are available at https://doi.org/10.5281/zenodo.17684466. The internal EdB event catalog, which includes events with magnitudes $M < 2$ used to test TieBeNN, is currently not publicly accessible (an FDSN event service is under development). We retrieved additional event data from the following sources:

- EDSW catalog from https://www.lgb-rlp.de/fachthemen-des-amtes/projekte/landeserdbebendienst-rheinland-pfalz/erdbebenereignisse-lokal
- TSN catalog from https://antares.thueringen.de/cadenza/index.xhtml
- BENS catalog from https://www.seismo.uni-koeln.de/catalog/2024.htm
- SED catalog from http://www.seismo.ethz.ch/en/earthquakes/switzerland/all-earthquakes

Waveform data used for testing and validation were retrieved via the European Integrated Data Archive (EIDA) and the GEOFON Data Center, from the following seismic networks: 1D (https://doi.org/10.14470/6Q705117); 2D (https://www.fdsn.org/networks/detail/2D_2021); 4C (https://doi.org/10.14470/9P982225); 8D (https://doi.org/10.12686/sed/networks/8d); 9E (https://doi.org/10.12686/sed/networks/9e); 9S (https://doi.org/10.12686/SED/NETWORKS/XP); AM (https://doi.org/10.7914/SN/AM); BE (https://doi.org/10.7914/SN/BE); BQ (https://doi.org/10.7914/SN/BQ); BW (https://doi.org/10.7914/SN/BW); C4 (https://doi.org/10.12686/sed/networks/c4); CH (https://doi.org/10.12686/sed/networks/ch); CZ (https://doi.org/10.7914/SN/CZ); FO (https://doi.org/10.15778/resif.fo); FR (https://doi.org/10.15778/RESIF.FR); G (https://doi.org/10.18715/GEOSCOPE.G); G2 (https://doi.org/10.12686/SED/NETWORKS/G2); GE (https://doi.org/10.14470/TR560404); GQ (https://www.fdsn.org/networks/detail/GQ); GR (https://doi.org/10.25928/mbx6-hr74); GX (https://www.fdsn.org/networks/detail/GX); HS (https://doi.org/10.7914/SN/HS); HU (https://doi.org/10.14470/UH028726); IV (https://doi.org/10.13127/sd/x0fxnh7qfy); LE (https://doi.org/10.7914/SN/LE); MN (https://doi.org/10.13127/sd/fbbbtdtd6q); NI (https://doi.org/10.7914/SN/NI); NL (https://doi.org/10.21944/e970fd34-23b9-3411-b366-e4f72877d2c5); OE (https://doi.org/10.7914/SN/OE); OX (https://doi.org/10.7914/SN/OX); RD (https://doi.org/10.15778/RESIF.RD); RF (https://doi.org/10.7914/SN/RF); RN (https://doi.org/10.7914/SN/RN); S (https://doi.org/10.12686/SED/NETWORKS/S); SI (https://www.fdsn.org/networks/detail/SI/); SL (https://doi.org/10.7914/SN/SL); ST (https://doi.org/10.7914/SN/ST); SX (https://doi.org/10.7914/SN/SX); TH (https://doi.org/10.7914/SN/TH); YD (https://doi.org/10.7914/SN/YD_2020); YV (https://doi.org/10.7914/brb1-cf34); Z3 (https://doi.org/10.12686/alparray/z3_2015); ZB (https://doi.org/10.14470/MO7576467356).

Competing interests

The authors have no competing interests.

References

AlpArray Seismic Network. AlpArray Seismic Network (AASN) temporary component, 2015. doi: 10.12686/alparray/z3_2015.

Armstrong, A. D., Claerhout, Z., Baker, B., and Koper, K. D. A deep-learning phase picker with calibrated Bayesian-derived uncertainties for earthquakes in the Yellowstone Volcanic Region. *Bulletin of the Seismological Society of America*, 113(6):2323–2344, 2023. doi: 10.1785/0120230068.

Arrowsmith, S. J., Trugman, D. T., MacCarthy, J., Bergen, K. J., Lumley, D., and Magnani, M. B. Big Data Seismology. *Reviews of Geophysics*, 60:e2021RG000769, 2022. doi: 10.1029/2021RG000769.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., and Wassermann, J. ObsPy: A Python toolbox for seismology. *Seismological Research Letters*, 81(3):530–533, 2010. doi: 10.1785/gssrl.81.3.530.

Bondár, I. and McLaughlin, K. L. A new ground truth data set for seismic studies. *Seismological Research Letters*, 80(3):465–472, 2009. doi: 10.1785/gssrl.80.3.465.

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

Braszus, B., Rietbrock, A., Haberland, C., and Ryberg, T. AI based 1-D P- and S-wave velocity models for the greater alpine region from local earthquake data. *Geophysical Journal International*, 237(2):916–930, 2024. doi: 10.1093/gji/ggae077.

CERN. CERN Seismic Network, 2016. doi: 10.12686/sed/networks/c4.

Charles University in Prague (Czech), Institute of Geonics, Institute of Geophysics, Academy of Sciences of the Czech Republic, Institute of Physics of the Earth Masaryk University (Czech) & Institute of Rock Structure and Mechanics. Czech Regional Seismic Network [Data set], 1973. doi: 10.7914/SN/CZ.

Dahm, T., Ohrnberger, M., Vollmer, D., Woith, H., and Isken, M. Longterm monitoring of swarm earthquakes in the western Eger rift [Data set], 2024. doi: 10.14470/6Q705117.

Department of Earth and Environmental Sciences, Geophysical Observatory, University of Munchen. BayernNetz [Data set], 2001. doi: 10.7914/SN/BW.

Department of Geosciences, Bensberg Observatory, University of Cologne. Bensberg Earthquake Network [Data set], 2016. doi: 10.7914/SN/BQ.

Diehl, T., Kissling, E., Herwegh, M., and Schmid, S. M. Improving absolute hypocenter accuracy with 3D P_g and S_g body-wave inversion procedures and application to earthquakes in the Central Alps region. *Journal of Geophysical Research: Solid Earth*, 126(1):e2021JB022155, 2021. doi: 10.1029/2021JB022155.

Epos-France. Epos-France Broad-band network (RLBP) [Data set], 1962. doi: 10.15778/RESIF.FR.

Epos-France Seismology. French Associated Seismological Network [Data set], 2020. doi: 10.15778/resif.fo.

Erdbebendienst Südwest Baden-Württemberg and Rheinland-Pfalz. Erdbebendienst Südwest [Data set], 2009. doi: 10.7914/SN/LE.

Federal Institute for Geosciences and Natural Resources. German Regional Seismic Network (GRSN), 1976. doi: 10.25928/mbx6-hr74.

Finger, C., Harrington, R. M., and Reinsch, T. Seismological Characterization of the Geothermal Field Laboratory Rhineland [Data set], 2022. doi: 10.14470/MO7576467356.

Friederich, W., Fischer, K. D., Rische, M., Hilgers, C., and Ruhr University Bochum. FloodRisk Seismic Network [Data set], 2020. doi: 10.7914/SN/YD_2020.

Gallacher, R., Garth, T., Harris, J., Bondár, I., McLaughlin, K., and Storchak, D. A. Revising the Seismic Ground Truth Reference Event Identification Criteria. *Seismica*, 4(1):1–13, 2025. doi: 10.26443/seismica.v4i1.1536.

GEOFON Data Centre. GEOFON Seismic Network [Data set], 1993. doi: 10.14470/TR560404.

Geological Survey-Provincia Autonoma di Trento. Trentino Seismic Network [Data set], 1981. doi: 10.7914/SN/ST.

Grad, M. and Tiira, T. The Moho depth map of the European Plate. *Geophysical Journal International*, 176(1):279–292, 2009. doi: 10.1111/j.1365-246X.2008.03919.x.

Hara, S., Fukahata, Y., and Iio, Y. P wave first-motion polarity determination of waveform data in western Japan using deep learning. *Earth Planets and Space*, 71(1):1–11, 2019. doi: 10.1186/s40623-019-1111-x.

Harrington, R. M., Roth, M. P., Fischer, K. D., and Finger, C. RUB Observational Geophysics Lab Deployment: Lower Rhine Embayment for project SIEGFRIED [Data set], 2024. doi: 10.7914/brb1-cf34.

Havskov, J., Bormann, P., and Schweitzer, J. Seismic source location. In Bormann, P., editor, *New Manual of Seismological Observatory Practice 2 (NMSOP-2)*, pages 1–36. Deutsches GeoForschungsZentrum GFZ, 2012. doi: 10.2312/GFZ.NMSOP-2_IS_11.1.

Hensch, M., Dahm, T., Ritter, J., Heimann, S., Schmidt, B., Stange, S., and Lehmann, K. Deep low-frequency earthquakes reveal ongoing magmatic recharge beneath Laacher See Volcano (Eifel, Germany). *Geophysical Journal International*, 216(3): 2025–2036, 2019. doi: 10.1093/gji/ggy532.

Hessian Agency for Nature Conservation, Environment and Geology. Hessischer Erdbebendienst [Data set]., 2012. doi: 10.7914/SN/HS.

Hetényi, G., Molinari, I., Clinton, J., Bokelmann, G., Bondár, I., Crawford, W. C., Dessa, J. X., Doubre, C., Friederich, W., Fuchs, F., Giardini, D., Gráczer, Z., Handy, M. R., Herak, M., Jia, Y., Kissling, E., Kopp, H., Korn, M., Margheriti, L., Meier, T., Mucciarelli, M., Paul, A., Pesaresi, D., Piromallo, C., Plenefisch, T., Plomerová, J., Ritter, J., Rümpker, G., Šipka, V., Spallarossa, D., Thomas, C., Tilmann, F., Wassermann, J., Weber, M., Wéber, Z., Wesztergom, V., Živčić', M., AlpArray Seismic Network Team, AlpArray OBS Cruise Crew, and AlpArray Working Group. The AlpArray Seismic Network: A Large-Scale European Experiment to Image the Alpine Orogen. *Surveys in Geophysics*, 39:1009–1033, 2018. doi: 10.1007/s10712-018-9472-4.

Hinzen, K. G., Reamer, S. K., and Fleischer, C. Seismicity in the Northern Rhine Area (1995-2018). *Journal of Seismology*, 25: 351–367, 2021. doi: 10.1007/s10950-020-09976-7.

Hunter, J. D. Matplotlib: A 2D graphics environment. *IEEE Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

Husen, S. and Hardebeck, J. L. Earthquake location accuracy. *Community Online Resource for Statistical Seismicity Analysis*, 2010. doi: 10.5078/corssa-55815573.

Institut de physique du globe de Paris (IPGP), & École et Observatoire des Sciences de la Terre de Strasbourg (EOST). GEOSCOPE, French Global Network of broad band seismic stations, 1982. doi: 10.18715/GEOSCOPE.G.

Institut für Geowissenschaften, Friedrich-Schiller-Universität Jena. Thüringer Seismologisches Netz [Data set], 2009. doi: 10.7914/SN/TH.

Institute of Geophysics, Academy of Sciences of the Czech Republic. West Bohemia Local Seismic Network [Data set], 1991. doi: 10.7914/SN/WB.

Istituto Nazionale di Geofisica e Vulcanologia (INGV). Rete Sismica Nazionale (RSN) [Data set], 2005. doi: 10.13127/sd/x0fxnh7qfy.

Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS. North-East Italy Seismic Network [Data set], 2016. doi: 10.7914/SN/OX.

Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, & University of Trieste. North-East Italy Broadband Network [Data set], 2002. doi: 10.7914/SN/NI.

KNMI. Netherlands Seismic and Acoustic Network, 1993. doi: 10.21944/e970fd34-23b9-3411-b366-e4f72877d2c5.

Kövesligethy Radó Seismological Observatory (Geodetic And Geophysical Institute, Research Centre For Astronomy And Earth Sciences, Hungarian Academy Of Sciences (MTA CSFK GGI KRSZO)). Hungarian National Seismological Network [Data set], 1992. doi: 10.14470/UH028726.

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C., and Wassermann, J. ObsPy: a bridge for seismology into the scientific Python ecosystem. *Computational Science & Discovery*, 8(1):014003, 2015. doi: 10.1088/1749-4699/8/1/014003.

Küperkoch, L., Olbert, K., and Meier, T. Long-term monitoring of induced seismicity at the Insheim geothermal site, Germany. *Bulletin of the Seismological Society of America*, 108(6):3668–3683, 2018. doi: 10.1785/0120170365.

Lengliné, O., Schmittbuhl, J., Drif, K., Lambotte, S., Grunberg, M., Kinscher, J., Sira, C., Schlupp, A., Schaming, M., Jund, H., and Masson, F. The largest induced earthquakes during the GEOVEN deep geothermal project, Strasbourg, 2018-2022: from source parameters to intensity maps. *Geophysical Journal International*, 234(3):2445–2457, 2023. doi: 10.1093/gji/ggad255.

Li, W., Chakraborty, M., Köhler, J., Quinteros-Cartaya, C., Rümpker, G., and Srivastava, N. Earthquake monitoring using deep learning with a case study of the Kahramanmaras Turkey earthquake aftershock sequence. *Solid Earth*, 15(2):197–213, 2024. doi: 10.5194/se-15-197-2024.

Lomax, A., Virieux, J., Volant, P., and Berge-Thierry, C. Probabilistic earthquake location in 3D and layered models. In Thurber, C. H. and Rabinowitz, N., editors, *Advances in Seismic Event Location. Modern Approaches in Geophysics*, volume 18, pages 101–134. Springer, 2000. doi: 10.1007/978-94-015-9536-0_5.

Lomax, A., Michelini, A., and Curtis, A. Earthquake Location, Direct, Global-Search Methods. In Meyers, R., editor, *Encyclopedia of Complexity and Systems Science*. Springer, 2014. doi: 10.1007/978-3-642-27737-5_150-2.

Mader, S. M., Ritter, J. R. R., Reicherter, K., and the AlpArray Working Group. Seismicity and seismotectonics of the Albstadt Shear Zone in the northern Alpine foreland. *Solid Earth*, 12(6): 1389–1409, 2021. doi: 10.5194/se-12-1389-2021.

Málek, J., Horálek, J., and Janský, J. One-dimensional qP-wave velocity model of the upper crust for the West Bohemia/Vogtland earthquake swarm region. *Studia Geophysica et Geodaet-*

ica, 49(4):501–524, 2005. doi: 10.1007/s11200-005-0024-2.

Málek, J., Brokešová, J., and Novotný, O. New velocity structure of the Nový Kostel earthquake-swarm region, West Bohemia, determined by the isometric inversion. *Pure and Applied Geophysics*, 180:2111–2134, 2023. doi: 10.1007/s00024-023-03250-w.

MedNet Project Partner Institutions. Mediterranean Very Broadband Seismographic Network (MedNet) [Data set], 1990. doi: 10.13127/sd/fbbbtdtd6q.

Megies, T., Beyreuther, M., Barsch, R., Krischer, L., and Wassermann, J. ObsPy—What can it do for data centers and observatories? *Annals of Geophysics*, 54(1):47–58, 2011. doi: 10.4401/ag-4838.

Michailos, K., Hetényi, G., Scarponi, M., Stipčević, J., Bianchi, I., Bonatto, L., Czuba, W., Di Bona, M., Govoni, A., Hannemann, K., Janik, T., Kalmár, D., Kind, R., Link, F., Lucente, F. P., Monna, S., Montuori, C., Mroczek, S., Paul, A., Piromallo, C., Plomerová, J., Rewers, J., Salimbeni, S., Tilmann, F., Środa, P., Vergne, J., and the AlpArray-PACASE Working Group. Moho depths beneath the European Alps: a homogeneously processed map and receiver functions database. *Earth System Science Data*, 15:2117–2138, 2023. doi: 10.5194/essd-15-2117-2023.

Mousavi, S. M. and Beroza, G. C. A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, 47(1):1–7, 2020. doi: 10.1029/2019GL085976.

Mousavi, S. M. and Beroza, G. C. Deep-learning seismology. *Science*, 377:6607, 2022. doi: 10.1126/science.abm4470.

Mousavi, S. M. and Beroza, G. C. Machine Learning in Earthquake Seismology. *Annual Review of Earth and Planetary Sciences*, 51: 105–129, 2023. doi: 10.1146/annurev-earth-071822-100323.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L., and Beroza, G. C. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11:3952, 2020. doi: 10.1038/s41467-020-17591-w.

Münchmeyer, J. PyOcto: A high-throughput seismic phase associator. *Seismica*, 3(1):1–15, 2024. doi: 10.26443/seismica.v3i1.1130.

Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., Michelini, A., Saul, J., and Soto, H. Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. *Journal of Geophysical Research: Solid Earth*, 127(1): e2021JB023499, 2022. doi: 10.1029/2021JB023499.

Myers, S. C., Begnaud, M. L., Ballard, S., Pasyanos, M. E., Phillips, W. E., Ramírez, 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 P_n Travel-Time Calculation. *Bulletin of the Seismological Society of America*, 100(2): 640–656, 2010. doi: 10.1785/0120090198.

Park, Y., Beroza, G. C., and Ellsworth, W. L. A mitigation strategy for the prediction inconsistency of neural phase pickers. *Seismological Research Letters*, 94(3):1603–1612, 2023. doi: 10.1785/0220230003.

Quinteros, J., Carter, J. A., Schaeffer, J., Trabant, C., and Pedersen, H. A. Exploring approaches for large data in seismology: user and data repository perspectives. *Seismological Research Letters*, 92(3):1531–1540, 2021. doi: 10.1785/0220200390.

Quinteros-Cartaya, C., Köhler, J., Li, W., Faber, J., and Srivastava, N. Exploring a CNN model for earthquake magnitude estimation using HR-GNSS data. *Journal of South American Earth Sciences*, 136:104815, 2024. doi: 10.1016/j.jsames.2024.104815.

Raspberry Shake, S.A. Raspberry Shake [Data set], 2016. doi: 10.7914/SN/AM.

Reamer, S. K. and Hinzen, K. G. An earthquake catalog for the Northern Rhine area, Central Europe (1975-2002). *Seismological Research Letters*, 75(6):713–725, 2004. doi: 10.1785/gssrl.75.6.713.

Renouard, A., Maggi, A., Grunberg, M., Doubre, C., and Hibert, C. Toward false event detection and quarry blast versus earthquake discrimination in an operational setting using semiautomated machine learning. *Seismological Research Letters*, 92(6): 3725–3742, 2021. doi: 10.1785/0220200305.

RESIF. CEA/DASE broad-band permanent network in metropolitan France [Data set], 2018. doi: 10.15778/RESIF.RD.

Retailleau, L., Saurel, J.-M., Zhu, W., Satriano, C., Beroza, G. C., Issartel, S., Boissier, P., OVPF Team, and OVSM Team. A wrapper to use a machine-learning-based algorithm for earthquake monitoring. *Seismological Research Letters*, 93(3):1673–1682, 2022. doi: 10.1785/0220210279.

Ritter, J. and Gassner, L. Inter-Wind [Data set], 2022. doi: 10.14470/6Q705117.

Ritter, J. R. R., Koushesh, K., Schmidt, B., Föst, J. P., Bühler, J., Hensch, M., and Mader, S. M. Seismological monitoring of magmatic and tectonic earthquakes in the East Eifel Volcanic Field, Germany. *Journal of Seismology*, 28:1325–1350, 2024. doi: 10.1007/s10950-024-10257-w.

Ross, Z. E., Meier, M.-A., Hauksson, E., and Heaton, T. H. Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, 108(5A):2894–2901, 2018a. doi: 10.1785/0120180080.

Ross, Z. E., Meier, M. A., and Hauksson, E. P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*, 123(6):5120–5129, 2018b. doi: 10.1029/2017JB015251.

Royal Observatory of Belgium. Belgian Seismic Network [Data set], 1985. doi: 10.7914/SN/BE.

Růžek, B. and Horálek, J. Three-dimensional seismic velocity model of the West Bohemia/Vogtland seismoactive region. *Geophysical Journal International*, 195(2):1251–1266, 2013. doi: 10.1093/gji/ggt295.

Ruhr University Bochum. RuhrNet - Seismic Network of the Ruhr-University Bochum [Data set], 2007. doi: 10.7914/SN/RN.

Schlittenhardt, J. Regional velocity models for Germany: a contribution to the systematic travel-time calibration of the International Monitoring System. In *Proceedings of the 21st DoD/-DOE Seismic Research Symposium: Technologies for Monitoring the comprehensive Nuclear-Test-Ban Treaty*, volume 1, pages 263–273, 1999.

Schober, P., Boer, C., and Schwarte, L. A. Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, 126 (5):1763–1768, 2018. doi: 10.1213/ANE.0000000000002864.

Shi, P., Grigoli, F., Lanza, F., Beroza, G. C., Scarabello, L., and Wiemer, S. MALMI: An automated earthquake detection and location workflow based on machine learning and waveform migration. *Seismological Research Letters*, 93(5):2467–2483, 2022. doi: 10.1785/0220220071.

Slovenian Environment Agency. Seismic Network of the Republic of Slovenia [Data set], 1990. doi: 10.7914/SN/SL.

Soto, H. and Schurr, B. DeepPhasePick: a method for detecting and picking seismic phases from local earthquakes based on highly optimized convolutional and recurrent deep neural networks. *Geophysical Journal International*, 227(2):1268–1294, 2021. doi: 10.1093/gji/ggab266.

Stammler, K. SeismicHandler—Programmable multichannel data handler for interactive and automatic processing of seismolog-

ical analyses. *Geophysical Journal International*, 19(2):135–140, 1993. doi: 10.1016/0098-3004(93)90110-Q.

Stammler, K., Bischoff, M., Brüstle, A., Ceranna, L., Donner, S., Fischer, K., Gaebler, P., Friederich, W., Funke, S., Hartmann, G., Homuth, B., Knapmeyer-Endrun, B., Korn, M., Megies, T., Pilger, C., Plenefisch, T., Pustal, I., Rappsilber, I., Schmidt, B., Sonnabend, L., Stange, S., Wassermann, J., and Wegler, U. German seismic and infrasound networks contributing to the European Integrated Data Archive (EIDA). *Seismological Research Letters*, 92 (3):1854–1875, 2021. doi: 10.1785/0220200401.

Steinberg, A., Vasyura-Bathke, H., Gaebler, P., Ohrnberger, M., and Ceranna, L. Estimation of seismic moment tensors using variational inference machine learning. *Journal of Geophysical Research: Solid Earth*, 126:e2021JB022685, 2021. doi: 10.1029/2021JB022685.

Strollo, A., Cambaz, D., Clinton, J., Danecek, P., Evangelidis, C. P., Marmureanu, A., Ottemøller, L., Pedersen, H., Sleeman, R., Stammler, K., Armbruster, D., Bienkowski, J., Boukouras, K., Evans, P. L., Fares, M., Neagoe, C., Heimers, S., Heinloo, A., Hoffmann, M., Kaestli, P., Lauciani, V., Michalek, J., Odon Muhire, E., Ozer, M., Palangeanu, L., Pardo, C., Quinteros, J., Quintiliani, M., Jara-Salvador, J. A., Schaeffer, J., Schloemer, A., and Triantafyllis, N. EIDA–The European Integrated Data Archive and service infrastructure within ORFEUS. *Seismological Research Letters*, 92(3):1788–1795, 2021. doi: 10.1785/0220200413.

Swiss Seismological Service (SED) At ETH Zurich. National Seismic Networks of Switzerland, 1983. doi: 10.12686/sed/networks/ch.

Swiss Seismological Service (SED) At ETH Zurich. Temporary deployments in Switzerland associated with aftershocks and other seismic sequences, 2005. doi: 10.12686/sed/networks/8d.

Swiss Seismological Service (SED) At ETH Zurich. GEOBEST Baseline Seismic Monitoring Networks for Deep Geothermal Energy Projects in Switzerland, 2006. doi: 10.12686/SED/NETWORKS/G2.

Swiss Seismological Service (SED) At ETH Zurich. Seismology at School Program, ETH Zurich, 2008. doi: 10.12686/SED/NETWORKS/S.

Swiss Seismological Service (SED) At ETH Zurich. Temporary deployments in Switzerland associated with landslides, 2012. doi: 10.12686/SED/NETWORKS/XP.

Swiss Seismological Service (SED) At ETH Zurich. Temporary deployments in Switzerland associated with building monitoring, 2014. doi: 10.12686/sed/networks/9e.

Theunissen, T., Chevrot, S., Sylvander, M., Monteiller, V., Calvet, M., Villaseñor, A., Benahmed, S., Pauchet, H., and Grimaud, F. Absolute earthquake locations using 3-D versus 1-D velocity models below a local seismic network: example from the Pyrenees. *Geophysical Journal International*, 212(3):1806–1828, 2018. doi: 10.1093/gji/ggx472.

Tian, D., Uieda, L., Leong, W. J., Fröhlich, Y., Schlitzer, W., Grund, M., Jones, M., Toney, L., Yao, J., Magen, Y., Jing-Hui, T., Materna, K., Belem, A., Newton, T., Anant, A., Ziebarth, M., Quinn, J., and Wessel, P. PyGMT: A Python interface for the Generic Mapping Tools, 2024. doi: 10.5281/zenodo.13679420.

Tiira, T., Uski, M., Kortström, J., Kaisko, O., and Korja, A. Local seismic network for monitoring of a potential nuclear power plant area. *Journal of Seismology*, 20:397–417, 2016. doi: 10.1007/s10950-015-9534-8.

University of Leipzig. SXNET Saxon Seismic Network [Data set], 2001. doi: 10.7914/SN/SX.

University of Trieste. Friuli Venezia Giulia Accelerometric Network [Data set], 1993. doi: 10.7914/SN/RF.

Uta, P. *Recent intraplate earthquakes in Northwest Germany: glacial isostatic adjustment and/or a consequence of hydrocarbon production?* PhD thesis, Gottfried Wilhelm Leibniz Universität Hannover, 2017. https://www.repo.uni-hannover.de/bitstream/123456789/9141/1/895461218.pdf.

Wang, T., Bian, Y., Zhang, Y., and Hou, X. Classification of earthquakes, explosions and mining-induced earthquakes based on XGBoost algorithm. *Computers & Geosciences*, 170:105242, 2023. doi: 10.1016/j.cageo.2022.105242.

Waskom, M. L. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021. doi: 10.21105/joss.03021.

Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., and Tian, D. The Generic Mapping Tools Version 6. *Geochemistry, Geophysics, Geosystems*, 20(11):5556–5564, 2019. doi: 10.1029/2019GC008515.

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., Michelini, A., Saul, J., and Soto, H. SeisBench—A toolbox for machine learning in seismology. *Seismological Research Letters*, 93(3):1695–1709, 2022. doi: 10.1785/0220210324.

Yin, J., Denolle, M. A., and He, B. A multitask encoder-decoder to separate earthquake and ambient noise signal in seismograms. *Geophysical Journal International*, 231(3):1806–1822, 2022. doi: 10.1093/gji/ggac290.

Yuan, C., Ni, Y., Lin, Y., and Denolle, M. A. Better together: ensemble learning for earthquake detection and phase picking. *IEEE Transactions on Geoscience and Remote Sensing*, 61:5920217, 2023. doi: 10.1109/TGRS.2023.3320148.

ZAMG - Zentralanstalt für Meterologie und Geodynamik. Austrian Seismic Network [Data set], 1987. doi: 10.7914/SN/OE.

Zhang, M., Liu, M., Feng, T., Wang, R., and Zhu, W. LOC-FLOW: An end-to-end machine learning-based high-precision earthquake location workflow. *Seismological Research Letters*, 93(5): 2426–2438, 2022. doi: 10.1785/0220220019.

Zhu, W. and Beroza, G. C. PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1):261–273, 2019. doi: 10.1093/gji/ggy423.

Zhu, W., Mousavi, S. M., and Beroza, G. C. Seismic signal denoising and decomposition using deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11):9476–9488, 2019. doi: 10.1109/TGRS.2019.2926772.

Zhu, W., Mousavi, S. M., and Beroza, G. C. Seismic signal augmentation to improve generalization of deep neural networks. In Moseley, M. and Krischer, L., editors, *Machine Learning in Geosciences*, volume 61 of *Advances in Geophysics*, pages 151–177. Elsevier, 2020. doi: 10.1016/bs.agph.2020.07.003.

Zhu, W., McBrearty, I. W., Mousavi, S. M., Ellsworth, W. L., and Beroza, G. C. Earthquake phase association using a Bayesian Gaussian Mixture Model. *Journal of Geophysical Research: Solid Earth*, 127:e2021JB023249, 2022. doi: 10.1029/2021JB023249.