

ArrayNet: A Combined Seismic Phase Classification and Back-Azimuth Regression Neural Network for Array Processing Pipelines

Andreas Köhler^{*1,2}  and Erik B. Myklebust¹ 

ABSTRACT

Array processing is an integral part of automatic seismic event detection pipelines for measuring apparent velocity and backazimuth of seismic arrivals. Both quantities are usually measured under the plane-wave assumption, and are essential to classify the phase type and to determine the direction toward the event epicenter. However, structural inhomogeneities can lead to deviations from the plane-wave model, which must be taken into account for phase classification and back-azimuth estimation. We suggest a combined classification and regression neural network, which we call ArrayNet, to determine the phase type and back-azimuth directly from the arrival-time differences between all combinations of stations of a given seismic array without assuming a plane-wave model. ArrayNet is trained using regional P- and S-wave arrivals of over 30,000 seismic events from reviewed regional bulletins in northern Europe from the past three decades. ArrayNet models are generated and trained for each of the ARCES, FINES, and SPITS seismic arrays. We observe excellent performance for the seismic phase classification (up to 99% accuracy), and the derived back-azimuth residuals are significantly improved in comparison with traditional array processing results using the plane-wave assumption. The SPITS array in Svalbard exhibits particular issues when it comes to array processing in the form of high apparent seismic velocities and a multitude of frost quake signals inside the array, and we show how our new approach better handles these obstacles. Furthermore, we demonstrate the performance of ArrayNet on 20 months of continuous phase detections from the ARCES array and investigate the results for a selection of regional seismic events of interest. Our results demonstrate that automatic event detection at seismic arrays can be further enhanced using a machine learning approach that takes advantage of the unique array data recorded at these stations.

KEY POINTS

- Deviations from the plane-wave model can affect seismic event detection using seismic arrays.
- We show that ArrayNet—a machine learning approach—allows improving the characterisation of seismic arrivals.
- ArrayNet in combination with other machine learning methods can enhance event detection pipelines worldwide.

Supplemental Material

INTRODUCTION

Array processing is routinely used to measure velocity and propagation direction of seismic arrivals (Ringdal and Husebye, 1982; Douglas, 2002; Rost and Thomas, 2002; Schweitzer *et al.*, 2012). Being an integral part of automatic processing pipelines for seismic event monitoring, for example, at the International Data Center (IDC) of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) or other

national data centers worldwide, this processing step usually follows seismic phase detection, and precedes seismic event association and location (e.g., Le Bras *et al.*, 1994, 2021; Fig. 1). Detection in continuous data streams is accomplished by power detectors (e.g., short-term average/long-term average [STA/LTA]) or the *F*-statistic (*F*-detector, Selby, 2008) applied to a set of deployed array beams covering the slowness space of interest (step 1). Subsequently, a refined apparent (horizontal) velocity and back-azimuth estimate of the seismic phase detection is generated in step 2 (Schweitzer *et al.*, 2012). Finally, arrivals detected on multiple stations are associated with

1. NORSAR, Kjeller, Norway,  <https://orcid.org/0000-0002-1060-7637> (AK);  <https://orcid.org/0000-0002-3056-2544> (EBM); 2. UiT The Arctic University of Norway (AK), Norway

*Corresponding author: andreas.kohler@norsar.no

Cite this article as Köhler, A., and E. B. Myklebust (2023). ArrayNet: A Combined Seismic Phase Classification and Back-Azimuth Regression Neural Network for Array Processing Pipelines, *Bull. Seismol. Soc. Am.* **113**, 2345–2362, doi: [10.1785/0120230056](https://doi.org/10.1785/0120230056)

© Seismological Society of America

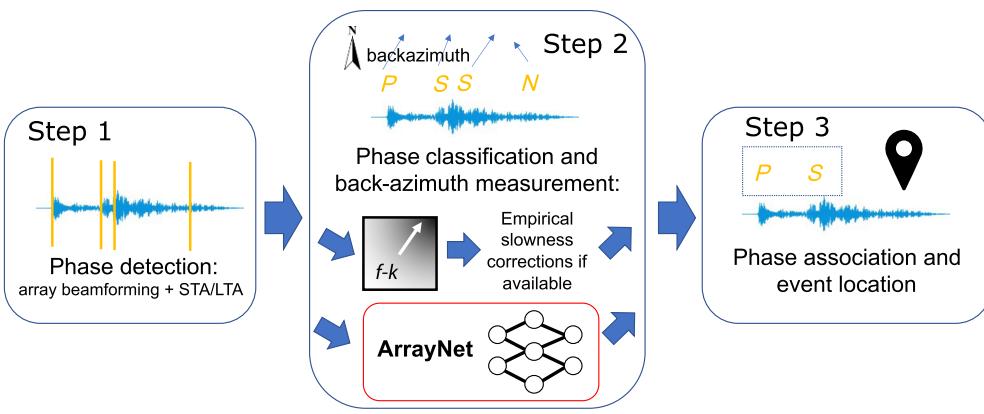


Figure 1. Conceptual sketch of an automatic array processing pipeline for seismic event detection. Our new method ArrayNet focuses on step 2, which is classification of detected seismic arrivals or phases, replacing frequency-wavenumber (f - k) analysis and any empirical slowness corrections that may be available. Letters in 2 and 3 indicate P - and S -wave arrivals and noise (N). The color version of this figure is available only in the electronic edition.

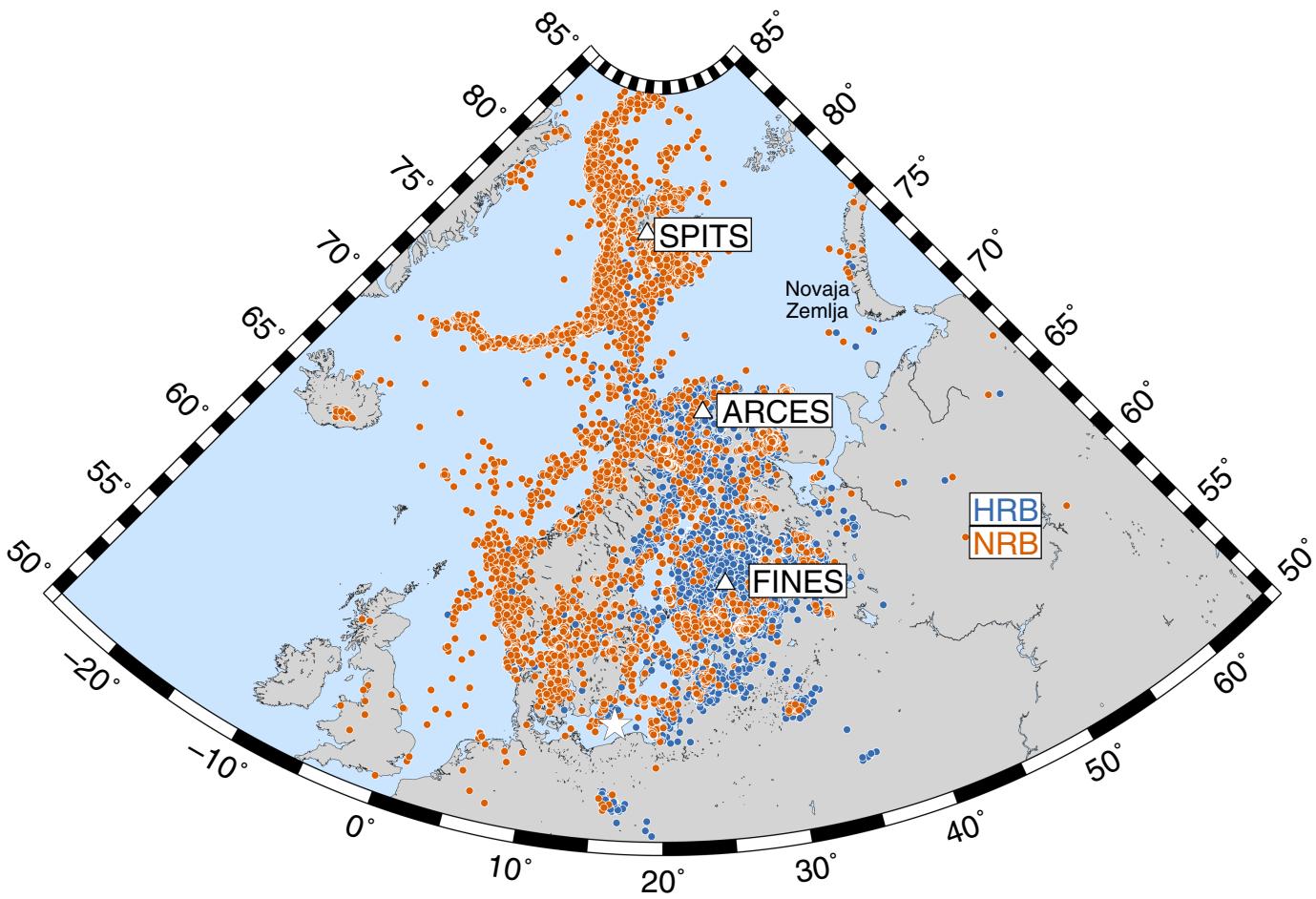
one another to form and locate events (step 3). The apparent velocity of a seismic arrival measured in step 2 is required in order to classify the type of the detected phase, that is, regional or teleseismic P - or S -wave phases, and the measured backazimuth is assumed to point toward the event epicenter, and is essential for event association and location in step 3. Phase type and backazimuth are usually measured under the plane-wave assumption using frequency-wavenumber (f - k) analysis (Capon, 1969) or other wavefront fitting algorithms such as the Progressive Multi-Channel Correlation method (PMCC, Cansi, 1995). However, local inhomogeneities below the seismic array, local topography, as well as subsurface structures between source and receiver often lead to frequency-dependent distortions of the plane-wave character, decorrelation of the wavefield between array stations, biased back-azimuth measurements with respect to the actual source direction, as well as significant deviations of the local apparent velocities compared to standard Earth models (Niazi, 1966; Berteussen, 1976; Schweitzer, 2001b; Gibbons *et al.*, 2011). The latter affects the accuracy of phase-type classification, because phase labels are assigned based on expected velocity ranges given by the chosen Earth model. In other words, step 2 of the array processing pipeline requires careful calibration and tuning of processing parameters, and often correction of the results, for each individual array (Gibbons *et al.*, 2010).

The previous attempts to take the aforementioned issues into account were based on empirical array-dependent slowness vector corrections (Schweitzer, 2001b; Ben Horin *et al.*, 2004), either in the form of back-azimuth-dependent analytic expressions or tabled slowness correction vectors for grid points in the slowness space. Seismic events with well-constrained locations were used to estimate the average slowness deviation for each slowness grid point. An alternative option is to consider empirical time or phase shifts on the coarray. The

coarray represents all combinations of two stations within an array. It includes the relative positions of each station pair, that is, the interstation distance and azimuth. A seismic event from a confined source region produces a characteristic time shift between incoming arrivals of a particular phase type for each station pair, often significantly different from the theoretically anticipated plane-wave time shifts (Kvaerna, Harris, *et al.*, 2023). This phenomenon is taking advantage of in Empirical Matched Field Processing (EMFP, Harris and Kvaerna, 2010; Gibbons *et al.*, 2017; Köhler *et al.*, 2022) to detect seismic arrivals with time or phase shift patterns on the coarray similar to that of a template event. This is accomplished by beamforming using the empirical phase shift of the template instead of using a plane-wave model and by calculating an appropriate detection statistic from the beam power evaluated at multiple frequencies (Kvaerna, Harris, *et al.*, 2023). Theoretically, EMFP can be implemented to detect and classify seismic phases from all possible source regions. However, this would require extensive effort to define empirical steering vectors from templates representing all events that can possibly be observed on a given array.

In the recent years huge advances have been made in integrating machine learning methods into seismological data analysis and processing pipelines, taking advantage of decades-long reviewed seismic event bulletins (Bergen *et al.*, 2019; Kong *et al.*, 2019; Mousavi and Beroza, 2023). The main focus is on seismic phase and event detection on single stations (e.g., Mousavi *et al.*, 2020). However, to our knowledge no machine learning methods have been developed, which explicitly take advantage of array data—using all array stations to replicate array processing (beamforming, back-azimuth measurements, and phase classification).

In this study, we focus on step 2 of the array processing pipeline by adapting a machine learning approach (Fig. 1). Instead of calculating empirical slowness corrections directly or using phase shifts from template events to build a detection statistic, we suggest a neural network architecture to learn from past event observations to determine the seismic phase type and backazimuth directly from the arrival-time differences measured on the coarray, without assuming a certain wavefield geometry. As input data we suggest to use phase shifts at multiple frequencies measured by estimating the cross spectrum of each coarray element. The neural network is a combined



classification (for phase type) and regression (for backazimuth) network, and is trained using regional *P* and *S* arrivals of seismic events from two reviewed regional bulletins in northern Europe, including events from the past three decades. We will show that this method is able to classify the phase type without first measuring the apparent velocity and without using preset velocity thresholds, and that an unbiased backazimuth is obtained pointing directly toward the event origin location. In the following section, we first introduce the array waveform and event data sets being used. After that, we describe the methodology including an approach to augment training data taking into account source directions underrepresented in the event bulletins. Next, we apply the method to different seismic arrays in northern Europe, and in the European Arctic and present the results. The subsequent section focuses on the evaluation of the method by comparing results to an existing automatic processing pipeline focusing on an array in northern Norway. Finally, we test the method on a selected set of seismic events of special interest.

SEISMIC EVENT AND WAVEFORM DATA

We use waveform data recorded on three International Monitoring System (IMS of the CTBTO) seismic arrays of different sizes (Fig. 2): the ARCES array in Finnmark, northern Norway (3 km aperture, 25 stations, Schweitzer *et al.*, 2021),

Figure 2. Map of seismic events used for training ArrayNet. Events from the NORSAR reviewed bulletin (NRB) and Helsinki reviewed bulletin (HRB) with arrivals picked on the ARCES, FINES, and SPITS arrays. The symbol and legend colors indicate the bulletin. Identical arrivals are removed when merging the bulletins. The white star and the labeled region in the northeast indicate the origins of seismic events of special interest (see text). The color version of this figure is available only in the electronic edition.

the FINES array in southern Finland (2 km aperture, 16 stations), and the SPITS array on the Arctic Archipelago of Svalbard (1 km aperture, nine stations, Schweitzer *et al.*, 2021). The instrument types for all arrays are CMG-3T broadband sensors with 40 Hz (ARCES and FINES) or 80 Hz sampling (SPITS). Several agencies routinely monitor seismicity in the Nordic countries and the surrounding regions, including the European Arctic, and produce regularly updated regional event bulletins. Here, we use events reviewed and located by analysts at NORSAR, Norway, (NORSAR Reviewed Bulletin, NRB in the following, NORSAR, 1971a) and at the Institute of Seismology, University of Helsinki, Finland (Helsinki Reviewed Bulletin, HRB in the following, Veikkolainen *et al.*, 2021). Both the bulletins have overlapping content, and a few common stations are used for phase picking, including ARCES and FINES (Fig. 2). The NRB focuses on seismicity for onshore and offshore Norway, the Barents Sea, and the

TABLE 1
Number of Arrivals in Training Data Set (Unbalanced)

Array	P	Pg	Pn	S	Sg	Sn	Rg	PT	Frost
ARCES	21,239	11,212	22,669	20,640	23,561	15,296	5387	2932	–
FINES	24,868	10,566	8,655	20,907	19,514	7,172	–	–	–
SPITS	6,104	–	–	3,497	–	–	–	–	3,500

P and S are crustal arrivals without Pg, Sg, Pn, or Sn label. Rg and teleseismic P waves (PT) correspond to synthetic data. “Frost” stands for frost-quake observations at SPITS. SPITS was trained without Pn, Pg, Sn, and Sg labels because of limited training data.

northern Atlantic region around Svalbard, the latter region not being covered by the HRB. The HRB on the other hand includes more seismic events in eastern Scandinavia and Finland, as well as a multitude of blasts from different mines that are not included in the NRB. We merge both the bulletins between September 1998 and May 2022 by removing double arrivals (<2 s time difference required), resulting in about 55,000 P- and 59,500 S-wave arrivals at ARCES, 44,000 P and 47,600 S phases at FINES, and 6000 P and 3500 S arrivals at SPITS (only NRB). Table 1 presents a more detailed overview of regional phase labels included in the training data set. The regional or crustal phase labels (Storchak *et al.*, 2003) are not necessarily provided directly by the analyst. Typically, the analyst picks only the first P and S arrival, which can be Pn/Pg or Sn/Sg, or possibly only the second S arrival (Sg) for events farther away than the crossover distance, and the location program provides the labels corresponding with the best hypocenter solution. However, in the HRB we found also arrivals of regional events with labels P and Pb, the latter one being a wave refracted at a middle-crustal discontinuity. We treat both as P waves in which the classification into Pn and Pg classes is not available. This is not an issue for our method, because it was designed to first classify P versus S waves and then attempt a classification into regional phases if possible (see next section). The epicenter locations of all the events are used to compute the theoretical backazimuth with respect to each array, which is used later as label for the regression.

Continuous data from all the three arrays mentioned earlier, and other regional arrays are routinely analyzed by NORSAR’s automatic array processing pipeline (Schweitzer *et al.*, 2012). Step 1 in Figure 1 produces phase detections from array beams (detection processing: DPX), which are subsequently in step 2 processed by f-k analysis to measure backazimuth and apparent velocity, and the arrival type is labeled (signal attribute processing or f-k analysis processing: FKX). Association and location of regional events (step 3) is done using the Generalized Beamforming method (GBF, Kværna *et al.*, 1999) and the HYPOSAT location program (Schweitzer, 2001a). Empirical slowness corrections have been previously found for ARCES and FINES as tabulated values (Schweitzer, 2001b), and are taken into account by the GBF. We use the FKX output, the empirical slowness corrections, and the GBF catalog for evaluating our new method.

The SPITS array exhibits particular issues when it comes to array processing. The challenges are twofold. First, there is a high number of local seismic events inside or very close to the array, making automatic detection and event association challenging, that is, prone to false associations. Most of these signals can be attributed to local mining events and cryoseisms (frost quakes) during sudden ground-freezing periods (Romeyn *et al.*, 2022). Second, local geology below the array featuring dipping sedimentary layers affects seismic arrivals, leading to back-azimuthal bias and high apparent velocities, that is, waves arriving with a steep angle (Schweitzer, 2001b). Using EMFP at SPITS, Köhler *et al.* (2022) showed that this effect can be compensated for when detecting repeating events from a localized, regional source. However, for general event detection, no empirical slowness corrections are available yet. Therefore, the phase classification is ambiguous, and direct crustal phases are labeled “SPG” by FKX in a certain apparent velocity range to keep the option open for the event associator to use the observation as a P or S wave.

To summarize, we have the ground-truth data of analyst-picked seismic arrivals associated with located events that provide us with the phase type and backazimuth from the array to the event epicenter. In addition, we have the output of the existing automatic event processing pipeline (FKX and GBF) based on the plane-wave assumption. We aim to reproduce the ground-truth data (training data), and then evaluate any improvements with respect to FKX and GBF.

METHOD

Our new approach is motivated by results of the EMFP method. It has been shown that the theoretical coarray phase shift expected for a plane wave often cannot explain the time differences of observed arrivals on arrays (Kværna, Harris, *et al.*, 2023). Arrivals from a given direction, with a particular slowness and frequency, exhibit a characteristic footprint of time shifts on the coarray. The discrepancies with respect to plane-wave time shifts become larger at higher frequencies, because waves get more affected by small-scale inhomogeneities. These effects are mainly due to local site conditions (geology, presence of scatterers, and change of rock types) and lead to plane-wave distortions dependent on backazimuth, incident angle, and frequency (Schweitzer, 2001b; Gibbons *et al.*, 2010, 2011). In addition, decorrelation at larger interstation distances

can lead to characteristic effects on phase shift measurements. Furthermore, larger scale subsurface structures below the array such as dipping layers affect the wave propagation (Niazi, 1966; Berteussen, 1976). Although such structures may not affect the plane-wave character, the apparent velocity and backazimuth expected from a given source location and a standard layered Earth model are distorted. Systematic back-azimuth deviations are also caused by large-scale structures between the array and the event location.

Our goal is to let the coarray phase shifts measured in narrow frequency bands speak for themselves to determine the seismic phase type and the source direction without assuming a particular wave propagation model. Hence, we deal with a combined classification and regression problem with high-dimensional input data, that is, the number of coarray elements multiplied by the number of frequency bands. Artificial neural networks (ANNs) are well suited for such a task. The ANN, which we call ArrayNet in the following, is trained in a supervised manner, meaning that training data, that is, the coarray phase shifts, are labeled with the phase type picked by the analyst and the backazimuth pointing from the array towards the reviewed event location. An ANN with large input dimension requires a large amount of training data that we have available using the almost 25 yr long bulletins of seismic events from northern Europe and the European Arctic.

Coarray phase pattern

We measure arrival-time differences on the coarray, that is, time differences of the phase arrival for each array station pair as phase shifts in the frequency domain using the cross spectral or covariance matrix $\mathbf{R}(\omega)$ for frequency ω :

$$\mathbf{R}(\omega) = E\{\mathbf{Y}(\omega)\mathbf{Y}(\omega)^H\}, \quad (1)$$

in which the vector $\mathbf{Y}(\omega)$ represents the complex spectral values of all array channels, and H denotes the Hermitian transpose. The expectation operator E indicates the estimation of the spectrum from a short-time window including the phase arrival. The time window starts 0.25 s before the picked arrival time and has a duration of 3.25 s for all arrays. This was found to be optimal for our data set, but can be adapted if needed. For spectral estimation we use the multitaper method of Prieto *et al.* (2009) and Prieto (2022), which provides more stable results for the short-time windows and narrow frequency bands used here (Gibbons *et al.*, 2017). We tested different ArrayNet models using phase shifts measured at different frequencies. From the 127 frequency sample points of the multiple-taper spectrum, we finally select four frequencies: 2.5, 4.0, 5.5, and 7.0 Hz, corresponding to the frequency band typically used for the analysis of regional seismic events. The selection of time window length and frequencies was based on model performance parameters (see metrics introduced later). The performance of ArrayNet did not change significantly

when including more frequencies. However, the choice of these model parameters may have to be optimized for each array. Furthermore, in case of the window length, it should be checked that it is not too short to cut the arrivals, and not too long to include too much noise and coda.

Figure 3 shows examples of seismic event waveforms, seismic arrivals, and their measured phase shifts visualized as color-coded symbols representing the coarray. The position of each coarray element corresponds to the interstation distance of the corresponding station pair and its orientation. Such plots are provided for two frequency bands for the P -wave arrival (2.5 and 5.5 Hz) and one for the S -wave arrival (2.5 Hz). We show examples for an event north of the Swedish mine at Kiruna observed on ARCES (Fig. 3a), for an earthquake in the Gulf of Bothenia on FINES (Fig. 3b), an event of the Storfjorden aftershock sequence, Svalbard, on SPITS (Fig. 3c), and a frost quake on SPITS (note, different time scale in Fig. 3d). The P -wave phase patterns in Figure 3a–c show a clear directionality pointing toward the event origin. There is a tendency that higher frequencies show a more complex pattern with lower coherency for greater distances. On the larger aperture arrays (ARCES and FINES) the S -wave loses coherency at larger interstation distances. In contrast, the frost quake does not show any directionality in the phase pattern, which is expected if the signal originates from within or close to the array. Forced plane-wave fitting would result in wrong velocity and back-azimuth estimates. The constant time window used to compute the cross spectrum includes the entire frost quake signal. Hence, even though this event is also composed of a weak P arrival followed by secondary arrivals, we treat it here as a single arrival—the phase pattern being probably dominated by the surface waves.

Preparation of training data set

Figure 4 shows that certain directions dominate and others are underrepresented in the event bulletin, especially in the case of arrivals picked on ARCES. The dominating back-azimuth bins are mainly due to almost daily mine blasts in northern Scandinavia, northern Finland, and on the Kola Peninsula. As with any supervised method, ArrayNet can only classify based on the range of data observed during training—in our case, phase types and backazimuths. Although we experimented with weighting the input data to account for bias in the observed data, the results were unsatisfactory. Instead, we subsample dominating directions and augment the training data with plane-wave phase shifts for unrepresented backazimuths. For such augmented data the results of ArrayNet will be equivalent to an f - k analysis. To generate these data, theoretical phase shifts are computed for about 18,000 randomly distributed points in the slowness space (maximum slowness 0.5 s/km) for a given array. In practice, to achieve a more balanced training data set, we divide the event bulletin into back-azimuth bins of 4°, randomly down-sample dominant bins corresponding to very active mines, and add synthetic data in bins

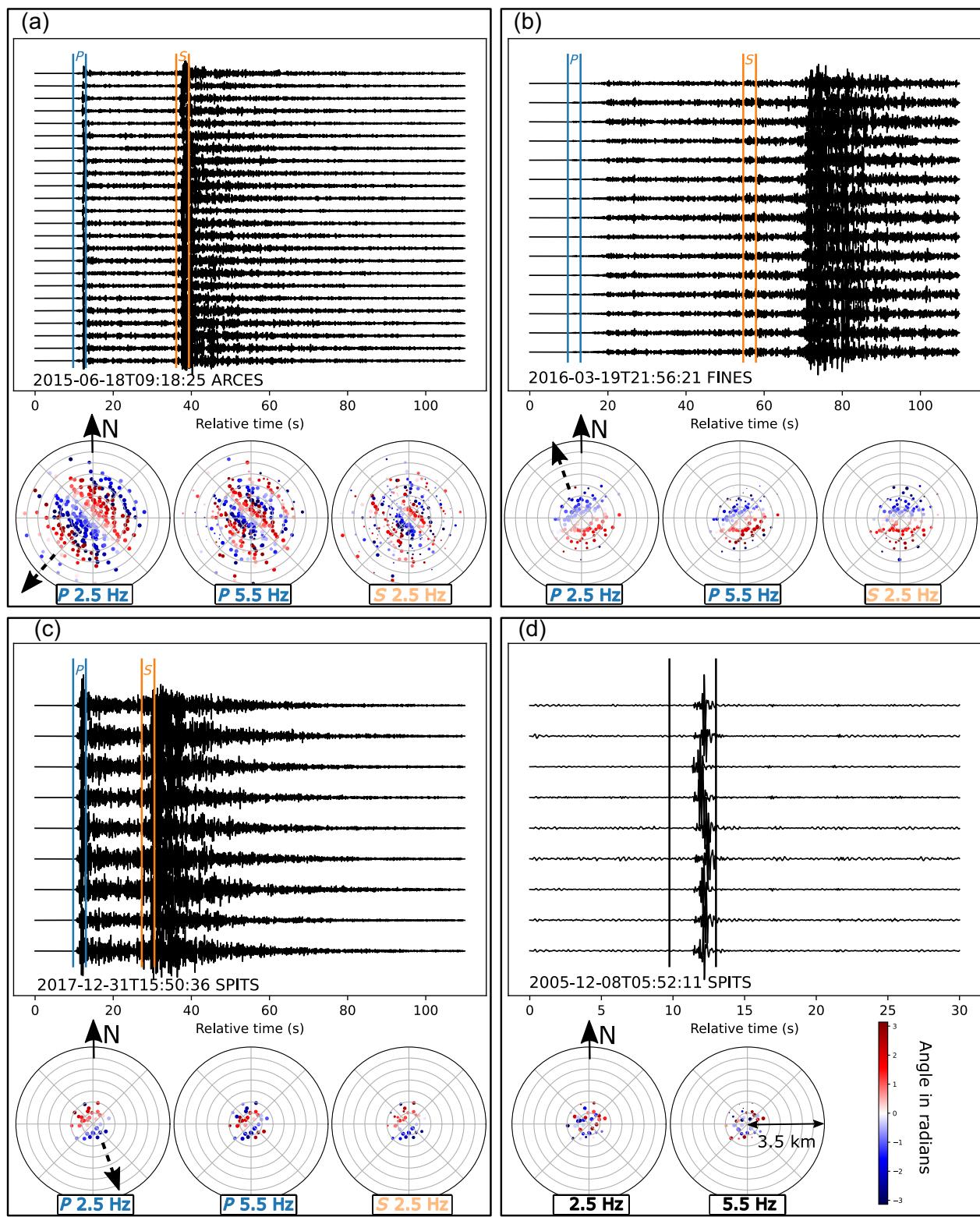
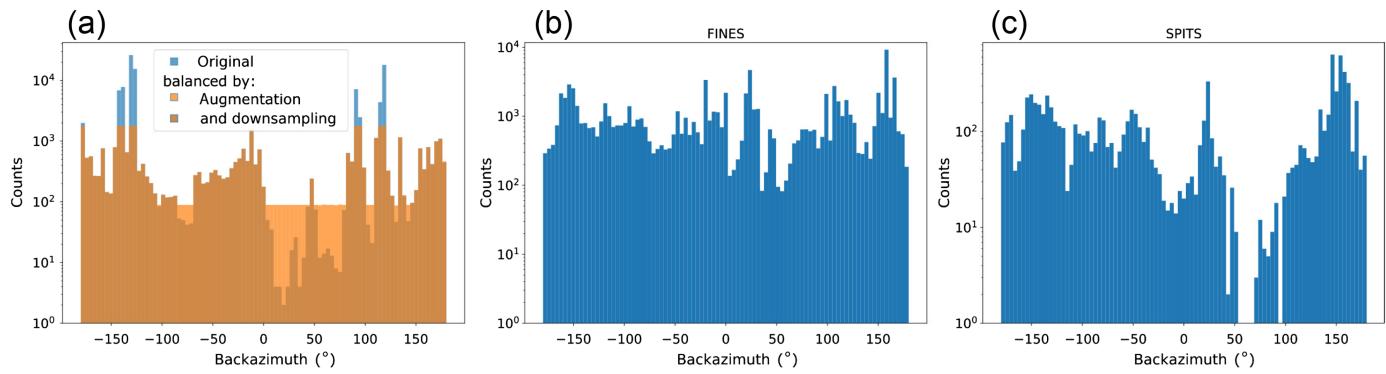


Figure 3. Examples of seismic arrivals in the training data sets. Vertical component waveforms on all array stations filtered between 2.5 and 8 Hz are shown. Time windows for computing the phase shifts of each arrival are shown as vertical lines. Subsequently, the coarray phase pattern for the indicated arrivals are presented for selected frequencies. Each symbol in the polar plot represents a coarray element. Its position corresponds to the station pair orientation and interstation distance. The symbol color indicates the phase shift value and the size to the coherency. North direction (N) and

backazimuth (dashed arrow) are shown. Note that the color scale and the radius are the same for all phase pattern plots and are given only once in panel (d). (a) An event in northern Sweden recorded on ARCES with P_n and S_n arrivals indicated $M_L = 3.8$. (b) An earthquake in the Gulf of Bothenia on FINES with P_n and S_n arrivals indicated $M_L = 4.7$. (c) An earthquake in the Storfjorden area, Svalbard, on SPITS with P and S arrivals indicated $M_L = 4.7$. (d) A frost quake very close to the SPITS array. The color version of this figure is available only in the electronic edition.



with fewer than 160 counts (see colored histograms in Fig. 4a). The latter number is simply the counts of available plane-wave data in each back-azimuth bin. The downsampling is controlled by a factor that is one if the number of real data to be kept in each bin should be equal to the number of data in each augmented bin (here 160 counts). The higher the factor, the more real data are kept, and the more unbalanced the data are. We found a balance factor of 20 to be optimal for ARCES. This means there can be up to 20 times more real data samples in each bin than the minimum number required given by the number of synthetic samples. For FINES no balancing or augmentation is required due to well-distributed seismicity around the array (Fig. 4b). In the case of SPITS directions between 50° and 70° are missing, but the total number of training data is also much lower than for ARCES and FINES (Fig. 4c). In addition, SPITS exhibits very strong slowness anomalies (Schweitzer, 2001b), and augmentation with synthetic data could therefore be inappropriate. In fact, we found in contrast to ARCES that when using an augmented SPITS data set ArrayNet performed worse, most likely because of these two reasons. Hence, the original SPITS data set is used in the following.

The training data set is labeled using the arrival type provided in the bulletin, which can be any regional phases (Pn , Pg , Sn , and Sg), or just a P or S phase label. The backazimuth is computed from the event epicenter location the pick is associated with. In the case of the synthetic data, the phase label is assigned based on typical apparent velocity ranges for seismic arrivals on the corresponding array, which are also used in FKX. Although we focus in this study on regional seismic data, we include also theoretical phase shift data (coarray phase patterns for plane waves) from teleseismic P and PKP phases as well as Rg phases to provide ArrayNet with categories not represented by the training data, which it will encounter during continuous data processing. Finally, we also need a class representing a nonarrival, that is, a noise class. Phase shifts calculated for a time window 10 s before each P arrival in the training data set are used as input data for this class.

We include an option to select training data above a coherency threshold to test the model performance with many or a few noisy arrivals (not done for the noise class). The coherency of an arrival is computed by averaging the coherency values of all phase shift measurements across the coarray and for all

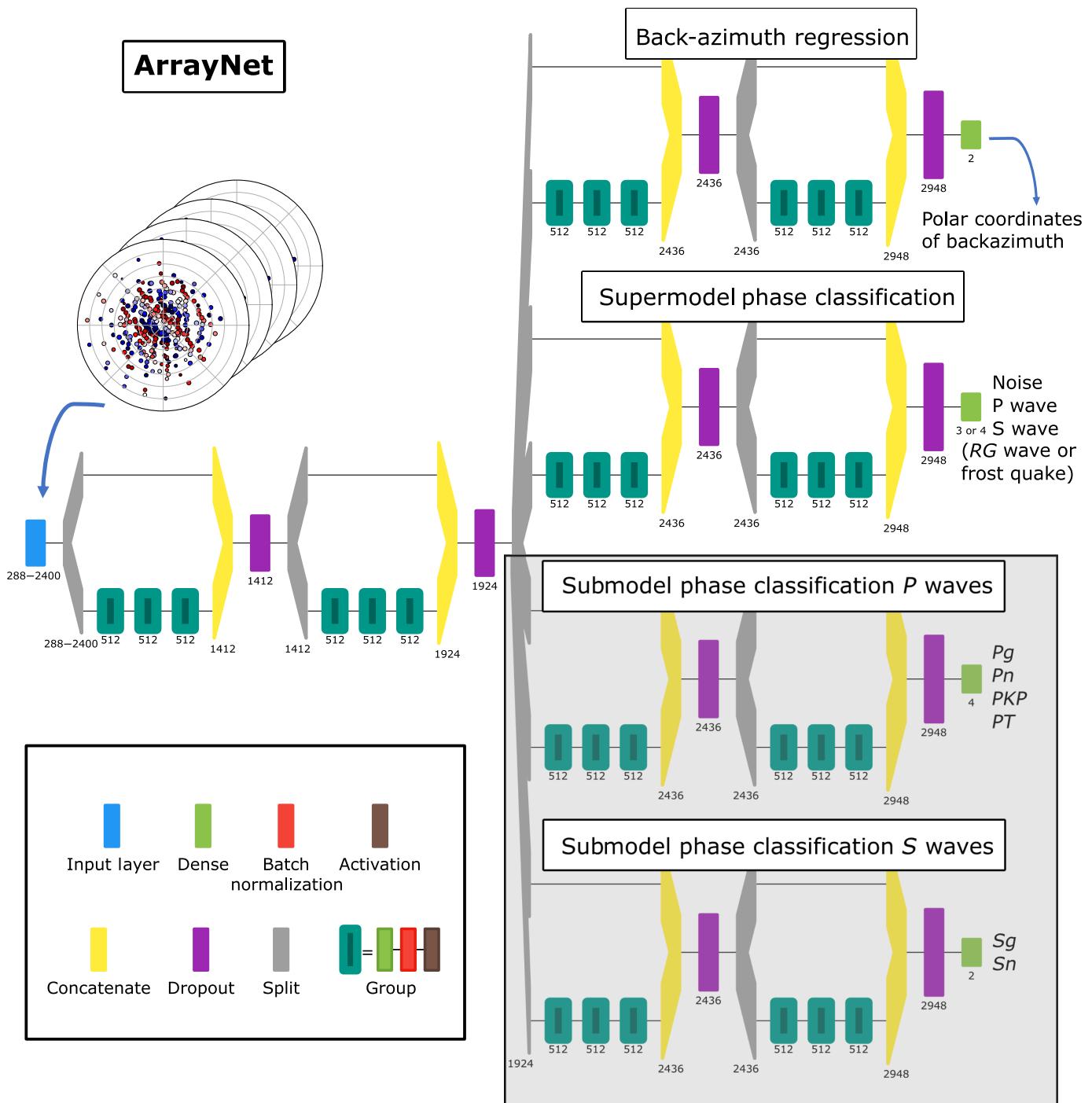
Figure 4. Back-azimuth distribution of seismic arrivals in the training data set for (a) ARCES, (b) FINES, and (c) SPITS array. Downsampling and augmentation with theoretical plane-wave phase shifts for underrepresented directions ensures a balanced data set for ARCES. Note, logarithmic scale for counts. The color version of this figure is available only in the electronic edition.

frequency bands. Finally, the (augmented) training data set is split into training (80%) and test sets (20%). The test set is reserved for final evaluation.

Neural network

The neural network used for the presented task is implemented in Keras (Chollet, 2015) and is based on a series of dense residual blocks (Fig. 5). These blocks are built up from subblocks before dropout is applied for regularization. Each subblock consists of a concatenation of all the previous layer outputs—a dense layer, batch normalization, and nonlinear activation. The 1D input layer has a dimension equal to two times the number coarray elements multiplied by the number of frequencies, that is, 2400 for ARCES, 960 for FINES, and 288 for SPITS. The phase differences for each coarray element are presented as complex numbers with absolute values equal to one. The factor of “two” originates from the fact that real and imaginary parts are simply concatenated as input nodes. If one (SPITS and FINES) or two stations (ARCES) of an array are missing, we simply set the input nodes to zero for the corresponding coarray elements. Data with more missing stations are skipped. After the first series of residual blocks, we split the output and pass it to two branches of a series of blocks: one for back-azimuth estimation and one for class prediction (supermodel). The output of the first branch is the backazimuth (complex decomposition, i.e., polar coordinates), whereas the output of the second branch is the probability for each class.

We expand this model to predict so-called subclasses for P and S arrivals (see the gray-shaded area in Fig. 5). These branches are identical for the superklass prediction, however, with a different number of outputs depending on the number subclasses. The submodel is trained simultaneously with the supermodel. The only difference is that weights of the submodel are only optimized if the training sample is a P or S wave in



the supermodel, and subclasses are available, that is, the corresponding loss is weighted by 1 if subclass exists and 0 if not. In practice, this means that, for example, P and Pn are not independent classes; but a Pn can only be predicted by the submodel if the arrival was already classified as P by the supermodel. We accept submodel results only if the probability of the winning class is larger than 70%. This threshold was found to be optimal for demonstrating the performance of ArrayNet but can be adapted according to the confidence level the user wants to achieve. Doing so, we can take into account that Pn and Pg are not always distinguishable, for example,

Figure 5. Architecture of ArrayNet. Input data are coarray phase patterns (dimension depends on array) from which the neural network learns through a series of blocks and subblocks of layers to estimate the back-azimuth (regression) and phase label (classification). The number of classes depends on the target arrival types. The model splits into four branches where the gray-shaded submodels are not implemented for all models presented in this study. The numbers below the layers in each subblock indicate the dimension, that is, the number of nodes. The color version of this figure is available only in the electronic edition.

TABLE 2

Results of Phase Classification for Test Data Sets (Models for ARCES, FINES, and SPITS)

Metric/Class	ARCES	FINES	SPITS	SPITS No Frost
Average accuracy	0.977	0.976	0.924	0.945
Accuracy frost	—	—	0.889	—
Accuracy noise	0.966	0.985	0.932	0.953
Accuracy <i>P</i>	0.984	0.968	0.962	0.980
Accuracy <i>S</i>	0.980	0.973	0.868	0.894
Accuracy <i>P</i> and <i>S</i> for FKX SPG	—	—	0.854	0.936
F1 score frost	—	—	0.933	—
F1 score noise	0.971	0.973	0.914	0.946
F1 score <i>P</i>	0.980	0.976	0.941	0.956
F1 score <i>S</i>	0.980	0.979	0.896	0.920
Accuracy FKX (<i>P</i> and <i>S</i>)	0.979	0.980	—	—
Accuracy FKX (<i>P</i>)	—	—	0.990	0.994
Accuracy FKX (<i>S</i>)	—	—	0.834	0.811

Accuracy and F1 score metrics for different classes are presented. For SPITS, metrics for two models are shown: one including and one excluding frost quakes (no frost).

around the crossover distance. In this case, the arrival will be labelled just as a *P* wave. To distinguish these arrivals from a teleseismic *P* wave, we could choose to label these phases as *Pgn*, which is the approach implemented in FKX. However, using the submodels architecture it is more convenient to introduce another subclass for teleseismic *P* waves, which we here call PT. For the *S* arrival submodel, only *Sg* and *Sn* phases are included.

The hyperparameters for the models (number of layers, number of units, dropouts, and activation functions) are based on reasonable choices. We use 512 units in each dense layer, 0.2 dropout, and rectified linear unit (ReLU) activation. We weight the loss functions for regression and classification equally during optimization, that is,

$$L = \text{MSE}(b, \hat{b}) + \text{CCE}(y, \hat{y}), \quad (2)$$

in which L is the combined loss function; MSE is the mean square error (regression); CCE is the categorical cross entropy losses (classification); b and y are the backazimuth and phase classification labels; and \hat{b} and \hat{y} denote their respective predictions. All models are trained using the Adam optimizer (Kingma and Ba, 2014) with a learning rate of 10^{-4} on a single Nvidia A30 graphic processing units (GPU), which takes a couple of hours depending on training data size and input data dimension. We use fivefold cross validation, that is, the training data set is split into five equal folds (subsets): four used for training (80%) and one for validation (20%). The training is repeated five times so that each fold serves as validation data once. This yields five slightly different models due to initialization, and the final prediction is averaged across these. The training is done for the maximum number of 200 epochs with an early stop if there is no improvement of loss after 15 epochs on the validation data. The learning rate is decreased by 50% if there is no decrease in validation loss after seven epochs. An example of a training curve, that is, decrease of training and

validation loss versus epoch, is provided in Figure S1, available in the supplemental material to this article, showing that the minimum loss is obtained after five to seven epochs.

We train multiple models for different seismic arrays using data sets with different balance factors and with or without using a coherency threshold. Furthermore, models are trained with only the supermodel branch (*P*, *S*, and noise), and using both supermodel and submodel branches for regional phase label prediction.

APPLICATION TO TEST DATA FROM REVIEWED BULLETINS

As a first step, we use only the supermodel branch for phase classification to test ArrayNet for ARCES, FINES, and SPITS. A coherency threshold of 0.3 was used for the training data selection. We evaluate the trained models with the test data not used in the training. In addition, we compare the results of ArrayNet models with phase labels and $f\text{-}k$ back-azimuth estimates obtained from the existing automatic processing pipeline (NORSAR FKX). Furthermore, the available empirical slowness vector corrections for ARCES and FINES are compared to ArrayNet's ability to correct back-azimuth bias and phase classification implicitly. In Table 2 and Figure 6 we present accuracy (ratio of the number of correct predictions to the total number of input samples), F1 score (the harmonic mean of precision and recall), and confusion matrices to evaluate the phase classification (see Berrar, 2019, for details about metrics). Back-azimuth regression is assessed by analyzing the distribution of back-azimuth residuals, that is, root mean square (rms) and median values with respect to true backazimuth.

In general, the performance of seismic phase-type classification is very good, with accuracy and F1 metrics between 0.97 and 0.98 for the ARCES and FINES models. These values are not significantly different compared to the FKX results, demonstrating that ArrayNet holds up to the well-established

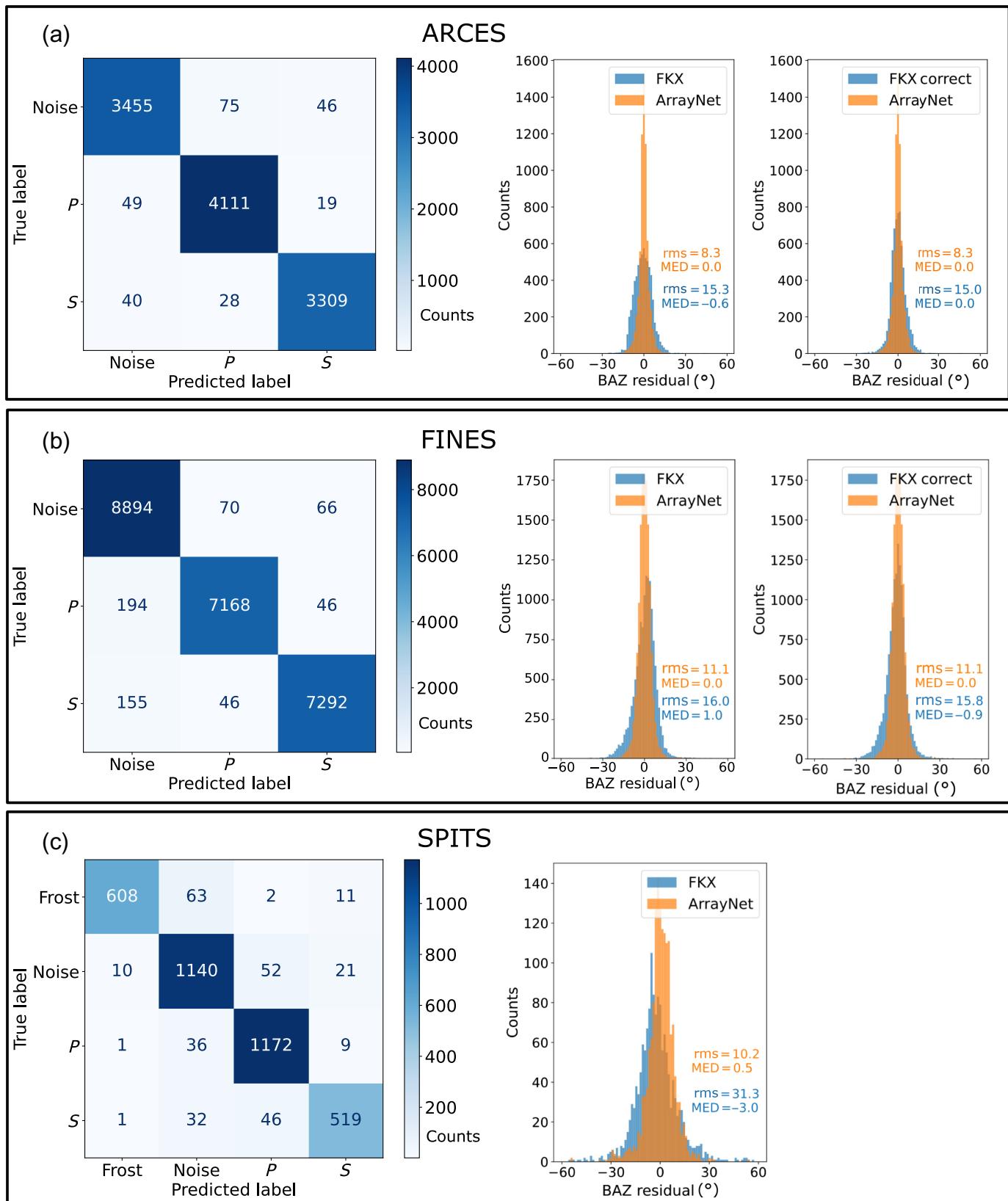


Figure 6. Results of phase classification and back-azimuth regression for test data sets. Models for (a) ARCES, (b) FINES, and (c) SPITS are shown. Confusion matrices for classification and root mean square (rms) and median (MED) for the distribution of back-azimuth residuals are presented.

The rightmost histograms show FFKX results taking into account empirical slowness corrections. (c) For SPITS no slowness corrections exist. The color version of this figure is available only in the electronic edition.

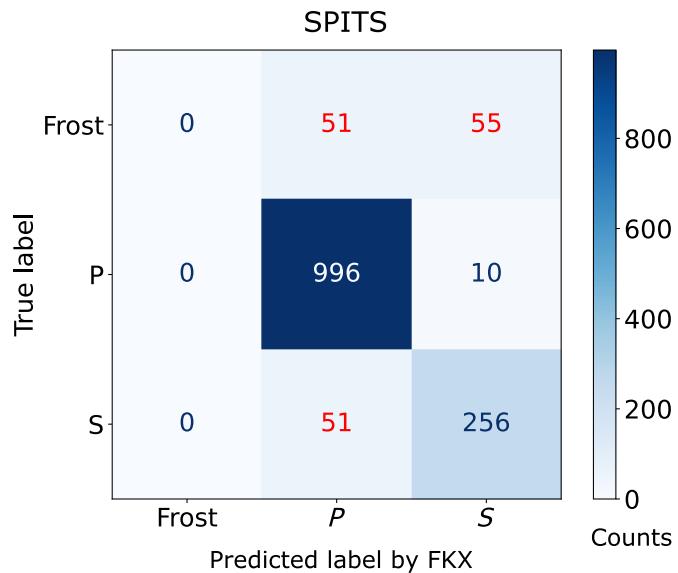


Figure 7. Confusion matrix for FNX at SPITS using the test data set. No frost quake class is included in FNX, and corresponding matrix elements are therefore zero. Highlighted in red are frost quakes falsely predicted as P or S arrivals, and S waves labeled P arrivals due to unusually high apparent velocities at SPITS. The color version of this figure is available only in the electronic edition.

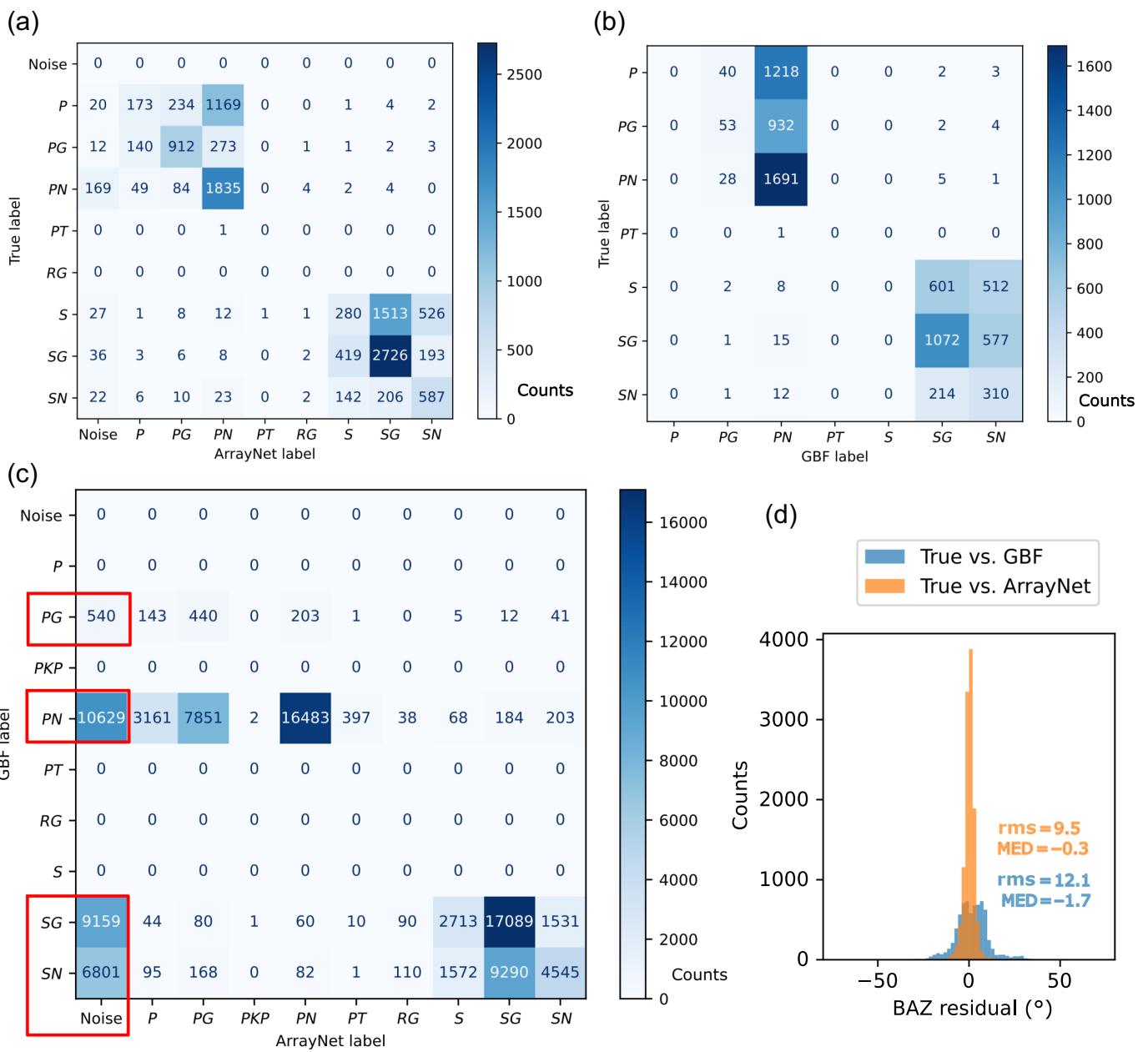
processing method (Fig. 6a,b and Table 2). On a closer look, ArrayNet using ARCES data performs slightly better than FNX when considering only P and S arrivals, while the FINES model produces values slightly below the FNX metrics. For reference, we also provide performance metrics for different back-azimuth ranges in case of ARCES (Table S1). At SPITS the accuracy and F1 score for each class show a bit more variability. We first only consider classification without using frost quakes. Although P-phase classification at SPITS has a high score for both ArrayNet and FNX (0.98–0.99), S-phase classification shows a lower accuracy of 0.89 for ArrayNet and 0.81 for FNX. This can be explained by the unusually high velocities for S waves at SPITS. As also shown in the confusion matrices (Figs. 6c and 7), S-wave arrivals have the tendency to be classified as P waves. However, ArrayNet does a much better job than FNX by increasing accuracy from 0.81 to 0.89, in the case of S-phase classification (Table 2).

ArrayNet achieves a clear improvement in the back-azimuth regression, that is, a significant decrease of residuals for all arrays (see rms values in Fig. 6). Moreover, the back-azimuth bias is reduced most prominently for FINES and SPITS (see median values in Fig. 6). When correcting FNX for ARCES and FINES for slowness deviations directly, residuals, and bias are reduced as well, however, not as much as achieved by ArrayNet. In summary, we find that ArrayNet produces comparable phase labeling results to FNX for all arrays, with SPITS showing slight improvements for the identification of S waves. For back-azimuth estimation, ArrayNet outperforms FNX for all arrays.

As mentioned earlier, SPITS FNX labels arrivals as SPG phases in ambiguous apparent velocity ranges. We did not include a SPG class into ArrayNet, since analyst-reviewed phase labels are used for training in which the context of the arrival clearly suggests if it is a P or S wave. The accuracy of ArrayNet for P and S arrivals labeled as SPG by FNX is 0.94 (without the frost quake class, see Fig. 6c), demonstrating that a precise discrimination using the coarray phase pattern is in fact possible for most of these arrivals. Although this achievement can contribute to improved phase association at SPITS, the presence of signals originating from inside or close to SPITS is another factor to be considered. Therefore, we added another class representing frost quakes to the supermodel (Fig. 5). Figure 6c shows that the discrimination between frost quakes, noise and P and S arrivals works well; although the accuracy and F1 score decrease by several percent compared to the three-class model. However, the confusion matrix shows that there are only a few P and S waves classified as frost quakes and vice versa. Furthermore, ArrayNet also reduces the number of frost quake detections previously classified as P or S waves. Figure 7 shows that FNX had assigned a P or S label for 106 frost quake signals in the test data set. In comparison, ArrayNet only misclassified 13 of these signals. Although ArrayNet confuses frost quakes and noise to some extent (Fig. 6c), this is not problematic when the main interest is the correct identification of P and S waves. Nevertheless, the inclusion of the frost quake class deteriorates the P- and S-waves accuracy, including the classification of arrivals labeled as SPG by FNX. This could be a result of the smaller training data set for SPITS compared to ARCES and FINES (see Table 1), because a multiple class model is expected to require more training data for good performance. However, overall, ArrayNet has several benefits compared to FNX at SPITS. This includes fewer frost quakes affecting event association, no need for a SPG label, and significantly reduced back-azimuth discrepancy.

APPLICATION TO CONTINUOUS ARRAY DETECTIONS AT ARCES

In the next step we train ArrayNet using both the supermodel and submodel by adding more classes from regional seismic arrivals at ARCES (see gray-shaded area in Fig. 5). The supermodel includes the noise class, P and S arrivals from the reviewed event bulletins, and Rg phases, the latter trained from synthetic coarray phase patterns using a plane-wave model. The submodel includes two classes for S waves (Sg and Sn) and four subclasses for P waves, that is, observed Pg/Pn waves and synthetic (plane wave) PKP or other teleseismic P arrivals (PT). We use the complete training data set without first applying the coherency threshold. We then apply the trained model to 20 months of DPX detections at ARCES between 7 May 2021 and 31 December 2022. This not only includes some of the reviewed arrivals seen during training but also new analyst-labeled data from 2022 that were not available when the model was trained.



The results are compared with the manually picked arrivals, and those Fkx arrivals automatically associated to form seismic events by the GBF method. The motivation for the latter is to assess if ArrayNet is able to better identify falsely associated spurious GBF arrivals by classifying them as noise.

Figure 8a,b shows confusion matrices for the phase classification done by ArrayNet and GBF. Figure 8c presents a comparison between GBF and ArrayNet labels. Figure 8d shows again that back-azimuth estimation is more precise with ArrayNet compared to GBF. Both ArrayNet and GBF perform well when discriminating P and S waves. GBF has obviously no noise class, because all arrivals are associated with potential events. Furthermore, all GBF phases are either Pg or Pn, or Sg or Sn, because the location program requires one of these labels in order to compute travel times for the first arrival. Pg

Figure 8. Results of applying ArrayNet to 20 months of DPX detections at ARCES. (a–c) Confusion matrices for comparing ArrayNet, GBF, and the true labels in the reviewed bulletin provided by the analyst. The red rectangles in panel (c) indicate noise classifications that GBF associated as P or S waves. (d) Back-azimuth residuals. The color version of this figure is available only in the electronic edition.

arrivals are relabeled as Pn or Pg depending on the epicenter distance. Furthermore, no Rg waves are associated. ArrayNet can provide a P or S label if the class probability for all sub-phases (e.g., Pg vs Pn) is lower than 70%. Many arrivals labeled as P or Pb in the HRB are in fact identified as Pn by ArrayNet (Fig. 8a). This can be explained by the fact that ArrayNet does not have a Pb class; and that Pb slowness, and thus the coarray

TABLE 3
Seismic Events Considered Critical for Detection

Origin Time (UTC) (yyyy/mm/dd hh:mm:ss)	M_L	BAZ (°)	FKX Phases	FKX BAZ (°)	ArrayNet Phases	ArrayNet BAZ (°)
2002/02/23 01:21:18	2.97	50.8	Pg/S	60.3	Pn/Sn	65.7
2002/11/10 11:04:47	2.0	73.0	Pn/S	78.4	Pn/S	78.8
2006/03/05 23:17:36	2.65	39.4	Pgn/Sg	47.0	Pn/Sg	36.8
2006/03/14 20:57:03	2.23	43.9	Pgn/Sn	57.2	Pn/Sn	11.1
2006/03/30 10:46:03	2.3	71.0	Pg/Sn/Sg	75.7	Pn/Sn/Sg	80.3
2007/06/26 03:19:20	2.76	53.7	Pgn/Sg	59.8	Pn/Sn	59.2
2009/11/11 04:18:21	3.17	65.4	Pn/Sg	73.4	Pn/Sn/Sg	69.0
2010/10/11 22:48:31	4.5	42.0	Pgn/Sg	46.2	Pn/Sn	47.1
2020/08/23 03:26:22	3.14	52.3	Pn/Sg	63.1	Pn/Sn/Sg	59.2
2022/09/26 17:03:50	2.7	201	Pn	197	Pn/Pn	192/208

Magnitude (M_L) and theoretical backazimuth (BAZ) are given. Upper rows are events in the Novaya Zemlya Region. The bottom row is the Nord Stream pipeline event.

phase pattern, is more similar to Pn than to Pg . Those arrivals classified as P waves most likely belong to events around the crossover distance, because Pn and Pg arrive almost simultaneously; and, thus, submodel probabilities are expected to be low and no label is provided. When comparing the confusion matrices we see that Pg and Pn discrimination of ArrayNet is not perfect, but much more precise than the GBF (Fig. 8a,b). Similar results are obtained for S phases: Sg is the most frequent label given by ArrayNet, many of those labeled as S in the bulletin (Fig. 8a). This makes sense, because Sn is often not recognizable as a first arrival, and consequently Sg is picked by the analyst. There is less confusion between Sg and Sn for ArrayNet compared to the GBF (Fig. 8b).

The confusion matrix in Figure 8c is more difficult to evaluate, since we do not have ground truth labels for all GBF events. However, a few important observations can be made. The red rectangles indicate a large number of noise classifications of ArrayNet, which are event-associated P and S phases in the GBF. The obvious question is whether these are false associations or real events, which would be missed if the associated arrivals would be sorted out falsely by ArrayNet. When we check ARCES waveforms for a randomly sampled subset of those arrivals (300), we find that 88% do not correspond to genuine phases for events. The remaining 12% are very-low-signal-to-noise ratio (SNR) seismic events observed only on ARCES. Further optimization of ArrayNet would be required to recognize these phases if such events were of interest. However, the ability of ArrayNet to identify a significant number of false arrivals would greatly benefit the GBF and the subsequent event processing by the analyst. There is a single phase labeled as teleseismic by the analyst in Figure 8a,b, which is presumably from a far-regional event in which Pn was labeled as P . In Figure 8c, there are 397 detections classified as teleseismic P waves by ArrayNet that GBF (regional events only) associated as Pn . This is related to the transition between far-regional and teleseismic arrivals. Furthermore, there are 200 associated Sn and Sg waves classified as Rg waves. Because Rg labels are not included in FKX/GBF, this result

is difficult to evaluate quantitatively. However, similar as for the noise class, this could be an indication that ArrayNet allows removing spurious Rg waves from the GBF input.

APPLICATION TO EVENTS OF INTEREST PREVIOUSLY DETECTED AT ARCES

Because there is the possibility that low-SNR-event arrivals might be classified as noise by ArrayNet, we investigate a set of events of interest that we consider critical for detection. Such events include those from the Novaya Zemlya region indicated in Figure 2 (Gibbons *et al.*, 2011)—the location of a former Soviet nuclear weapon test site and the recent seismic events from the Nord Stream gas pipeline explosions. Table 3 presents nine earthquakes that occurred in the Novaya Zemlya region during the past 21 yr. P - and S -wave arrivals were detected by DPX and FKX for each of these events. Applying ArrayNet to these data, we are able to correctly classify all arrivals. Furthermore, the back-azimuth estimates are in the range of about $\pm 15^\circ$ around the true locations. The FKX back-azimuth estimates vary to the same extent. This is expected given that those event locations are somewhat uncertain, since they were only located using the ARCES and SPITS array, the SNRs are in general low, and the regional velocity model is more uncertain (Kværna, Dando, *et al.*, 2023). An exception is the event on 14 March 2006 with the second smallest magnitude where ArrayNet is off by more than 30° . This is most likely due to low SNR but could also be a result of a mislocated event epicenter. There are not many observing stations in that region, and the velocity model is not well known.

On 26 September 2022, seismic events were detected on stations throughout northern Europe related to the explosions and subsequent leakages at the Nord Stream gas pipelines in the Baltic Sea (Köhler *et al.*, 2023; Lund *et al.*, 2023). The strongest event had an origin time at 17:03:50 UTC, and an associated Pn arrival was detected by DPX and FKX at ARCES at a distance of 1660 km at 17:07:13.5 UTC with a low SNR = 9 (Fig. 9, see location in Fig. 2). The GBF method used this phase for event association in addition to arrays closer to the epicenter (e.g.,

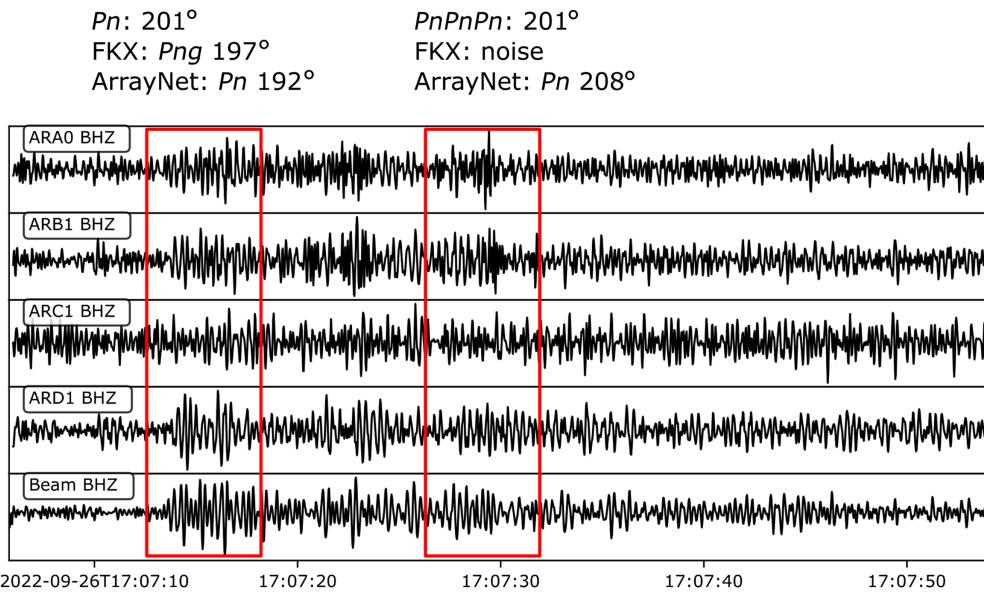


Figure 9. Pn phases of a seismic event related to the Nord Stream explosions. Vertical-component waveforms on four array stations of ARCES and the Pn beam (7.5 km/s) toward the event epicenter filtered between 2.5 and 8 Hz are shown. Pn arrivals are indicated in the red boxes. FNX and ArrayNet phase labels, and back-azimuth estimates are given. Last arrival is consistent with $PnPnPn$ phase but is more likely a first arrival of a subsequent event. The color version of this figure is available only in the electronic edition.

FINES). ArrayNet correctly classifies the first Pn arrival and estimates the backazimuth as 192°, which is 9° from the theoretical value (Table 3). In addition, DPX detected a subsequent phase consistent with theoretical arrival times and slowness of a $PnPnPn$ phase (Fig. 9). ArrayNet labels it correctly as Pn with a backazimuth of 208°, whereas FNX did not provide P -wave classification and backazimuth for this arrival. A possible $PnPn$ arrival is visible as well but was not detected by DPX due to low SNR. However, a recent study suggests that these three arrivals originate from multiple events (Köhler *et al.*, 2023; Lund *et al.*, 2023).

DISCUSSION

As with other applications in detection seismology, the motivation for developing a machine learning-based method for processing arrivals detected on a seismic array is to take advantage of existing long bulletins with labeled seismic events. The result is that we are able to effectively account for seismic wave propagation over an array that is too complex to be accurately represented by a plane-wave model. ArrayNet automatically discovers and generalizes empirical relationships between observed arrival times and the phase type and backazimuth without a plane-wave model in the background, which is usually assumed in array processing. We can think of ArrayNet as a generalized framework for the previously used empirical slowness corrections for arrays (Schweitzer, 2001b). The stronger the plane wave and/or slowness deviations are at a particular array, the more beneficial is our approach for subsequent seismic event

association compared to, for example, classic f - k -based processing. See, for example, the improvement of back-azimuth estimation at the SPITS array (Fig. 6c). We find that the phase classification of ArrayNet clearly matches the results of the well-established FNX at ARCES, FINES, and SPITS. Moreover, a consistent finding for all arrays is a clear improvement when it comes to back-azimuth estimation, including in comparison with using existing empirical slowness corrections. The results for the SPITS array need particular consideration. First, the training data set is significantly smaller than for ARCES and FINES, which slightly deteriorates the performance. This demonstrates the need for a long recording period for our

method. Second, strong slowness deviations and frequent events inside or very close to SPITS have caused issues for automatic event processing in the past. We show that ArrayNet can help to overcome these issues by identifying frost quake signals, better discriminating between P and S wave arrivals, as well as removing the strong back-azimuth bias.

In the case of arrays without significant slowness deviations and clearly defined velocity ranges for P - and S -wave arrivals, ArrayNet will not likely make significant improvements to back-azimuth estimates or the discrimination between P and S waves. However, the method can still be beneficial for identifying certain detections, for example, local noise bursts or correlated continuous ocean wave-generated noise, which should not be associated to any event, but which seem to have characteristic and distinct coarray phase patterns. We have shown that many such detections previously classified as P or S waves at ARCES are now categorized as noise. It may be worth mentioning in this context that also the F -detector has been shown to avoid the detection of signals in the ocean wave-generated background noise at ARCES (Selby, 2008).

ArrayNet can also contribute to speeding up processing. From our experience, a trained ArrayNet model runs up to ten times faster than FNX. For both ArrayNet and f - k analysis, the cross-spectral matrix of the array data has to first be estimated. However, ArrayNet does not require the time expensive grid search in the slowness space as performed by the f - k analysis. Instead, ArrayNet relies on matrix multiplications using the neural network weights for predicting phase type and

backazimuth. Running ArrayNet on a single central processing unit CPU (1.7 GHz) took 1.6 s for a single arrival, including the time of reading data from a local data base and computing the phase pattern.

It is also worth conceding that classic FkX remains more flexible in certain scenarios. ArrayNet requires a large training data set and must be trained for each array individually. However, for FkX to be effective, tuning of the parameters for individual arrays is also necessary and also requires labeled data. This is typically possible only after several months of data collection. Factoring in the need for applying empirical slowness corrections for certain arrays, which may take years of data collection to accurately derive, the comparable effort of training ArrayNet becomes greatly diminished. Furthermore, with the availability of reviewed event bulletins and the waveform data from seismic arrays now stretching back for decades, for example, from the IMS arrays, we have the necessary data to implement ArrayNet at many of the world's existing permanent seismic arrays. For the deployment of new arrays, one option is to use synthetic data as a preliminary training data set, as we did for data augmentation in this study. The initial model would then mimic *f-k* analysis and phase classification using preset slowness limits, however, still with the advantage of being faster than *f-k*. ArrayNet can then be regularly retrained when reviewed observations become available, adapting itself gradually to the characteristics of real arrivals. It is worth mentioning that the synthetic data we augment with here are pure plane-wave time differences without adding noise. This is intentional, because we want ArrayNet to provide the best plane-wave solution of arrivals for underrepresented or missing back-azimuth ranges or arrival classes. However, if a model should be trained from synthetic data alone, adding noise to the arrival times may be an option to be tested in the future work.

The impact of training data augmentation as we do for ARCES needs a few more considerations. If we train ArrayNet models without augmentation, we would like the output (back-azimuth and phase type) to be an interpolation constrained by the data surrounding the regions in the slowness space with missing or underrepresented data. Ideally, this will result in a smooth continuity; however, there is no guaranty that the neural network will not overfit the few available data points, which would result in undesirable behavior for unseen data in the corresponding slowness space region. In our case, events to the northeast of ARCES are rare (see Fig. 4). To test the effect of augmentation, we compare an ArrayNet model without augmentation (but still subsampling dominant directions) and with augmentation. The classification accuracy does not change significantly. However, Figure S2 shows that there is a difference for the back-azimuth prediction in the range of few training data. The unaugmented model tends to predict similar values in certain dominant direction ranges (around -5° and 47°). This is visible as a local "plateaus" in the figure.

Using the same data set for prediction with the augmented model does not result in these plateaus, although there is still some scatter of the predicted backazimuth compared to the ground truth. The most likely explanation is that the synthetic data helps ArrayNet to better interpolate at gaps and prevents dominant directions to bias the predictions. No artificial artifacts (jumps in backazimuth) are introduced. These results suggest that augmentation of back-azimuth gaps (in addition to subsampling and adding teleseismic or *Rg* arrivals) seems to be beneficial for ARCES or at least does not negatively affect the results. We anticipate that augmentation will become even more important for smaller training data sets with major data gaps as discussed earlier.

Another disadvantage of ArrayNet might be that it is trained for coarray phase patterns at fixed frequencies (here four bands between 2.5 and 7 Hz). Decreasing the lowermost frequency or increasing the uppermost requires training a new model since the type of input data changes, that is, the coarray phase patterns at fixed frequencies. However, training an ArrayNet model is a matter of hours only on GPUs, and one could for example train multiple models for different frequency bands. In the processing pipeline the model for the frequency band that has the highest SNR could then be applied, similar to adaptive band-pass filters in the existing FkX. A test of this approach is the subject of an upcoming study.

Data issues can affect the performance of ArrayNet similar as for any conventional array processing method. If one or two stations are missing, we simply set the input nodes to zero for the corresponding coarray elements. If more stations are missing, then the arrival is skipped. For arrays having more frequent station outages, one may relax the skipping criterion and allow for more zero input nodes. This has to be evaluated in the future studies. The case when a station provides faulty data is more difficult to handle. Because of the vast amount of training data, a short time period of bad data should not have a significant effect on the trained model. Because ARCES and FINES are primary IMS stations, and data quality is generally high, we do not see that the models are affected. However, not all arrays have such a high quality at all times. In fact, there have been longer time periods of faulty or missing sensors for the SPITS array in the past (timing errors). To better handle such data, one could for example adjust the coherency threshold for training data and prediction, however, at the expense of less available data. Figure S3 shows the coherency for SPITS coarray phase patterns over time plotted versus the number of operational sensors. Indeed, time periods of missing sensors and/or with timing issues can be identified before 2015. To investigate this issue, we trained a new SPITS model for a time period after May 2015 in which the array was operating without issues after an instrument upgrade. Furthermore, we compare the prediction results for data before and after 2015 separately. Table S2 shows that accuracy and F1 score values decrease for the new model due to less training data available.

However, the main point to emphasize here is that performance for predicting with data after May 2015 increases for both the models, most significantly for the new model. The results for back-azimuth estimation (Fig. S4) are similar for both the time periods with slight improvement of the rms value after 2015. Thus, we can conclude that the model we trained originally handles both the time periods very well with a slightly better performance for times periods with good data, whereas a model only trained with good data fails to handle the time period with more faulty data. Hence, as a recommendation, training data for ArrayNet should represent different states of data quality if an array is known to be prone to station failures.

There is obviously much more work to be done to evaluate ArrayNet, but this is beyond the scope of this article. More testing can be done for choosing the most optimal input data for different arrays, that is, frequencies and window length. We plan to systematically test how ArrayNet improves event association, in particular, if it will reduce the number of false associations. At the same time we have to ensure that no arrivals are missed. We have tested this for a few events of interest, but this needs a more thorough evaluation. Furthermore, the next steps are training models for more arrays including IMS stations and using training data from other regional or global reviewed bulletins, including the IDC's Reviewed Event Bulletin (REB) or Late Event Bulletin (LEB). Global bulletins would also require extending ArrayNet's phase classification to teleseismic arrivals, which is straight forward, as we have shown when extending the model to more regional phases in this study. Moreover, classification uncertainties are not considered so far, apart from requiring a 70% threshold for the submodel phase classification. Uncertainties can be incorporated in the output of ArrayNet, for example, by Monte Carlo dropout ([Gal and Ghahramani, 2016](#)).

It is worth noting that a machine learning approach could also be used to correct slowness vector directly. One would first compute the theoretical slowness vector for a wave type at a given distance and backazimuth using a global 1D velocity model. A neural network could then be trained to map the slowness vector measured by f - k analysis to the theoretical slowness vector. Such a model would not require the coarray patterns to be included as input data. However, this approach would not be able to handle sources close to or inside the array such as frost quakes. There would also be no noise class that can potentially be used to clean the input of the event associator, and f - k analysis is still needed, which is computationally slower.

Most of the deep learning models developed recently for seismic event detection and classification make use of raw seismic waveforms and convolutional neural networks (CNNs, [Zhu and Beroza, 2019](#); [Mousavi et al., 2020](#); [Köhler et al., 2022](#); [Mousavi and Beroza, 2023](#)). ArrayNet is different,

because it does not process a time series including information about the temporal context of the arrival. Hence, a standard neural network layout without convolutional layers is used. The coarray phase patterns are computed from the waveforms of already detected arrivals beforehand. On the other hand, methods like PhaseNet ([Zhu and Beroza, 2019](#)) or EQTransformer ([Mousavi et al., 2020](#)) detect and classify P and S waves using CNNs and raw waveforms; however, they do not take advantage of array data yet. We envision a combination of these methods for future array processing pipelines. PhaseNet- or EQTransformer-like models could replace STA/LTA-based phase detection on array beams or individual array channels (step 1), whereas ArrayNet could then in step 2 give a more refined phase classification and back-azimuth estimate. Another option is to add convolutional layers applied to all array channels to implicitly extract the coarray phase differences, or any other feature that the network finds to be useful, prior to the ArrayNet-dense layers. A disadvantage with such an approach is that synthetic phase pattern cannot be included so easily anymore. CNN-based classification could also be added after event association (step 3) to sort out false associations ([Köhler et al., 2022](#)). There are many ways to integrate machine learning into array processing, and we anticipate more work to be done in this field in the coming years.

CONCLUSIONS

We have developed a combined classification and regression neural network called ArrayNet, which labels seismic arrivals recorded on arrays and estimates their backazimuth. In contrast to classic array processing, no plane-wave model is assumed for wave propagation. ArrayNet uses coarray phase patterns and, thus, can take into account structural inhomogeneities below and farther away from the array as well as local topography affecting the waveform. We have shown that models trained with regional P - and S -wave arrivals in northern Europe and the European Arctic on the ARCES, FINES, and SPITS arrays perform with high accuracy for phase classification, comparable to existing array processing results. Moreover, the back-azimuth residuals are clearly reduced with respect to plane-wave array processing, including where empirical slowness corrections have been previously used. For the SPITS array in Svalbard, we have shown that ArrayNet can handle issues related to unusually high apparent velocities for incoming phases and the presence of events from inside the array. Furthermore, we evaluated an ArrayNet model for the ARCES array with arrivals detected over a duration of 20 months and found clear advantages compared to the existing processing. We also found that a selection of seismic events of special interest are correctly identified by ArrayNet.

Array processing is an integral part of automatic pipelines for seismic event detection, but little work has been done on integrating machine learning into this process so far. We demonstrated that using such tools has potential benefits

for the generation of automatic event bulletins. We consider ArrayNet in the present form as a first step in this direction, and we suggested further potential improvements and extension of our novel method. We envision that ArrayNet could be one of several machine or deep learning methods assisting or replacing modules in the existing processing schemes.

DATA AND RESOURCES

Waveform figures were generated, and seismic data processing was partly done using Obspy (Beyreuther *et al.*, 2010). Figure 2 was generated using the Generic Mapping Tools (Wessel and Smith, 1995). ARCES and SPITS waveform data are available via Incorporated Research Institutions for Seismology (IRIS) (Albuquerque Seismological Laboratory [ASL]/U.S. Geological Survey [USGS], 1988) or the Norwegian node of The European Integrated Data Archive (EIDA) (Ottemöller *et al.*, 2021). All data are stored at Norwegian Seismic Array (NORSAR) (NORSAR, 1971b). Reviewed seismic event bulletins are available from the Finish National Seismic Network (Institute of Seismology, 1980a,b; Veikkolainen *et al.*, 2021) and from NORSAR (NORSAR, 1971a). The code of ArrayNet is available at <https://github.com/NorwegianSeismicArray/arraynet> (last accessed August 2023). Supplemental Tables S1 and S2 show phase classification metrics for more ArrayNet models and data ranges. Supplemental Figures S1–S4 show training curve, details on back-azimuth prediction for the ARCES models, data quality at SPITS, and more details about back-azimuth prediction for SPITS models.

DECLARATION OF COMPETING INTERESTS

The authors acknowledge that there are no conflicts of interest recorded.

ACKNOWLEDGMENTS

The authors thank the Associate Editor Junghyun Park, two anonymous reviewers, and Ben Dando for their helpful comments. The authors thank Karoline Aastrup for spellchecking the article. Author contribution: Andreas Köhler developed the idea, trained and tested the models, and wrote the article. Erik B. Myklebust implemented the models and contributed to the article text.

REFERENCES

- Albuquerque Seismological Laboratory (ASL)/U.S. Geological Survey (USGS) (1988). Global seismograph network (GSN–IRIS/USGS), doi: [10.7914/sn/iu](https://doi.org/10.7914/sn/iu).
- Ben Horin, Y., K. Koch, and Y. Bartal (2004). Use of GSETT-3 gamma data in the slowness-azimuth calibration of IMS primary arrays at regional distances, *J. Seismol.* **8**, 129–142.
- Bergen, K. J., T. Chen, and Z. Li (2019). Preface to the focus section on machine learning in seismology, *Seismol. Res. Lett.* **90**, no. 2A, 477–480.
- Berrar, D. (2019). Performance measures for binary classification, in *Reference Module in Life Sciences, Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, M. Grabskov, K. Nakai, and C. Schönbach (Editors), Vol. 1, Academic Press, Oxford, England, 546–560, doi: [10.1016/B978-0-12-809633-8.20351-8](https://doi.org/10.1016/B978-0-12-809633-8.20351-8).
- Berteussen, K. A. (1976). The origin of slowness and azimuth anomalies at large arrays, *Bull. Seismol. Soc. Am.* **66**, 719–741.
- Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A Python toolbox for seismology, *Seismol. Res. Lett.* **81**, no. 3, 530–533.
- Cansi, Y. (1995). An automatic seismic event processing for detection and location: The PMCC method, *Geophys. Res. Lett.* **22**, no. 9, 1021–1024.
- Capon, J. (1969). High-resolution frequency-wavenumber spectrum analysis, *Proc. IEEE* **57**, no. 8, 1408–1418.
- Chollet, F. (2015). Keras, available at <https://github.com/fchollet/keras> (last accessed August 2023).
- Douglas, A. (2002). Seismometer arrays—Their use in earthquake and test ban seismology, in *International Handbook of Earthquake and Engineering Seismology*, W. H. K. Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger (Editors), Academic Press, London, England, 357–367.
- Gal, Y., and Z. Ghahramani (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning, *International Conf. on Machine Learning, PMLR*, New York, 19–24 June 2016, 1050–1059.
- Gibbons, S. J., D. B. Harris, T. Dahl-Jensen, T. Kværna, T. Larsen, B. Paulsen, and P. Voss (2017). Locating seismicity on the Arctic plate boundary using multiple-event techniques and empirical signal processing, *Geophys. J. Int.* **211**, no. 3, 1613–1627.
- Gibbons, S. J., T. Kværna, and F. Ringdal (2010). Considerations in phase estimation and event location using small-aperture regional seismic arrays, *Pure Appl. Geophys.* **167**, 381–399, doi: [10.1007/s00024-009-0024-1](https://doi.org/10.1007/s00024-009-0024-1).
- Gibbons, S. J., J. Schweitzer, F. Ringdal, T. Kværna, S. Mykkeltveit, and B. Paulsen (2011). Improvements to seismic monitoring of the European Arctic using three-component array processing at SPITS, *Bull. Seismol. Soc. Am.* **101**, no. 6, 2737–2754.
- Harris, D. B., and T. Kvaerna (2010). Superresolution with seismic arrays using empirical matched field processing, *Geophys. J. Int.* **182**, no. 3, 1455–1477.
- Institute of Seismology (1980a). *The Finnish National Seismic Network*, GFZ Data Services, Other/Seismic Network, Finland, doi: [10.14470/UR044600](https://doi.org/10.14470/UR044600).
- Institute of Seismology (1980b). The Finnish national seismic network [Data set], available at <https://www.seismo.helsinki.fi/bulletin/list/norBull.html> (last accessed August 2023).
- Kingma, D. P., and J. Ba (2014). Adam: A method for stochastic optimization, *Proc. of the Third International Conf. for Learning Representations*, San Diego, 2015, doi: [10.48550/arxiv.1412.6980](https://arxiv.org/abs/1412.6980).
- Köhler, A., C. Alvizuri, B. Dando, B. Goertz-Allmann, J. Schweitzer, V. Oye, B. Lund, P. Schmidt, M. Roth, G. Eggertsson, *et al.* (2023). Relative locations and moment tensors of the Nord stream pipeline events, *Technical Rept. Copernicus Meetings, EGU General Assembly 2023*, Vienna, Austria, 24–28 April 2023, EGU23-7019, doi: [10.5194/egusphere-egu23-7019](https://doi.org/10.5194/egusphere-egu23-7019).
- Köhler, A., E. Myklebust, and S. Mæland (2022). Enhancing seismic calving event identification in Svalbard through empirical matched field processing and machine learning, *Geophys. J. Int.* **230**, no. 2, 1305–1317.

- Kong, Q., D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft (2019). Machine learning in seismology: Turning data into insights, *Seismol. Res. Lett.* **90**, 1, 3–14.
- Kværna, T., B. D. E. Dando, and S. J. Gibbons (2023). Seismic monitoring of Novaya Zemlya: Progress, challenges, and prospects, *Seismol. Res. Lett.* **94**, no. 3, 1495–1508.
- Kværna, T., D. B. Harris, S. P. Nåsholm, A. Köhler, and S. J. Gibbons (2023). Tracking aftershock sequences using empirical matched field processing, *Geophys. J. Int.* doi: [10.1093/gji/ggad297](https://doi.org/10.1093/gji/ggad297).
- Kværna, T., J. Schweitzer, L. Taylor, and F. Ringdal (1999). *Monitoring of the European Arctic Using Regional Generalized Beamforming, NORSAR Scientific Report: Semiannual Technical Summary, 1 October–31 March 1999 No. 2-1998/1999*, NORSAR, Kjeller, Norway, 78–94.
- Le Bras, R., N. Arora, N. Kushida, P. Mialle, I. Bondar, E. Tomuta, F. K. Alamneh, P. Feitio, M. Villarroel, B. Vera, et al. (2021). NET-VISA from cradle to adulthood. A machine-learning tool for seismo-acoustic automatic association, *Pure Appl. Geophys.* **178**, 2437–2458.
- Le Bras, R., H. Swanger, T. Sereno, G. Beall, R. Jenkins, W. Nagy, and A. Henson (1994). Global Association, *Final Rept.*, Science Applications International Corporation, San Diego, California, available at <https://apps.dtic.mil/sti/pdfs/ADA304805.pdf> (last accessed August 2023).
- Lund, B., G. Eggertsson, A. Tryggvason, P. Schmidt, M. Roth, T. Larsen, P. Voss, T. Dahl-Jensen, N. Rinds, A. Köhler, et al. (2023). The Nord stream underwater explosions: Location, classification and yield estimation, *Technical Rept. Copernicus Meetings, EGU General Assembly 2023*, Vienna, Austria, 24–28 April 2023, EGU23-6775, doi: [10.5194/egusphere-egu23-6775](https://doi.org/10.5194/egusphere-egu23-6775).
- Mousavi, S. M., and G. C. Beroza (2023). Machine learning in earthquake seismology, *Annu. Rev. Earth Planet. Sci.* **51**, 105–129.
- 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, *Nat. Commun.* **11**, no. 1, 1–12.
- Niazi, M. (1966). Corrections to apparent azimuths and travel-time gradients for a dipping Mohorovicic discontinuity, *Bull. Seismol. Soc. Am.* **56**, 491–509.
- NORSAR (1971a). NORSAR seismic bulletins, doi: [10.21348/b.0001](https://doi.org/10.21348/b.0001).
- NORSAR (1971b). NORSAR station network [Data set], doi: [10.21348/d.no.0001](https://doi.org/10.21348/d.no.0001).
- Ottmöller, L., J. Michalek, J.-M. Christensen, U. Baadshaug, F. Halpaap, Ø. Natvik, T. Kværna, and V. Oye (2021). UiB-NORSAR EIDA node: Integration of seismological data in Norway, *Seismol. Soc. Am.* **92**, no. 3, 1491–1500.
- Prieto, G. A. (2022). The multitaper spectrum analysis package in Python, *Seismol. Res. Lett.* **93**, 1922–1929.
- Prieto, G. A., R. Parker, and F. Vernon III (2009). A Fortran 90 library for multitaper spectrum analysis, *Comput. Geosci.* **35**, no. 8, 1701–1710.
- Ringdal, F., and E. S. Husebye (1982). Application of arrays in the detection, location, and identification of seismic events, *Bull. Seismol. Soc. Am.* **72**, S201–S224.
- Romeyn, R., A. Hanssen, and A. Köhler (2022). Long-term analysis of cryoseismic events and associated ground thermal stress in Adventdalen, Svalbard, *Cryosphere* **16**, no. 5, 2025–2050.
- Rost, S., and C. Thomas (2002). Array seismology: Methods and applications, *Rev. Geophys.* **40**, no. 3, 2–1–2–27, doi: [10.1029/2000RG000100](https://doi.org/10.1029/2000RG000100).
- Schweitzer, J. (2001a). HYPOSAT—An enhanced routine to locate seismic events, *Pure Appl. Geophys.* **158**, 277–289.
- Schweitzer, J. (2001b). Slowness corrections—One way to improve IDC products, *Pure Appl. Geophys.* **158**, 375–396.
- Schweitzer, J., J. Fyen, S. Mykkeltveit, S. Gibbons, M. Pirli, D. Kühn, and T. Kværna (2012). Seismic arrays, in *New Manual of Seismological Observatory Practice (NMSOP-2)*, Second Ed., (revised), P. Bormann (Editor), Deutsches GeoForschungsZentrum GFZ, Potsdam, Germany, 1–80.
- Schweitzer, J., A. Köhler, and J. M. Christensen (2021). Development of the NORSAR Network over the Last 50 Yr, *Seismol. Res. Lett.* **92**, no. 3, 1501–1511.
- Selby, N. D. (2008). Application of a generalized F detector at a seismometer array, *Bull. Seismol. Soc. Am.* **98**, no. 5, 2469–2481.
- Storchak, D. A., J. Schweitzer, and P. Bormann (2003). The IASPEI standard seismic phase list, *Seismol. Res. Lett.* **74**, no. 6, 761–772.
- Veikkolainen, T., J. Kortström, T. Vuorinen, I. Salmenperä, T. Luhta, P. Mäntyniemi, G. Hillers, and T. Tiira (2021). The Finnish national seismic network: Toward fully automated analysis of low-magnitude seismic events, *Seismol. Res. Lett.* **92**, no. 3, 1581–1591.
- Wessel, P., and W. H. F. Smith (1995). New version of the Generic Mapping Tools, *EOS Trans. AGU* **76**, 329–329.
- Zhu, W., and G. C. Beroza (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method, *Geophys. J. Int.* **216**, no. 1, 261–273.

Manuscript received 3 April 2023
Published online 23 August 2023