ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR-
UND MEERESFORSCHUNG

Pilot, Matthias; Schlindwein, Vera (2024)

A Practical Approach to Automatic Earthquake Catalog Compilation in Local OBS Networks using Deep-Learning and Network-Based Algorithms. Seismological Research Letters, 95 (4), 2124-2140.

https://doi.org/10.1785/0220230182

A Practical Approach to Automatic Earthquake Catalog Compilation in Local OBS Networks using Deep-Learning and Network-Based Algorithms

Matthias Pilot*[,1,2], Vera Schlindwein[1,2]

[1] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Am Alten Hafen 26, 27568 Bremerhaven, Germany

[2] Also at: University of Bremen, Faculty of Geosciences, Klagenfurter Straße 2-4, 28359 Bremen, Germany

* Corresponding author. Adress: Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Am Alten Hafen 26, 27568 Bremerhaven, Germany; E-mail: matthias.pilot@awi.de; ORCID iD: https://orcid.org/0000-0003-1704-5443

Declaration of Competing Interests

The authors declare that they have no competing interests.

Abstract

In land-based seismology modern automatic earthquake detection and phase picking algorithms have already proven to outperform classic approaches, resulting in more complete catalogs while only taking a fraction of the time needed for classic methods. For marine-based seismology similar advances have not been made yet. For Ocean Bottom Seismometer (OBS) data additional challenges arise, such as a lower signal-to-noise ratio and fewer labelled datasets available for training deep-learning models. However, the performance of available deep-learning models has not yet been extensively tested on marine-based datasets. Here, we apply three different modern event detection and phase picking approaches to a ~12-months local OBS dataset and compare the resulting earthquake catalogs and location results. Additionally, we evaluate their performance by comparing different sub-catalogs of manually detected events and visually revised picks to their automatic counterparts. The results show that seismicity patterns from automatically compiled catalogs are comparable to a manually revised catalog after applying strict location quality control criteria. However, the number of such well-constrained events varies between the approaches and catalog completeness can not be reliably determined. We find that PhaseNet is more suitable for local OBS networks compared to EQTransformer and propose a pick-independent event detection approach, such as Lassie, as the preferred choice for an initial event catalog compilation. Depending on the aim of the study different schemes of manual re-picking should be applied, as the automatic picks are not yet reliable enough for developing a velocity model or interpreting small-scale seismicity patterns.

Introduction

In passive seismology, data typically consist of continuous recording of ground motion by seismometers in three spatial directions. Catalogs of reliably located earthquakes are compiled from these time series which are subsequently used for, e.g., geological interpretation, hazard assessment or earthquake tomography (Douilly *et al.*, 2016; Parnell-Turner *et al.*, 2020; Meier *et al.*, 2022; Yaroshenko *et al.*, 2022). Within the processing workflow a number of steps, including considerable manual work, have to be accomplished. Events first need to be identified by their characteristic waveforms and the onset times of P and S phases need to be accurately determined. Classic methods such as the short-term average to long-term average ratio (STA/LTA) approach (Allen, 1978), the use of kurtosis-based characteristic functions (Baillard *et al.*, 2014) and template matching approaches (Gibbons and Ringdal, 2006) have reduced the amount of time needed by seismologists to detect earthquakes and pick phases. To reduce the influence of misidentified or missing phase picks inherent in these automatic approaches, strict quality control criteria are applied after location, restricting the events in a catalog to, e.g., events that have phase picks at a minimum number of stations while producing a low root-mean square (RMS) residual (e.g., Parnell-Turner *et al.*, 2020). In recent years, automatic event detectors and phase pickers, including deep-learning approaches, have received increasing interest (Mousavi and Beroza, 2022). Due to this fast-evolving field much of the time-consuming manual work in classic seismological workflows can potentially be saved in future. Additionally, the use of deep-learning methods has resulted in more complete earthquake catalogs (Seydoux *et al.*, 2020; Majstorović *et al.*, 2021; Park *et al.*, 2022; Wu *et al.*, 2022; Scotto di Uccio *et al.*, 2023). With the ever-increasing amount of data available, e.g., due to the recent use of fiber optic cables for earthquake detection (Lindsey and Martin, 2021; Spica *et al.*, 2022), automated methods will be essential for the future. However, due to the black-box nature of deep-learning approaches, a better understanding of the limitations and effects on the results is necessary (Mousavi and Beroza, 2022; Park *et al.*, 2023), especially when applying these methods to datasets with characteristics that differ from the underlying training datasets.

Additional challenges arise with Ocean Bottom Seismometer (OBS) recordings. So far, deep-learning models have been trained on land-based earthquake catalogs (Ross *et al.*, 2018; Zhu and Beroza, 2019; Mousavi *et al.*, 2020). Past attempts to train the models on OBS data did not show significant improvements yet (Chen *et al.*, 2022) or are currently underway (Bornstein *et al.*, 2023, prepr.). While the land-based catalogs are large, the performance of the trained models strongly depends on the datasets they are applied on (Münchmeyer *et al.*, 2022). Compared to land-based datasets, OBS datasets typically show a lower signal-to-noise ratio. In addition, they are subject to region-specific noise such as induced tremor by ocean-bottom currents (Hilmo and Wilcock, 2020; Essing *et al.*, 2021), abundance of short-duration events (SDE) (Tary *et al.*, 2012; Domel *et al.*, 2022), OBS self-noise (Stähler *et al.*, 2018), marine mammal vocalisations (Brodie and Dunn, 2015), and anthropogenic noise related to seismic surveying and ship noise (Trabattoni *et al.*, 2023). The resulting plethora of seismic signals have a similar frequency range and duration as earthquake signals and cause abundant false detections of, e.g., STA/LTA detectors (Williams *et al.*, 2010) and even modern deep-learning models have difficulties in correctly identifying earthquakes in marine data sets (Domel *et al.*, 2023). Instead, OBS surveys often either rely on classic methods (Chen *et al.*, 2023), resample the input data to better fit the training datasets (Gong *et al.*, 2022), or return to manual phase picking (Meier *et al.*, 2021). While Wu *et al.* (2022) developed a workflow including deep-learning methods for OBS data, the effects of different automatic approaches on the resulting earthquake catalog remains unclear.

Here, we evaluate to what extent modern event detection and phase picking approaches can be used to automatically compile a consistent earthquake catalog from local OBS networks. For this purpose, we use data from a ~12-months OBS deployment in the Norwegian-Greenland Sea in a seismically active area of the Knipovich Ridge (Figure 1, a). The network consists of eight OBS with an instrument spacing of 5-8 km (Figure 1, b). We compare different sub-catalogs of manually detected and picked events to their automatic counterparts to evaluate their performance and the effect of manual re-picking on the resulting earthquake catalog. Additionally, we test three different automatic approaches, show their limitations, and evaluate the resulting earthquake catalogs and location results after applying quality control criteria. Thus, this study provides a practical approach on how to employ deep-learning and network-based earthquake detection and phase picking algorithms for similar marine seismological datasets, showing their specific limitations and highlighting where manual re-evaluation is still necessary.

Data and Methods

For this study we used the Loki dataset, consisting of eight four-channel OBS which were deployed around Loki's Castle hydrothermal vent field and an active fault zone (Johansen *et al.*, 2019) at the Mohn-Knipovich Ridge bend (Figure 1) between July 2019 and July 2020. All OBS were equipped with Trillium Compact broadband seismometers, HighTech Inc hydrophones, and K.U.M. 6D6 data loggers (Schmidt-Aursch and Haberland, 2017) and sampled at 100 Hz for all stations except LOK01 and LOK06 which sampled at 250 Hz. A geological interpretation of the data is subject of a different manuscript, here we will solely focus on the automatic approaches to yield an earthquake catalog.

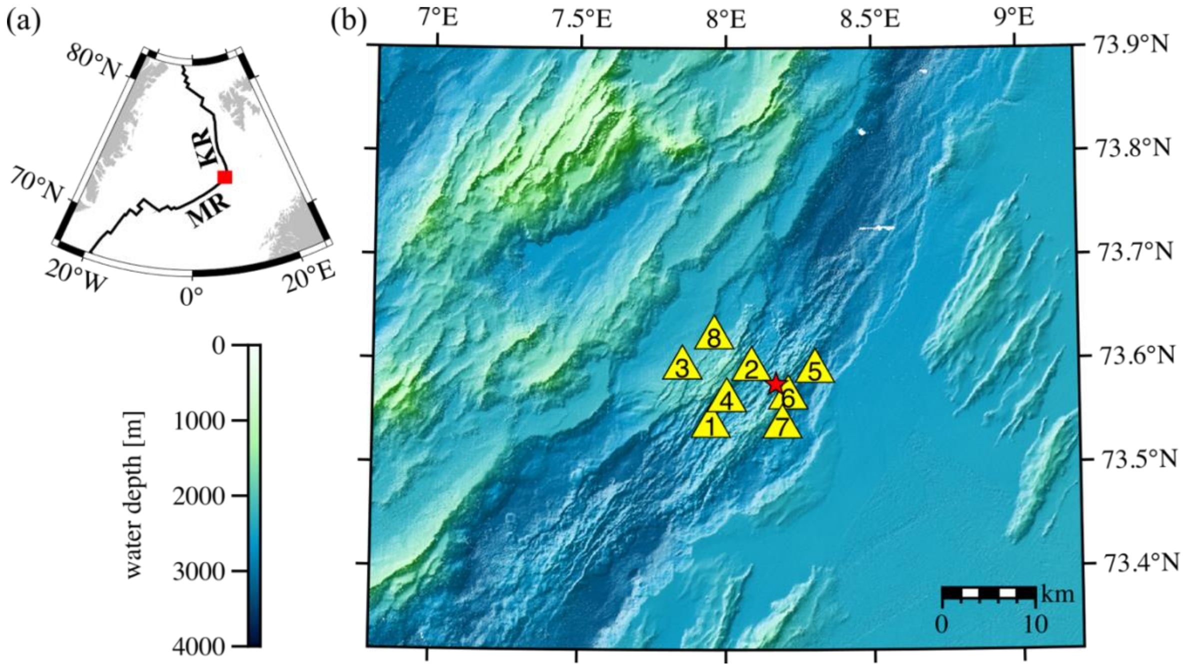

Figure 1: a): Overview map showing the study area (red square) in the Norwegian-Greenland Sea. MR = Mohns Ridge, KR = Knipovich Ridge. b): Bathymetry of the Mohn-Knipovich Ridge bend including the positions and numbers of the OBS stations (yellow triangles) and the position of Loki's Castle (red star). Bathymetry data from Kartverket (www.kartverket.no).

Event Detection

From this ~12-months continuous recording we selected 11 days, evenly distributed throughout the deployment duration, to manually create a reference event database (Figure 2, a, left; Figure S1 in Supplement). The reference days were manually screened for earthquakes by looking at 3-15 Hz bandpass filtered seismic traces of all components and stations using SEISAN (Havskov and Ottemöller, 1999; Havskov *et al.*, 2020). If an earthquake signal was observed at three or more stations an event was registered into the reference database. This served as a "ground-truth" database of locatable earthquakes for later comparison. Typically the event waveform was manually cut ~15 s before and ~30 s after the first arrival unless another event was within this time window, then it was cut shorter. For all days we marked the most prominent events that were visible at seven or eight stations. In addition, for a single day (Figure S2 in Supplement), we counted the number of stations with discernible signal amplitude for all manually detected events.

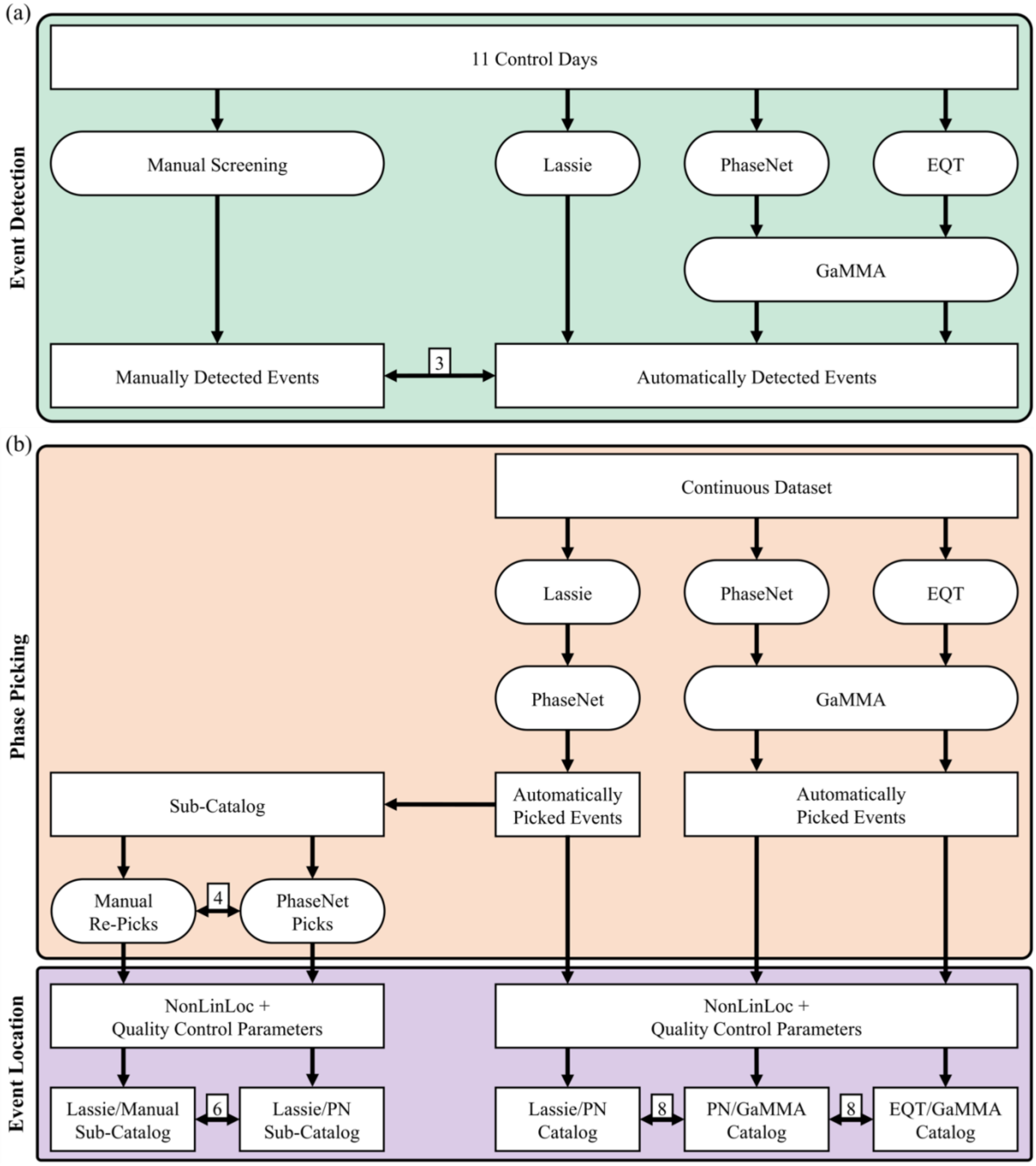

Figure 2: Schematic workflows used for this study. a): For event detection we used a subset of 11 control days and compared a manually compiled event catalog to the event detections of three automatic detectors (Lassie, PhaseNet/GaMMA, and EQTransformer (EQT)/GaMMA). b): From the continuous dataset three automatically detected and picked event catalogs were compiled (top box, right). A sub-catalog of Lassie detected and PhaseNet picked, best-constrained events was manually re-picked for comparison with the original PhaseNet picks

(top box, left). For event location, both sub-catalogs (bottom box, left) and the three automatically compiled catalogs (bottom box, right) were located and location quality-control criteria were applied before comparing the location results. Numbers in squares refer to the corresponding figures.

The same 11 control days were then used for the automatic event detection approaches (Figure 2, a, right) to compile event catalogs and compare them to the manually compiled event catalog. First, we used the migration-based Lassie earthquake detector (Heimann *et al.*, 2017). Lassie computes a characteristic function for each station individually which is then back-shifted by the expected travel time within a grid covering the seismic network. The characteristic function of each station is then stacked to obtain an image function for each possible source location. An event is detected if the detection threshold of the image function is exceeded. For Lassie we used a preliminary velocity model that is based on seismic profiles from the study area (Jeddi *et al.*, 2021). By comparison with the manually detected events, we found that a detection threshold of 36 is high enough to not detect continuous noise as events (Figure S3, d, in Supplement) but low enough to not exclude smaller events.

Additionally, we used the GaMMA associator (Zhu *et al.*, 2022) in combination with the deep-learning-based PhaseNet (Zhu and Beroza, 2019) and EQTransformer (Mousavi *et al.*, 2020) to automatically pick P and S phases on the 11 control days and associate them to seismic events. We chose these models due to their good cross-domain performance (Münchmeyer *et al.*, 2022). For PhaseNet we used the model which was trained on the Northern California Earthquake Catalog and kept the default P and S detection thresholds of 0.3. The data from stations LOK01 and LOK06 was resampled to 100 Hz. For EQTransformer we used the original model and a detection threshold of 0.3, a P threshold of 0.3 and a S threshold of 0.5. The general detection threshold is used to detect earthquake signals within the data while the P and S thresholds are used for phase picking (Mousavi *et al.*, 2020). For the subsequent association of phases to seismic events with the GaMMA associator, a constant P velocity of 6 km/s and Vp/Vs 1.75 was assumed and we required at least three associated P picks for an event detection (further settings in Table S1). Comparable to the Lassie detection value, GaMMA calculates a probability value (GaMMA score) for each associated event.

To evaluate the performance of the automatic event detection procedures we compared the origin times of the automatically detected events to start times of the waveforms in the manually created reference dataset. Events were considered as matched if origin times occurred between 5 s before to 20 s after the start of the waveform, equivalent to origin times 20 s before to 5 s after the roughly determined first arrivals in manual screening.

Phase Picking and Velocity Model

For the subsequent evaluation of the phase picking and location results, we ran the three automatic approaches on the ~12-months continuous dataset (Figure 2, b) with the same settings as described in the previous chapter. We cut event waveforms from the continuous data with a time window of 45 s around the Lassie and GaMMA origin times (- 15 s, + 30 s). We operated PhaseNet to pick phases in all Lassie-detected events. As PhaseNet often picked more than one P or S phase on a single station for the same event, we only kept picks that were within ± 2 s of the theoretical Lassie phase arrival times. For the two GaMMA catalogs phase picking was already done in the previous step (PhaseNet, EQTransformer; Figure 2, b, top box).

The events from the Lassie catalog were located with HYPOSAT (Schweitzer, 2001, 2018) using the preliminary velocity model. From this catalog we selected a subset of best-constrained events which had picks at seven or eight stations, were within the network (gap ≤ 120 °) and had a RMS ≤ 0.2 s. This resulted in a sub-catalog of 1534 events which were then manually re-picked and compared to the sub-catalog with the original PhaseNet phase picks (Figure 2, b, top box, left).

To find an appropriate velocity model and station correction terms we selected strong, well-observed events within the network from the manually re-picked sub-catalog, using the Lassie detection value as a proxy. A value of ≥ 130 yielded 386 events, which were inverted by PyVelest (Kissling *et al.*, 1995) in an iterative approach. From 1900 randomly created velocity models the best fitting model was chosen based on the minimum total RMS. Station correction terms were determined by locating the subset of 386 events with NonLinLoc (Lomax *et al.*, 2000, 2009) iteratively. The mean station corrections of a location run were used as a priori station corrections for the subsequent location run. Since the lowest RMS solution does not necessarily represent a stable, optimal solution to the inverse problem (Schlindwein, 2020), we also considered the average length of the three axes of the error-ellipsoid (abbreviated as error-ellipsoid length from here on), the average hypocenter depth, the average difference between maximum likelihood and expectation hypocenter (abbreviated as hypocentral spread from here on) and the difference between S and P phase station correction terms in the selection of the final station correction terms. These parameters stabilized after three NonLinLoc location iterations and yielded the final station correction terms which were used in combination with the minimum RMS velocity model for all subsequent earthquake locations.

Event Location and Quality Control Criteria

For event location (Figure 2, b, bottom box) we used NonLinLoc with the Oct-Tree sampling algorithm (Lomax and Curtis, 2001) and the least square GAU_ANALYTIC inversion approach (Tarantola and Valette, 1981). For both sub-catalogs (Figure 2, b, bottom box, left) we used a velocity grid with 551 x 551 x 421 (x,y,z) nodes with a spacing of 0.1 km in each direction. We used a search grid with 222 x 222 x 161 (x,y,z) nodes with a spacing of 0.25 km in each direction. As some initial event locations from the three automatically compiled event catalogs (Figure 2, b, bottom box, right) were outside of this grid, the grid size for the these approaches on the continuous dataset was increased to 2551 x 2551 x 421 nodes with a spacing of 0.1 km in each direction (search grid with 1021 x 1021 x 161 (x,y,z) with a spacing of 0.25 km spacing in each direction). In an automated catalog compilation procedure, cut-off thresholds are typically applied to discard poorly located events containing potentially mis-identified or mis-picked phases (e.g., Parnell-Turner *et al.*, 2020). For all approaches we applied the same location quality control criteria for well-constrained events. After inspection of the individual frequency distributions of the sub-catalogs and catalogs, we chose as thresholds for this study a maximum RMS residual of 0.2 s, a maximum average error-ellipsoid length of 1.6 km, and a maximum hypocentral spread of 0.6 km.

Magnitude Calculation

To calculate the magnitudes for all catalogs we used the Automag routine of SEISAN. It automatically picks event amplitudes on Wood-Anderson simulated data from both horizontal components within a 5 s window length around the picked S phase. Amplitudes were only kept if the signal-to-noise ratio was at least 1.5. The local magnitude (ML) was calculated after the equation by Hutton and Boore (1987), using the hypocentral distance instead of the epicentral distance:

$$ML = \log_{10}(amplitude) + 1.11 \cdot \log_{10}(distance) + 0.00189 \cdot distance - 2.09 \quad (1)$$

For the calculation of the magnitude of completeness (Mc) we used the maximum curvature and goodness-of-fit methods (Wiemer and Wyss, 2000).

Results

Automatic Event Detection of Manual Reference Database

Within the 11 screened days a total of 1746 reference events were manually found. Comparing the event detections from the automatic approaches with the manual reference database shows that Lassie performs best at detecting the reference events with clear signal at ≥ 7 stations (≥ 98 %, Table 1). From these events Lassie missed only regional earthquakes which were not targeted. Lassie and EQTransformer in combination with the GaMMA associator also detected a considerable number of events that are not part of the reference events (Figure 3, d; Table 1, see also Figure S4 in Supplement). For Lassie, most of these detections have image function values just above the threshold of 36 (see also Figure S5 in Supplement) and could be removed by raising the threshold value, e.g., to 38, which would however remove some reference events (Figure 3, d). Among the additional events Lassie detected very weak events with discernible arrivals at two stations which were not included in the reference database. The majority of the additional detections were SDEs that reached detection values of up 70 (Figure S3 in Supplement). These signals are visible as high-amplitude, impulsive arrivals on a single station only (Figure S3 in Supplement). Therefore, in the subsequent location procedure SDEs will effectively be removed from the catalog due to the requirement of a minimum of four phase picks. Only in a few cases, Lassie detected events that were missed in the reference database. Most of these earthquakes occurred shortly before, after, or in between two larger events and can easily be missed during manual screening (Figure S6 in Supplement).

Table 1*: Overview of the performance of the automatic event detection approaches for the 11 reference days. Detection rate (det. rate) refers to the percentage of detected events from the 1746 manually detected reference events. Detection rates for the reference events with signal at 7 and 8 stations are also indicated. Additional detections are events not present in the reference events and in parentheses their proportion relative to the total number of automatic detections.*

	Total detections	Overall det. rate [%]	Det. rate for 7 stations [%]	Det. rate for 8 stations [%]	Additional detections
Lassie	2634	57%	98%	99%	1632 (62%)
PhaseNet + GaMMA	707	38%	88%	96%	50 (7%)
EQTransformer + GaMMA	2317	62%	92%	95%	1228 (53%)

While the EQTransformer/GaMMA approach has the highest overall detection rate of 62 %, less reference events with clear signal at ≥ 7 stations were detected compared to Lassie (Figure 3, f; Table 1). The number of false detections (for example see Figure S7 in Supplement) among the additional 1228 events is larger compared to Lassie, where many of the additional detections are SDEs that will be removed during event location. Using the GaMMA score as a proxy to remove most of the potential false detections (e.g., ≥ 8) would result in many reference events being removed (Figure 3, f). PhaseNet in combination with the GaMMA associator detected only 50 events that were not part of the reference database (Figure 3, b; Table 1). Both events from the single test day that were not included in the reference database (Figure 3, e) are true events that were missed manually. However, PhaseNet's overall detection rate of reference events is lowest out of all approaches (38 %, Table 1), with most events visible at < 6 stations not being detected (Figure 3, h).

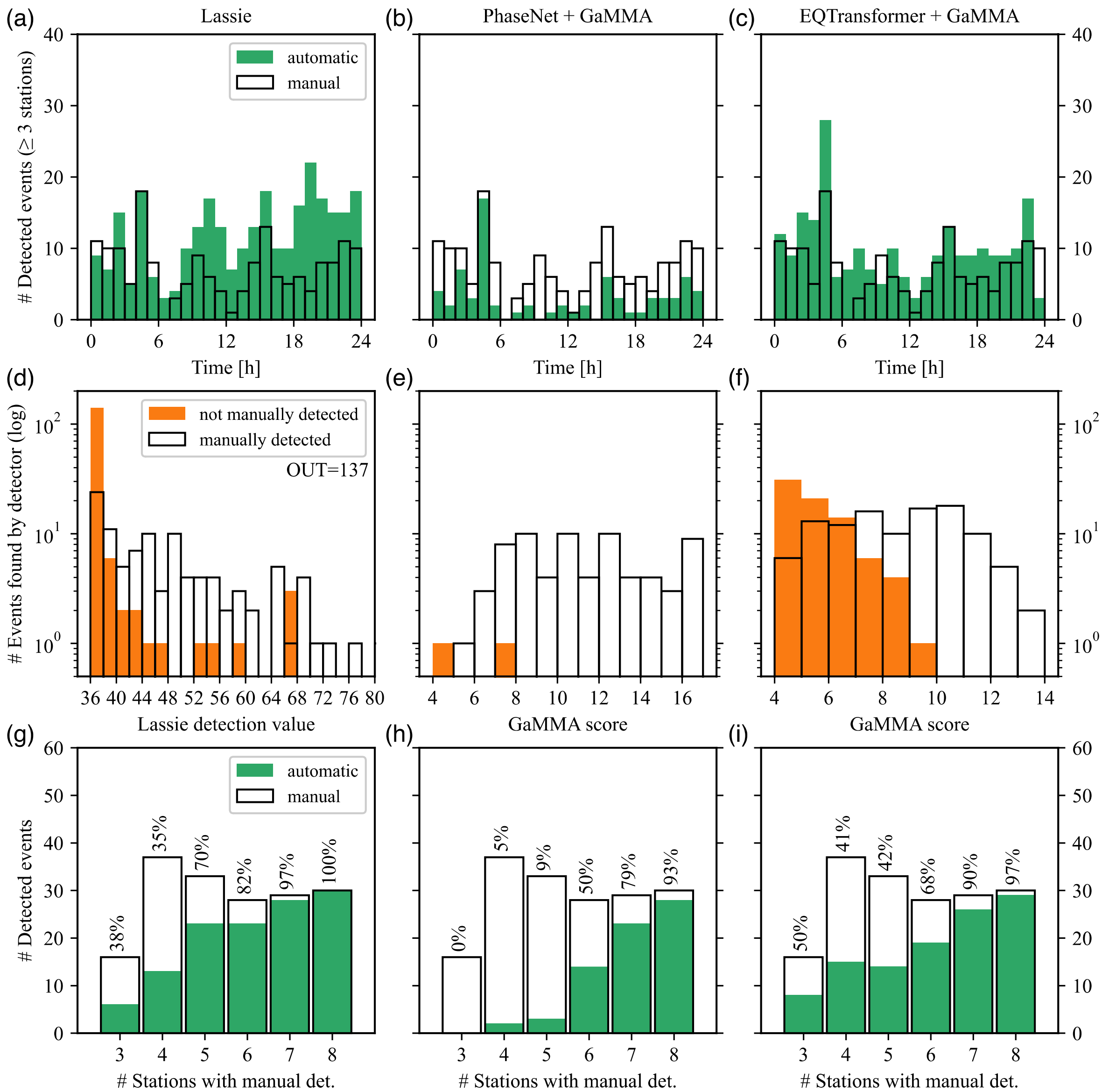

Figure 3: Single day comparison of the three different event detection approaches with the reference event database. (a-c): Number of events automatically (green bars) and manually detected (black line, reference database) at ≥ 3 stations over time. (d-f): Number of automatically detected events that are also in the reference database (black line,) and additional events (orange bars) in relation to the Lassie detection value or GaMMA score. OUT indicates the number of manually detected event above the shown values. (g-i): Number of automatically detected events (green bars) that are part of the reference database (black line) in relation to the number of stations where the event was seen.

Manual Re-Picking of the PhaseNet Picked Lassie Sub-Catalog

During the manual re-picking of the PhaseNet picked Lassie sub-catalog of the 1534 best-constrained events, spanning the entire study period (Figure 2, b, top box, left), we observed systematic mis-picks of the P and S phases at station LOK03. Most prominently, PhaseNet often picked the P phase around 0.2 to 0.6 s too late (Figure 4, a, e). Additionally at this station, 13.6 % of the PhaseNet S picks and 4.2 % of the P picks had to be deleted as incorrect. For the other stations the number of incorrectly picked phases ranged between 0.2 to 2.7 % for the S phases and 0.3 to 3.6 % for the P phases (Figure 4, c-e). Of all final PhaseNet P picks 0.7 to 8.8 % (11 to 101 picks) were manually added during repicking, most notably at stations LOK05 and LOK07 (Figure 4, g, i). Overall, only a few S picks were added manually (0 to 1.8 % or 0 to 13 picks). Most of the manually re-evaluated P and S phase picks are within ± 0.1 s of the original PhaseNet picks and thus only minor or no adjustments of these picks were done assuming that PhaseNet picks the onset of a correctly identified phase arrival in a more systematic manner throughout the dataset than a human analyst.

For the continuous dataset, PhaseNet picked 673,342 P and 929,836 S phases, while EQTransformer picked 1,679,108 P and 370,696 S phases. Both pickers had issues with the P phases at station LOK03, but no other general trends could be observed. While sometimes PhaseNet picked a phase correctly, EQTransformer missed it or vice versa. Also, they sometimes both picked one phase correctly and mis-picked the other (Figure 4, b).

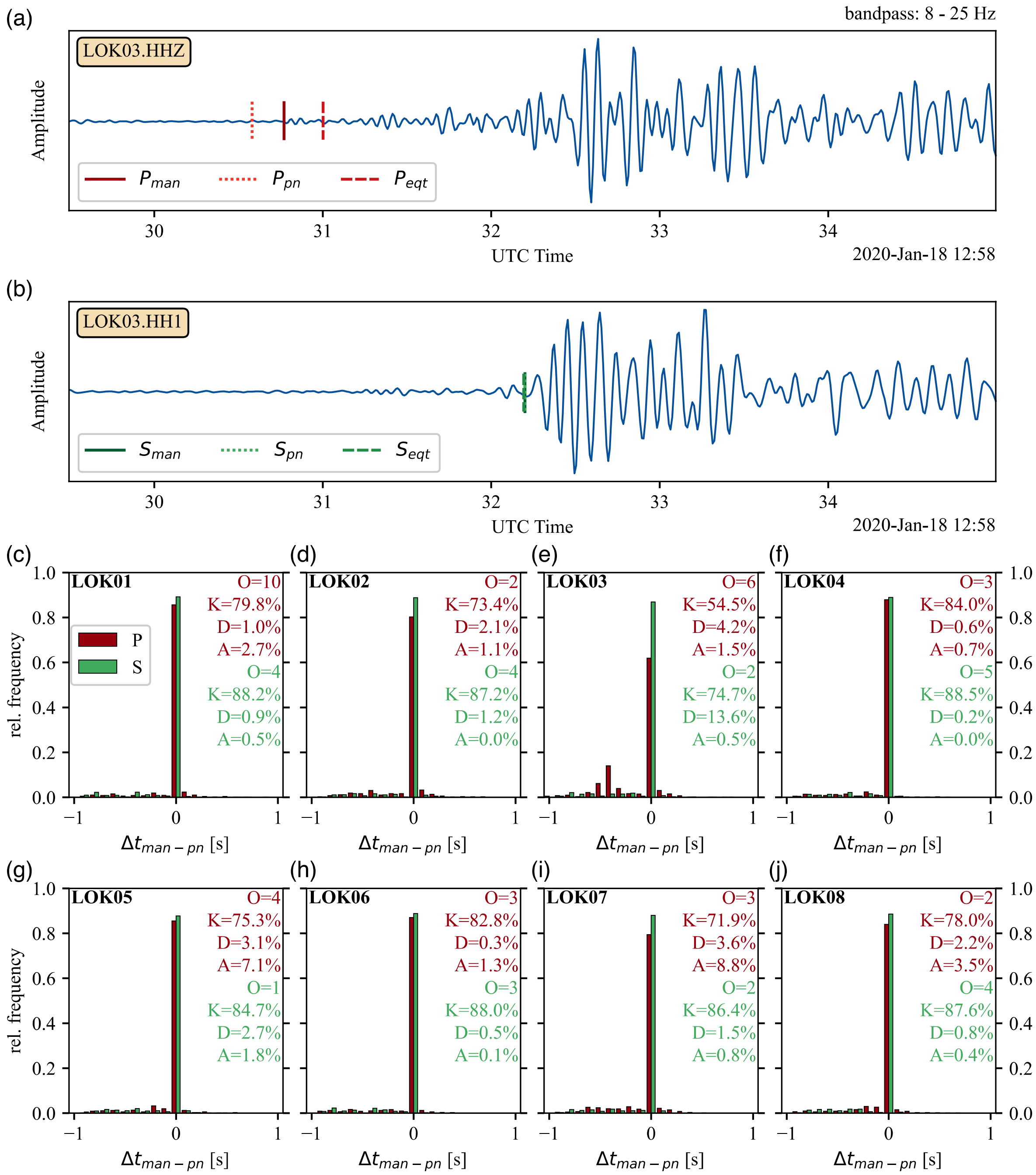

Figure 4: (a-b): Example showing the automatic (PN = PhaseNet, EQT = EQTransformer) and manual P and S phase picks at station LOK03. Data is 8-25 Hz bandpass filtered and amplitude is scaled to the maximum trace value. (c-j): Time difference between the manual and PhaseNet P (red) and S (green) phase picks for each station (bin width 0.1 s). O: Number of picks outside the shown x-limits, K: Percentage of PhaseNet picks that were kept unchanged, D: Percentage of deleted PhaseNet picks, A: Percentage of manually added picks in final picks.

The Effect of Manual Re-Picking on Location Quality, Magnitude Distribution, and Seismicity Pattern

To assess the effect of manual re-picking on the quality of the location results we located the sub-catalog of 1534 events both with the original PhaseNet picks and the manually refined picks. The effect of the manual re-picking on the quality of the location results is most significant for the resulting RMS residual distribution (Figure 5 a,b). Apart from a single event, the location RMS residual is below 0.2 s for the re-picked phases, with the histogram maximum around 0.05 s. When using the original PhaseNet picks, the RMS residual distribution is much broader with many RMS residuals between 0.1 and 0.2 s. When applying location quality control thresholds to select well-constrained events, the manually re-picked dataset retains 1391 events while 968 events are left for the PhaseNet picked catalog (Figure 5). Two of the automatically picked events are missing in the manual catalog as too many erroneous picks were deleted for these events to be located by NonLinLoc. Manual re-picking of the sub-catalog does not have a significant effect on the magnitude of completeness (Figure 5, g, h). The difference from the maximum curvature method is only marginal as for each catalog the number of events in the non-cumulative 1.0 and 1.2 bins is very close to each other. The goodness-of-fit method gives the same magnitude of completeness for both catalogs (Mcg = 1.0). However, after applying the quality control thresholds to the sub-catalogs, the number of PhaseNet picked events with ML ≥ 1.0 is only 581 compared to 1028 events within the manually refined sub-catalog such that a catalog completeness above Mc 1.0 appears unlikely for the PhaseNet picked event catalog. Judging from the similar completeness estimates, it seems that erroneous picks in the PhaseNet picked event catalog, that result in failing to meet the location quality control thresholds, are not limited to small magnitudes but occur throughout the entire range of magnitudes (see also Figure S8 in Supplemental Material).

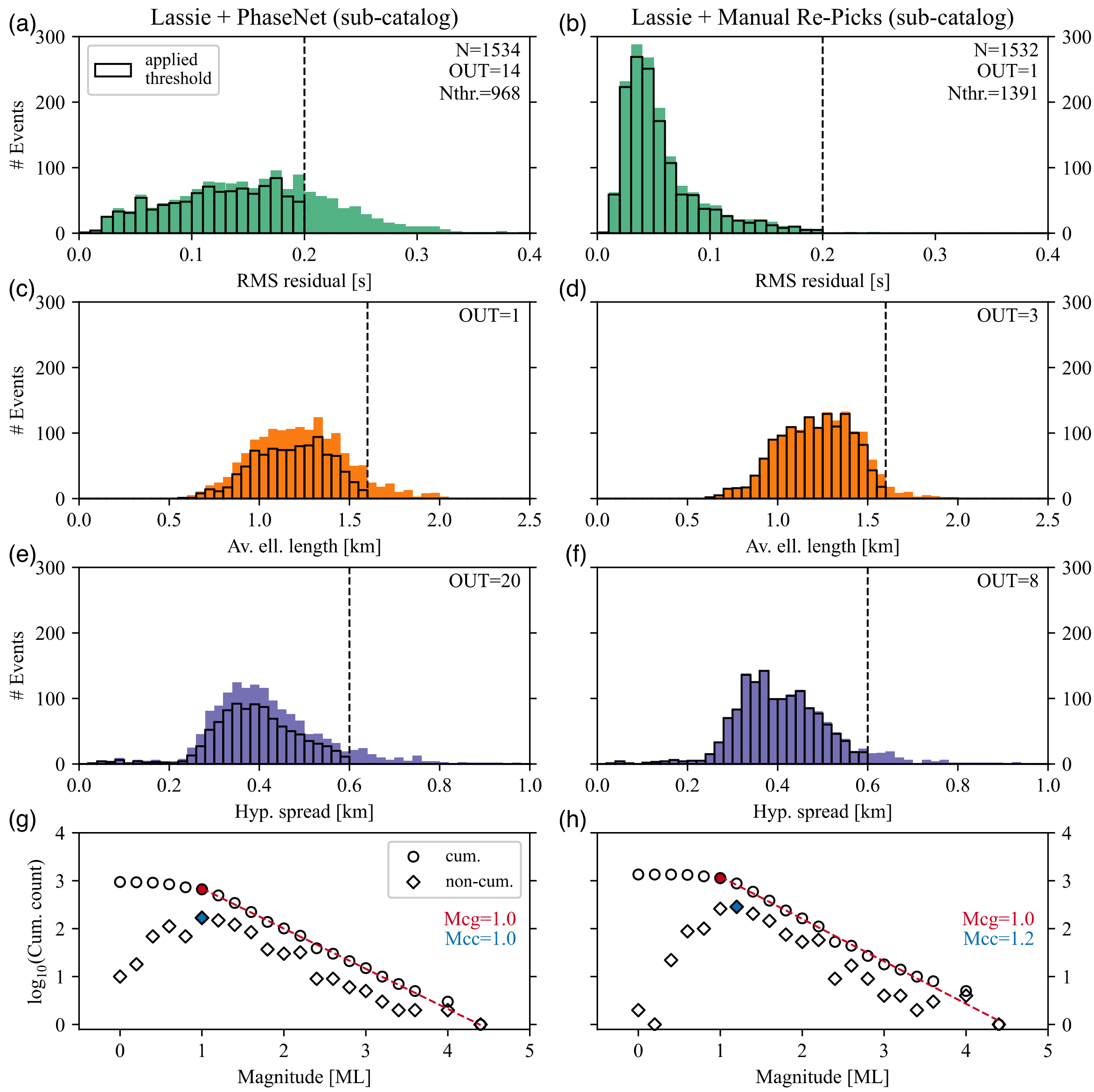

Figure 5: Statistics for the NonLinLoc location results comparing the catalogs based on automatic PhaseNet picks (a,c,e,g) and manual re-picks (b,d,f,h). Shown are the RMS residual (a,b, bin width 0.01 s), average error-ellipsoid length (c,d, bin width 0.05 km), hypocentral spread (e,f, bin width 0.02 km), and local magnitude ML (g,h, bin width 0.2, Mcg = goodness of fit; Mcc = maximum curvature) distributions after applying the location quality control thresholds (dashed lines). Black bars show the histogram distributions after applying thresholds. OUT indicates the number of events above the shown range. Nthr. = number of events below thresholds.

Plotting both sub-catalogs and comparing the resulting seismicity patterns shows that the main features of the manually re-picked sub-catalog are also visible when using the automatic PhaseNet phase picks: A central main band of seismicity, two clusters of events towards the Northwest and sparse seismicity Southeast of it (Figure 6, a, c). Similarly both cross-sections show two distinct clusters of seiscmicity at depths of ~3-7 km. However, seismicity around these main features appears more scattered in the automatically created sub-catalog compared to the sharp boundaries of these features in the manually re-picked sub-catalog (Figure 6, b, d, Figure S9 in Supplement).

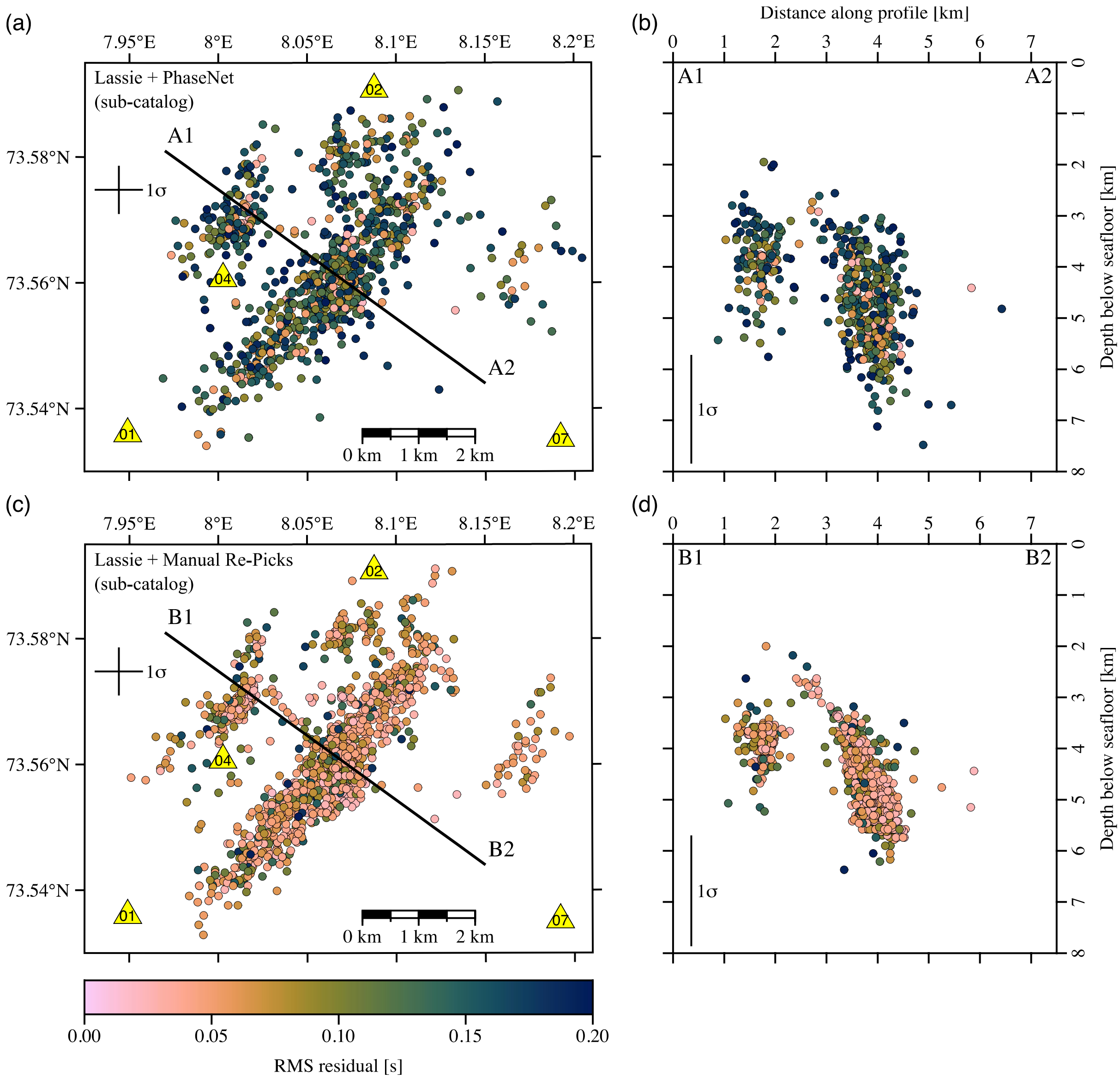

Figure 6: Close-up map and cross-section of well-constrained earthquake locations of the PhaseNet picked (a-b, 968 events) and the manually re-picked sub-catalogs (c-d, 1391 events). Yellow triangles indicate the OBS stations. Cross-sections include earthquakes within 1 km distance of the profiles in (a) and (c). 1σ-uncertainty in map view is the average uncertainty of the error-ellipses of all plotted events, for the cross-sections it is the average of the vertical error of all events.

Although the differences between the PhaseNet and manual picks are mainly within the range of ± 0.1 s (Figure 4, c-h), there are overall more well-constrained events with a significantly lower RMS residual after manual re-picking (Figure 5, a-b). Therefore, a geological interpretation from the manual re-picks would be able to describe features more precisely (e.g., more accurate dipping angle of seismicity or spatial cluster characteristics).

Fully Automatic Earthquake Catalog Compilation

Here, we compare the located earthquake catalogs from the three automatic approaches, each applied to the ~12-months continuous dataset. We imposed the same quality control criteria after location that were applied to the sub-catalogs (Figure 5). For both approaches using PhaseNet as a phase picker, we see similar results in location quality (Figure 7). When using Lassie as an event detector and subsequently only retaining PhaseNet picks on the detected events that can be associated with the theoretical arrival times calculated by Lassie, a total of 20,626 Lassie-detected events with a sufficient number of phase picks could be located out of a total of initial 112,315 Lassie detections. 2745 events are considered as well-constrained according to the quality thresholds (Figure 7, a). Using PhaseNet followed by the GaMMA associator yielded a total of 19,450 events of which 19,031 events could be located and 3011 are left as well-constrained events (Figure 7, b). Using EQTransformer followed by the GaMMA associator, NonLinLoc located 36,021 of initially 70,722 detected events with 1769 of them being left as well-constrained events (Figure 7, c). The RMS residual distribution of the located events from EQTransformer and GaMMA catalog is much broader compared to the other two approaches and 8966 events have RMS residuals beyond the shown x-limit of 0.6 s (Figure 7, c). However, the average error ellipsoid length and hypocentral spread do not show such a broad distribution with many events below the applied thresholds (Figure 7, f, i) suggesting that the RMS residual criterium contributes most to the quality control.

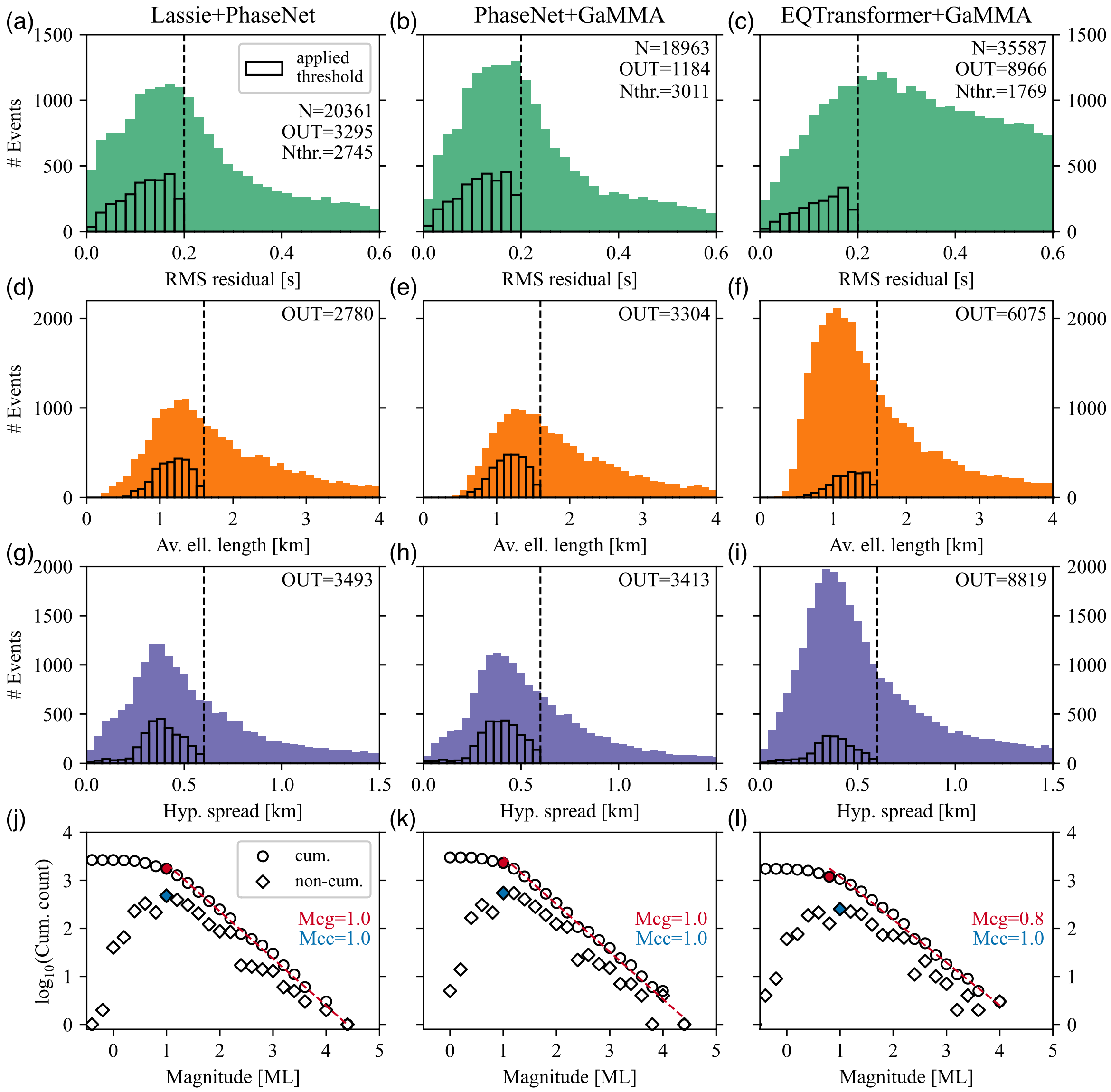

Figure 7: Same as Figure 5 but for the automatic catalog compilation approaches ran on the ~12-months continuous dataset: Lassie with PhaseNet (a,d,g,j), PhaseNet with GaMMA (b,e,h,k), and EQTtransformer with GaMMA (c,f,i,l).

Similar to the observations between the automatically and manually re-picked sub-catalogs, the three automatic approaches result in a magnitude of completeness of ML = 1.0. The only difference being the goodness-of-fit method for the EQTransformer and GaMMA approach (Mcc = 0.8), where the difference to the ML = 1.0 bin is small (Figure 7, l). After applying the quality control thresholds on the three catalogs, the number of events with ML ≥ 1.0 is 1510 for the Lassie and PhaseNet catalog, 2037 for the PhaseNet and GaMMA catalog, and 947 for the EQTransformer and GaMMA catalog. As observed in Figure 5, e-f, this shows that neither approach seems to systematically dismiss or favor small or large magnitude events and location quality control criteria affect events of all magnitudes (see also Figure S8 in Supplement).

From the resulting seismicity patterns based on the well-constrained events of the automatic approaches (Figure 8) we can see similar features as for the manually refined sub-catalog (Figure 6): a central, clearly dipping main band of seismicity and a cluster of seismicity towards the Northwest (Figure 8). However, for the EQTransformer and GaMMA approach, the clusters appear more scattered compared to the PhaseNet picked approaches. The RMS residuals of all three approaches hardly differ.

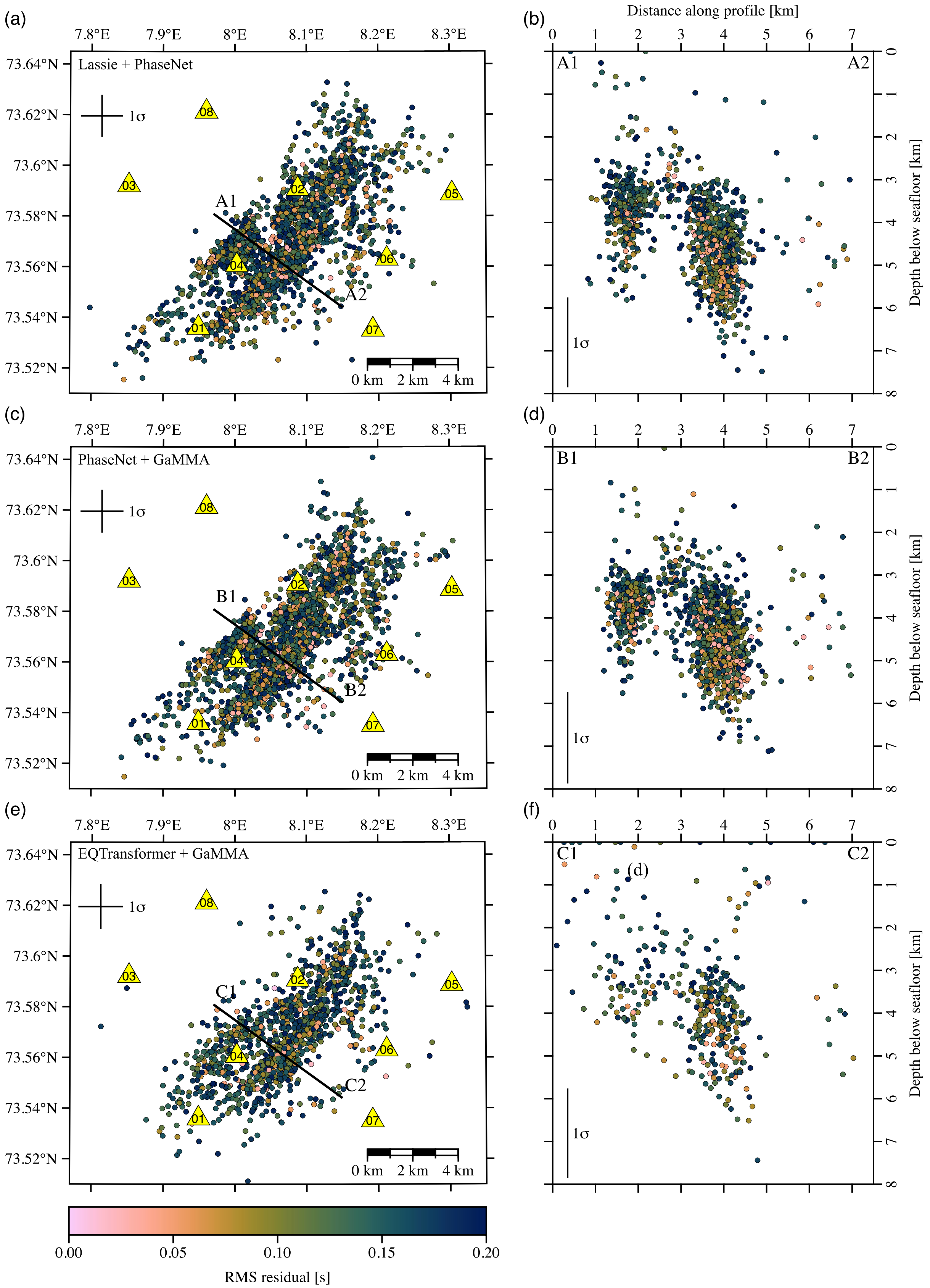

Figure 8: Similar to Figure 6 but shown here are the well-constrained events from the automatic catalog compilation approaches: (a-b) Lassie and PhaseNet, (c-e) PhaseNet and GaMMA, (e-f) EQTransformer and GaMMA.

Discussion

In this study we showed how modern deep-learning and network-based alogrithms can effectively be utilized in workflows for automatic earthquake catalog compilation from local OBS datasets. Implementing Lassie, PhaseNet, and the GaMMA associator into the workflow can save a lot of time compared to the manual work while resulting in interpretable seismicity after applying strict location quality control criteria (Figure 8, a-d). However, finding the suitable approach depending on the dataset and aim of the study requires extra care. The performance of the automatic approaches can vary strongly, the limits of interpretable seismicity have to be considered, and manual re-evaluation of automatic detections and picks can still be necessary.

Performance of Automatic Approaches in Catalog Compilation

With the strongly dataset-dependent performance of deep-learning phase pickers (H. Chen *et al.*, 2022, J. Chen et al., 2023; Münchmeyer *et al.*, 2022) and the wide range of noise conditions in OBS datasets (Stähler *et al.*, 2018; Trabattoni *et al.*, 2023), using an event detection algorithm that does not rely on phase pickers is favorable. In this study both deep-learning approaches resulted in either a catalog with many additional detections and poor pick quality (EQTransformer and GaMMA) or a catalog that systematically excludes smaller events (PhaseNet and GaMMA). Both approaches share in common, that the initial processing step is performed on single stations without using the concurrent record of waveforms by a seismic network and therefore entirely rely on accurate phase picking at this stage. The network-based Lassie detector in turn first exploits the contribution of the wave amplitudes of several seismic stations of a network to a joint detection function without having to rely on accurate phase picks. With Lassie, we thus obtained an unpicked catalog of 112,315 events.

While the majority of these events could not be picked well enough by PhaseNet to obtain a sufficient number of phases for event location (Figure 7, see Figure S10 in Supplement), the advantage with this approach is that the events are registered in the catalog and preserved for subsequent processing steps. For example, manual refinement of picks during swarm activity could be envisaged. Despite its network-based character, Lassie includes many SDEs visible only on single stations in the initial catalog. These will only be removed after phase picking by requiring a minimum number of phases for subsequent location. The catalogs based on single station phase picking as initial processing steps effectively discard SDEs already during event association by the GaMMA associator with the chosen requirement of at least three P picks from different stations for event detection.

We observed a high detection rate with EQTransformer and GaMMA, but this is mostly due to EQTransformer picking three times the amount of P phases compared to PhaseNet while the location results show that the pick quality is worse compared to the other approaches. In this study, PhaseNet as a picker combined with the GaMMA associator resulted in less false-positives compared to EQTransformer. However, the performance of the used pickers on OBS data cannot be generalized. For example Chen *et al.* (2022) reported less false-positives when using EQTransformer compared to PhaseNet on an OBS dataset from the Southern Mariana Trench. With many events in their study located at distances > 10 km from the nearest station, and EQTransformer being trained on a global event catalog with events at distances of hundreds of kilometers, EQTransformer may not be applicable to local OBS studies with small aperture networks while performing well in regional studies. This agrees with the better pick performance of EQTransformer for a local OBS study when artificially increasing S-P travel times of the input data for EQTransformer by resampling it to 200 Hz (Gong *et al.*, 2022).

Future models trained on OBS datasets could improve the performance of both event detection and phase picking. With the wide range of OBS instruments, network sizes, and regional specific noise levels even within a deployment (Stähler *et al.*, 2016; Parnell-Turner *et al.*, 2020; Meier *et al.*, 2021; Chen *et al.*, 2022; Trabattoni *et al.*, 2023), large, manually picked training datasets are needed. As already observed for land-based datasets, the performance of deep-learning models can vary depending on the dataset they are trained and used on (García *et al.*, 2022). Thus, future approaches utilizing OBS-trained deep-learning models will likely still require manual supervision depending on the dataset and aim of the study.

Schemes of Manual Re-Picking depending on the Aim of the Study

We showed that available land-based deep-learning models can already be utilized during the workflow to save much of the time needed for manual phase picking. When using a suitable approach for the dataset (for this study: PhaseNet as a phase picker) the majority of P and S picks are within ± 0.1 s of a manual pick (Figure 4, c-j). Manual revision mainly includes removing false or misidentified picks and adding missed picks. Many badly picked events can be automatically removed from the catalog after location by applying strict location quality criteria. Resulting seismicity patterns from automatically picked well-constrained events are comparable to manually revised well-constrained events (Figure 6). Using a location algorithm that reliably recognizes and downweighs outlier picks during location can further improve the location results of fully automatic approaches, provided the majority of automatic picks is of good quality (Figure 4, c-j). For example, the EDT_OT_WT inversion scheme implemented in NonLinLoc downweighs outliers and the location results are very similar for both the PhaseNet picked and manually re-evaluated sub-catalogs with picks at ≥ 7 stations (see Figure S11 in Supplement). Location quality is improved compared to the GAU_ANALYTICAL inversion scheme (see Figure S12 in Supplement). However, if only few stations contribute phase picks, outliers may not be correctly identified and potentially high-quality picks may be rejected unless a measure of the pick quality (e.g., signal to noise ratio, phase probability) is considered during inversion. Therefore, weighing schemes as the EDT_OT_WT inversion scheme are best used for large networks, where a sufficient number of picks per event are available and outliers can be reliably identified as such.

A fully automatic approach can give a good general overview of the recorded seismicity. However, manual phase refinement retains more high-quality events and achieves an overall better location quality. Additionally, even badly picked events in automatic approaches may result in location parameters that pass the quality control thresholds and thus, extra care should be taken when interpreting the resulting seismicity patterns. For an in-depth analysis of small-scale features (e.g., intrusions, fluid flow, aseismic areas) manual re-picking should still be applied. This becomes especially important with small-scale local OBS deployments, where one or two false picks significantly impact the location result. Here, a detailed analysis of the location results and phase picks is needed to, e.g., identify a systematically mis-picked P phase at a single station (Figure 4, e). Furthermore, for a robust velocity model and station correction terms a high-quality, manually re-picked sub-catalog should be compiled as both impact the location results (Grevemeyer *et al.*, 2019; Schlindwein, 2020) and the automatic pick accuracy is not sufficient yet (Chen *et al.*, 2023). However, our study shows that fully automatic procedures and a preliminary location can give a good overview catalog that serves as a basis for subsequent detailed analysis. Depending on the aim of the study a sub-catalog of the best-constrained events to develop a velocity model can be extracted and refined by manual re-picking. Likewise, the manual labor of phase pick refining can be concentrated on previously poorly constrained events. Distributions of RMS (as in Figure 7, a-c) or station residuals (as in Figure 4, c-j) based on the automatically compiled preliminary catalog give an excellent overview of the data quality and can help to tailor dataset specific criteria for optimally targeting manual labor in order to retain more events that pass the quality control criteria.

Catalog Completeness

When comparing the magnitude ranges of the resulting well-constrained catalogs (Figure 5, g-h; Figure 7, j-l) we observed no correlation between the event magnitude and the pick quality, e.g., events of smaller magnitudes being more prone to bad phase picking than larger events (see also Figure S8 and Figure S13 in Supplement). Instead, events with magnitudes of all sizes may be discarded on the basis of bad location parameters. Hence, magnitude of completeness thresholds and b-values determined from automatically compiled catalogs have to be considered with care. If catalog completeness is one aim of a study, it is best to use a detection approach independent of phase picks to detect as many events as possible and manually re-pick events above a targeted completeness threshold. Subsequent event location and magnitude determination results in robustly located events with only few events being removed by location quality criteria (Figure 5), such that completeness can be achieved and robust estimates of b-values can be obtained.

Conclusions and Recommendations

We showed that modern deep-learning and geometry-based earthquake detection and phase picking algorithms can already be utilized to obtain located earthquake catalogs from a local OBS dataset. All automatic approaches result in a similar seismicity pattern and seemingly catalog completeness. The main differences are the location quality and thus, the number of well-constrained events after applying quality control criteria varies. Good, geologically interpretable results were achieved with the combination of Lassie and PhaseNet as well as PhaseNet and GaMMA. EQTransformer is not working as well for local seismicity in the marine environment. The sharpest seismicity patterns can be achieved by manually re-picking the automatic picks. Manual picks should also be the base for developing a velocity model or for local tomography, as the quality of available pickers is still insufficient for these purposes. For large OBS networks with a sufficient number of picks per event, using an inversion scheme that identifies and downweighs outliers (e.g., EDT_OT_WT in NonLinLoc) can further improve the location results. Using a network-based and pick-independent event detection software, like Lassie, results in an initial event catalog that includes also weak events for further analysis that go undetected in phase-pick dependent automatic procedures.

When applying fully automatic catalog compilation approaches, we recommend to evaluate the performance of the used event detectors and phase pickers on a reference subset of manually detected and picked events. This way, appropriate algorithms can be tested and chosen based on the aim of the study, e.g., if false-positives or a potential systematic omission of small events are a concern. Additionally, station-specific systematic picking errors can be identified and their impact on the resulting seismicity pattern evaluated. Dataset- and purpose-tailored schemes of manual re-picking can then be developed to minimize manual work while optimizing the resulting catalog. Eventhough land-based deep-learning approaches in marine seismology still show limitations and additional supervisional steps during the automatic catalog compilation are necessary, the amount of time that can be saved compared to a completely manually compiled earthquake catalog is considerable. Datasets from large OBS networks can be automatically processed more effectively compared to small OBS networks, where the impact of a few mis-picked phases on the final location quality is more significant.

Data and Resources

The Loki dataset is accessible at GEOFON (doi:10.14470/3Z326135), Bathymetry data in Figure 1 is from Kartverket (www.kartverket.no), Figure 4 (a,b), S1, S2, S5, S6 and S7 were prepared using ObsPy (Beyreuther *et al.*, 2010; Tobias Megies *et al.*, 2011; Krischer *et al.*, 2015), Figure 1 Figure 6, Figure 8, S9, and S11 were prepared using Generic Mapping Tools version 6 (Wessel *et al.*, 2019). Supplemental Material for this article includes waveform plots of the reference days, examples of event detections from different approaches, and a plot showing the number of events manually detected from the reference day which are automatically detected and left as well-constrained events after location. Further it includes plots for the magnitude distributions of events rejected by the location quality thresholds and the relation of the magnitude to the average error ellipsoid length for all catalogs, a plot showing the distance between located events for both sub-catalogs, and two plots showing the location results using NonLinLoc's EDT_OT_WT inversion scheme. Electronic Supplement includes the five located earthquake catalogs used in this manuscripts.

Acknowledgments

The authors thank crew and scientist of the R/V G.O. Sars cruises for deployment and recovery of the OBS stations and Przemyslaw Domel for helpful discussions on the manuscript. The constructive comments of editor Romain Jolivet and two anonymous reviewers are gratefully acknowledged. OBS were provided by the DEPAS Pool (Schmidt-Aursch and Haberland, 2017).

References

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces, *Bulletin of the Seismological Society of America* **68**, no. 5, 1521–1532, doi: 10.1785/BSSA0680051521.

Baillard, C., W. C. Crawford, V. Ballu, C. Hibert, and A. Mangeney (2014). An automatic kurtosis-based P-and S-phase picker designed for local seismic networks, *Bulletin of the Seismological Society of America* **104**, no. 1, 394–409, doi: 10.1785/0120120347.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A Python Toolbox for Seismology, *Seismological Research Letters* **81**, no. 3, 530–533, doi: 10.1785/gssrl.81.3.530.

Bornstein, T., D. Lange, J. Münchmeyer, J. Woollam, A. Rietbrock, G. Barcheck, I. Grevemeyer, and F. Tilmann (2023). PickBlue: Seismic phase picking for ocean bottom seismometers with deep learning, arXiv, doi: 10.48550/arXiv.2304.06635.

Brodie, D. C., and R. A. Dunn (2015). Low frequency baleen whale calls detected on ocean-bottom seismometers in the Lau basin, southwest Pacific Ocean, *The Journal of the Acoustical Society of America* **137**, no. 1, 53–62, doi: 10.1121/1.4904556.

Chen, J., W. C. Crawford, and M. Cannat (2023). Microseismicity and lithosphere thickness at a nearly-amagmatic oceanic detachment fault system, *Nat Commun* **14**, no. 1, 430, doi: 10.1038/s41467-023-36169-w.

Chen, H., H. Yang, G. Zhu, M. Xu, J. Lin, and Q. You (2022). Deep Outer-Rise Faults in the Southern Mariana Subduction Zone Indicated by a Machine-Learning-Based High-Resolution Earthquake Catalog, *Geophysical Research Letters* **49**, no. 12, doi: 10.1029/2022GL097779.

Domel, P., C. Hibert, V. Schlindwein, and A. Plaza-Faverola (2023). Event recognition in marine seismological data using Random Forest machine learning classifier, *Geophysical Journal International* **235**, no. 1, 589–609, doi: 10.1093/gji/ggad244.

Domel, P., S. Singhroha, A. Plaza-Faverola, V. Schlindwein, H. Ramachandran, and S. Bünz (2022). Origin and Periodic Behavior of Short Duration Signals Recorded by Seismometers at Vestnesa Ridge, an Active Seepage Site on the West-Svalbard Continental Margin, *Frontiers in Earth Science* **10**, doi: 10.3389/feart.2022.831526.

Douilly, R., W. L. Ellsworth, E. Kissling, A. M. Freed, A. Deschamps, and B. Mercier De Lépinay (2016). 3-D velocity structure in southern Haiti from local earthquake tomography, *JGR Solid Earth* **121**, no. 12, 8813–8832, doi: 10.1002/2016JB013123.

Essing, D., V. Schlindwein, M. C. Schmidt-Aursch, C. Hadziioannou, and S. C. Stähler (2021). Characteristics of Current-Induced Harmonic Tremor Signals in Ocean-Bottom Seismometer Records, *Seismological Research Letters* **92**, no. 5, 3100–3112, doi: 10.1785/0220200397.

García, J. E., L. M. Fernández-Prieto, A. Villaseñor, V. Sanz, J. Ammirati, E. A. Díaz Suárez, and C. García (2022). Performance of Deep Learning Pickers in Routine Network Processing Applications, *Seismological Research Letters* **93**, no. 5, 2529–2542, doi: 10.1785/0220210323.

Gibbons, S. J., and F. Ringdal (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophysical Journal International* **165**, no. 1, 149–166, doi: 10.1111/j.1365-246X.2006.02865.x.

Gong, J., W. Fan, and R. Parnell-Turner (2022). Microseismicity Indicates Atypical Small-Scale Plate Rotation at the Quebrada Transform Fault System, East Pacific Rise, *Geophysical Research Letters* **49**, no. 3, doi: 10.1029/2021GL097000.

Grevemeyer, I., N. W. Hayman, D. Lange, C. Peirce, C. Papenberg, H. J. A. Van Avendonk, F. Schmid, L. G. de La Peña, and A. Dannowski (2019). Constraining the maximum depth of brittle deformation at slow- and ultraslow-spreading ridges using microseismicity, *Geology* **47**, no. 11, 1069–1073, doi: 10.1130/G46577.1.

Havskov, J., and L. Ottemöller (1999). SeisAn Earthquake Analysis Software, *Seismological Research Letters* **70**, no. 5, 532–534, doi: 10.1785/GSSRL.70.5.532.

Havskov, J., P. H. Voss, and L. Ottemöller (2020). Seismological observatory software: 30 yr of SEISAN, *Seismological Research Letters* **91**, no. 3, 1846–1852, doi: 10.1785/0220190313.

Heimann, S., M. Kriegerowski, M. Isken, S. Cesca, S. Daout, F. Grigoli, C. Juretzek, T. Megies, N. Nooshiri, A. Steinberg, *et al.* (2017). Pyrocko - An open-source seismology toolbox and library, *GFZ Data Services*, doi: 10.5880/GFZ.2.1.2017.001.

Hilmo, R., and W. S. D. Wilcock (2020). Physical Sources of High-Frequency Seismic Noise on Cascadia Initiative Ocean Bottom Seismometers, *Geochem. Geophys. Geosyst.* **21**, no. 10, doi: 10.1029/2020GC009085.

Hutton, L. K., and D. M. Boore (1987). The ML scale in Southern California, *Bulletin of the Seismological Society of America* **77**, no. 6, 2074–2094, doi: 10.1785/BSSA0770062074.

Jeddi, Z., L. Ottemöller, M. B. Sørensen, S. Rezaei, S. J. Gibbons, M. L. Strømme, P. H. Voss, and T. Dahl-Jensen (2021). Improved seismic monitoring with obs deployment in the arctic: A pilot study from offshore western svalbard, *Seismological Research Letters* **92**, no. 5, 2705–2717, doi: 10.1785/0220200471.

Johansen, S. E., M. Panzner, R. Mittet, H. E. F. Amundsen, A. Lim, E. Vik, M. Landrø, and B. Arntsen (2019). Deep electrical imaging of the ultraslow-spreading Mohns Ridge, *Nature* **567**, no. 7748, 379–383, doi: 10.1038/s41586-019-1010-0.

Kissling, E., U. Kradolfer, and H. Maurer (1995). Velest user's guide - short introduction, *Institute of Geophysics, ETH Zurich*.

Krischer, L., T. Megies, R. Barsch, M. Beyreuther, T. Lecocq, C. Caudron, and J. Wassermann (2015). ObsPy: a bridge for seismology into the scientific Python ecosystem, *Comput. Sci. Disc.* **8**, no. 1, 014003, doi: 10.1088/1749-4699/8/1/014003.

Lindsey, N. J., and E. R. Martin (2021). Fiber-Optic Seismology, *Annu. Rev. Earth Planet. Sci.* **49**, no. 1, 309–336, doi: 10.1146/annurev-earth-072420-065213.

Lomax, A., and A. Curtis (2001). Fast probabilistic earthquake location in 3D models using Oct-Tree importance sampling, *Geophys. Res. Abstr.* **3**.

Lomax, A., A. Michelini, and A. Curtis (2009). Earthquake Location, Direct, Global-Search Methods, in *Encyclopedia of Complexity and Systems Science* R. A. Meyers(Editor), Springer, New York, NY, 1–33, doi: 10.1007/978-3-642-27737-5_150-2.

Lomax, A., J. Virieux, P. Volant, and C. Berge-Thierry (2000). Probabilistic Earthquake Location in 3D and Layered Models, in *Advances in Seismic Event Location* C. H. Thurber, and N. Rabinowitz(Editors), Springer Netherlands, Dordrecht, Modern Approaches in Geophysics, 101–134, doi: 10.1007/978-94-015-9536-0_5.

Majstorović, J., S. Giffard-Roisin, and P. Poli (2021). Designing Convolutional Neural Network Pipeline for Near-Fault Earthquake Catalog Extension Using Single-Station Waveforms, *JGR Solid Earth* **126**, no. 7, e2020JB021566, doi: 10.1029/2020JB021566.

Meier, M., V. Schlindwein, and F. Schmid (2022). Magmatic Activity and Dynamics of Melt Supply of Volcanic Centers of Ultraslow Spreading Ridges: Hints from Local Earthquake Tomography at the Knipovich Ridge, *Geochemistry, Geophysics, Geosystems*, doi: 10.1029/2021gc010210.

Meier, M., V. Schlindwein, J.-R. Scholz, J. Geils, M. C. Schmidt-Aursch, F. Krüger, W. Czuba, and T. Janik (2021). Segment-scale seismicity of the ultraslow spreading Knipovich Ridge, *Geochemistry, Geophysics, Geosystems*, doi: 10.1029/2020gc009375.

Mousavi, S. M., and G. C. Beroza (2022). Deep-learning seismology, *Science* **377**, no. 6607, eabm4470, doi: 10.1126/science.abm4470.

Mousavi, S. M., W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking, *Nature Communications* **11**, no. 1, doi: 10.1038/s41467-020-17591-w.

Münchmeyer, J., J. Woollam, A. Rietbrock, F. Tilmann, D. Lange, T. Bornstein, T. Diehl, C. Giunchi, F. Haslinger, D. Jozinović, *et al.* (2022). Which Picker Fits My Data? A Quantitative Evaluation of Deep Learning Based Seismic Pickers, *Journal of Geophysical Research: Solid Earth* **127**, no. 1, 1–22, doi: 10.1029/2021JB023499.

Park, Y., G. C. Beroza, and W. L. Ellsworth (2023). A Mitigation Strategy for the Prediction Inconsistency of Neural Phase Pickers, *Seismological Research Letters* **94**, no. 3, 1603–1612, doi: 10.1785/0220230003.

Park, Y., G. C. Beroza, and W. L. Ellsworth (2022). Basement Fault Activation before Larger Earthquakes in Oklahoma and Kansas, *The Seismic Record* **2**, no. 3, 197–206, doi: 10.1785/0320220020.

Parnell-Turner, R., R. A. Sohn, C. Peirce, T. J. Reston, C. J. MacLeod, R. C. Searle, and N. M. Simão (2020). Seismicity trends and detachment fault structure at 13°N, Mid-Atlantic Ridge, *Geology* **49**, no. 3, 320–324, doi: 10.1130/G48420.1.

Ross, Z. E., M. Meier, E. Hauksson, and T. H. Heaton (2018). Generalized Seismic Phase Detection with Deep Learning, *Bulletin of the Seismological Society of America* **108**, no. 5A, 2894–2901, doi: 10.1785/0120180080.

Schlindwein, V. (2020). Constraining the maximum depth of brittle deformation at slow- and ultraslow-spreading ridges using microseismicity: Comment, *Geology* **48**, no. 5, 501–501, doi: 10.1130/G47444C.1.

Schmidt-Aursch, M. C., and C. Haberland (2017). DEPAS (Deutscher Geräte-Pool für amphibische Seismologie): German Instrument Pool for Amphibian Seismology, *JLSRF* **3**, A122, doi: 10.17815/jlsrf-3-165.

Schweitzer, J. (2001). Hyposat-An enhanced routime to locate seismic events, *Pure and Applied Geophysics* **158**, no. 1–2, 277–289, doi: 10.1007/pl00001160.

Schweitzer, J. (2018). User Manual HYPOSAT 6 and HYPOMOD 2, no. May, 1–38, doi: 10.2312/GFZ.NMSOP-3.

Scotto di Uccio, F., A. Scala, G. Festa, M. Picozzi, and G. C. Beroza (2023). Comparing and integrating artificial intelligence and similarity search detection techniques: application to seismic sequences in Southern Italy, *Geophysical Journal International* **233**, no. 2, 861–874, doi: 10.1093/gji/ggac487.

Seydoux, L., R. Balestriero, P. Poli, M. de Hoop, M. Campillo, and R. Baraniuk (2020). Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning, *Nat Commun* **11**, no. 1, 3972, doi: 10.1038/s41467-020-17841-x.

Spica, Z. J., J. C. Castellanos, L. Viens, K. Nishida, T. Akuhara, M. Shinohara, and T. Yamada (2022). Subsurface Imaging With Ocean-Bottom Distributed Acoustic Sensing and Water Phases Reverberations, *Geophysical Research Letters* **49**, no. 2, doi: 10.1029/2021GL095287.

Stähler, S. C., M. C. Schmidt-Aursch, G. Hein, and R. Mars (2018). A self-noise model for the German DEPAS OBS pool, *Seismological Research Letters* **89**, no. 5, 1838–1845, doi: 10.1785/0220180056.

Stähler, S. C., K. Sigloch, K. Hosseini, W. C. Crawford, G. Barruol, M. C. Schmidt-Aursch,

M. Tsekhmistrenko, J.-R. Scholz, A. Mazzullo, and M. Deen (2016). Performance report of the RHUM-RUM ocean bottom seismometer network around La Réunion, western Indian Ocean, *Adv. Geosci.* **41**, 43–63, doi: 10.5194/adgeo-41-43-2016.

Tarantola, A., and B. Valette (1981). Inverse problems = Quest for information, *J Geophys* **50**, no. 1, 159–170.

Tary, J. B., L. Géli, C. Guennou, P. Henry, N. Sultan, N. Çağatay, and V. Vidal (2012). Microevents produced by gas migration and expulsion at the seabed: a study based on sea bottom recordings from the Sea of Marmara: Microevents produced by gas migration, *Geophysical Journal International* **190**, no. 2, 993–1007, doi: 10.1111/j.1365-246X.2012.05533.x.

Tobias Megies, Moritz Beyreuther, Robert Barsch, Lion Krischer, and Joachim Wassermann (2011). ObsPy – What can it do for data centers and observatories?, *Annals of Geophysics* **54**, no. 1, doi: 10.4401/ag-4838.

Trabattoni, A., G. Barruol, R. Dréo, and A. Boudraa (2023). Ship detection and tracking from single ocean-bottom seismic and hydroacoustic stations, *The Journal of the Acoustical Society of America* **153**, no. 1, 260–273, doi: 10.1121/10.0016810.

Wessel, P., J. F. Luis, L. Uieda, R. Scharroo, F. Wobbe, W. H. F. Smith, and D. Tian (2019). The Generic Mapping Tools Version 6, *Geochem. Geophys. Geosyst.* **20**, no. 11, 5556–5564, doi: 10.1029/2019GC008515.

Wiemer, S., and M. Wyss (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan, *Bulletin of the Seismological Society of America* **90**, no. 4, 859–869, doi: 10.1785/0119990114.

Williams, M. C., A. M. Trehu, and J. Braunmiller (2010). Local Earthquake Detection in

Marine Environments Using Seismic Signal Parameters, in *AGU Fall Meeting Abstracts*, S53A-1965.

Wu, X., S. Huang, Z. Xiao, and Y. Wang (2022). Building Precise Local Submarine Earthquake Catalogs via a Deep-Learning-Empowered Workflow and its Application to the Challenger Deep, *Front. Earth Sci.* **10**, 817551, doi: 10.3389/feart.2022.817551.

Yaroshenko, G., I. Koulakov, N. Al-Arifi, S. Qaysi, and S. E. Khrepy (2022). Structure of the magma plumbing system beneath Semisopochnoi Island (Aleutian Arc) inferred from seismic tomography, *Scientific Reports* **12**, no. 1, 1–12, doi: 10.1038/s41598-022-14794-7.

Zhu, W., and G. C. Beroza (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method, *Geophysical Journal International* **216**, no. 1, 261–273, doi: 10.1093/gji/ggy423.

Zhu, W., I. W. McBrearty, S. M. Mousavi, W. L. Ellsworth, and G. C. Beroza (2022). Earthquake Phase Association Using a Bayesian Gaussian Mixture Model, *JGR Solid Earth* **127**, no. 5, doi: 10.1029/2021JB023249.

Co-Author Mailing List

Vera Schlindwein. Adress: Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Am Alten Hafen 26, 27568 Bremerhaven, Germany; E-mail: vera.schlindwein@awi.de; ORCID iD: https://orcid.org/0000-0001-5570-2753

Figure Captions List

Figure 1: a): Overview map showing the study area (red square)) in the Norwegian-Greenland Sea. MR = Mohns Ridge, KR = Knipovich Ridge. b): Bathymetry of the Mohn-

Knipovich Ridge bend including the positions and numbers of the OBS stations (yellow triangles) and the position of Loki's Castle (red star). Bathymetry data from Kartverket (www.kartverket.no).

Figure 2: Schematic workflows used for this study. a): For event detection we used a subset of 11 control days and compared a manually compiled event catalog to the event detections of three automatic detectors (Lassie, PhaseNet/GaMMA, and EQTransformer (EQT) /GaMMA). b): From the continuous dataset three automatically detected and picked event catalogs were compiled (top box, right). A sub-catalog of Lassie detected and PhaseNet picked, best- constrained events was manually re-picked for comparison with the original PhaseNet picks (top box, left). For event location, both sub-catalogs (bottom box, left) and the three automatically compiled catalogs (bottom box, right) were located and location quality-control criteria were applied before comparing the location results. Numbers in squares refer to the corresponding figures.

Figure 3: Single day comparison of the three different event detection approaches with the reference event database. (a-c): Number of events automatically (green bars) and manually detected (black line, reference database) at ≥ 3 stations over time. (d-f): Number of automatically detected events that are also in the reference database (black line,) and additional events (orange bars) in relation to the Lassie detection value or GaMMA score. OUT indicates the number of manually detected event above the shown values. (g-i): Number of automatically detected events (green bars) that are part of the reference database (black line) in relation to the number of stations where the event was seen.

Figure 4: (a-b): Example showing the automatic (PN = PhaseNet, EQT = EQTransformer) and manual P and S phase picks at station LOK03. Data is 8-25 Hz bandpass filtered and amplitude

is scaled to the maximum trace value. (c-j): Time difference between the manual and PhaseNet P (red) and S (green) phase picks for each station (bin width 0.1 s). O: Number of picks outside the shown x-limits, K: Percentage of PhaseNet picks that were kept unchanged, D: Percentage of deleted PhaseNet picks, A: Percentage of manually added picks in final picks.

Figure 5: Statistics for the NonLinLoc location results comparing the catalogs based on automatic PhaseNet picks (a,c,e,g) and manual re-picks (b,d,f,h). Shown are the RMS residual (a,b, bin width 0.01 s), average error-ellipsoid length (c,d, bin width 0.05 km), hypocentral spread (e,f, bin width 0.02 km), and local magnitude ML (g,h, bin width 0.2, Mcg = goodness of fit; Mcc = maximum curvature) distributions after applying the location quality control thresholds (dashed lines). Black bars show the histogram distributions after applying thresholds. OUT indicates the number of events above the shown range. Nthr. = number of events below thresholds.

Figure 6: Close-up map and cross-section of well-constrained earthquake locations of the PhaseNet picked (a-b, 968 events) and the manually re-picked sub-catalogs (c-d, 1391 events). Yellow triangles indicate the OBS stations. Cross-sections include earthquakes within 1 km distance of the profiles in (a) and (c). 1σ-uncertainty in map view is the average uncertainty of the error-ellipses of all plotted events, for the cross-sections it is the average of the vertical error of all events.

Figure 7: Same as Figure 5 but for the automatic catalog compilation approaches ran on the ~12-months continuous dataset: Lassie with PhaseNet (a,d,g,j), PhaseNet with GaMMA (b,e,h,k), and EQTtransformer with GaMMA (c,f,i,l).

Figure 8: Similar to Figure 6 but shown here are the well-constrained events from the automatic catalog compilation approaches: (a-b) Lassie and PhaseNet, (c-e) PhaseNet and GaMMA, (e-f) EQTransformer and GaMMA.