Modelling uncertainty in *P*-wave arrival-times retrieved from DAS data: case-studies from 15 fibre optic cables

E. Bozzi [1], N. Piana Agostinetti,[1,2] A. Fichtner [3], S. Klaasen,[3] A. Ugalde [4], B. Biondi,[5] S. Yuan,[5] T. Dahm,[6] M. Isken,[6] P. Paitz,[7] F. Walter,[7] A. F. Baird [8], C. Becerril,[9] T. Nishimura,[10] J. Shen [11], T. Zhu [11] and G. Saccorotti[12]

[1]*Department of Environmental and Earth Sciences, University of Milano-Bicocca, 20126 Milano, Italy. E-mail: e.bozzi3@campus.unimib.it*
[2]*Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Nazionale Terremoti (ONT), 00143 Rome, Italy.*
[3]*Department of Earth Sciences, Institute of Geophysics, ETH Zürich, 8092 Zurich, Switzerland*
[4]*Department of Marine Geosciences, Barcelona Center for Subsurface Imaging Institut de Ciències del Mar, CSIC, 08003 Barcelona, Spain*
[5]*Department of Geophysics, Stanford University, Stanford, 94305 CA, USA*
[6]*Department of Geophysics, Helmholtz Centre Potsdam, GFZ, German Research Centre for Geosciences, 14473 Potsdam, Germany*
[7]*Swiss Federal Institute for Forest, Snow and Landscape Research, WSL, 8903 Birmensdorf, Switzerland*
[8]*NORSAR, 2007 Kjeller, Norway*
[9]*Université Côte d'Azur, CNRS, Observatoire de la Côte d'Azur, IRD, Géoazur, 06560 Valbonne, France*
[10]*Department of Geophysics, Tohoku University, 980-0845 Sendai, Japan*
[11]*Department of Geosciences, Pennsylvania State University, State College 16802, PA, USA*
[12]*Istituto Nazionale di Geofisica e Vulcanologia [Pisa section], 56125 Pisa, Italy*

SUMMARY

Distributed acoustic sensing (DAS) technology enables the detection of waves generated by seismic events, generally as uniaxial strain/strain rate time-series observed for dense, subsequent, portions of a Fibre Optic Cable (FOC). Despite the advantages in measurement density, data quality is often affected by uniaxial signal polarization, site effects and cable coupling, beyond the physical energy decay with distance. To better understand the relative importance of these factors for data inversion, we attempt a first modelling of noise patterns affecting DAS arrival times for a set of seismic events. The focus is on assessing the impact of noise statistics, together with the geometry of the problem, on epicentral location uncertainties. For this goal, we consider 15 'real-world' cases of DAS arrays with different geometry, each associated with a seismic event of known location. We compute synthetic *P*-wave arrival times and contaminate them with four statistical distributions of the noise. We also estimate *P*-wave arrival times on real waveforms using a standard seismological picker. Eventually, these five data sets are inverted using a Markov chain Monte Carlo method, which offers the evaluation of the relative event location differences in terms of posterior probability density (PPD). Results highlight how cable geometry influences the shape, extent and directionality of the PPDs. However, synthetic tests demonstrate how noise assumptions on arrival times often have important effects on location uncertainties. Moreover, for half of the analysed case studies, the observed and synthetic locations are more similar when considering noise sources that are independent of the geometrical characteristics of the arrays. Thus, the results indicate that axial polarization, site conditions and cable coupling, beyond other intrinsic features (e.g. optical noise), are likely responsible for the complex distribution of DAS arrival times. Overall, the noise sensitivity of DAS suggests caution when applying geometry-only-based approaches for the *a priori* evaluation of novel monitoring systems.

Key words: Bayesian inference; Monte Carlo methods; Distributed acoustic sensing.

1 INTRODUCTION

Understanding the origins of uncertainty in observational data is a crucial aspect of seismology. Such uncertainty is potentially generated by different sources, which can influence our ability to use geophysical data for making inferences about the Earth's interior. This knowledge is especially critical for the evaluation of new geophysical instruments and methodologies. Among novel methods, distributed acoustic sensing (DAS), following pioneering industry tests in the early 2010s (Mestayer *et al.* 2011; Molenaar *et al.* 2011; Parker *et al.* 2014; Hill 2015) is now often used as a monitoring apparatus by the seismological community (Zhan 2020). DAS exploits laser interrogation of FOCs to obtain a very-dense array of axial strain/strain-rate sensors. Such high-sampling capability is unreachable with a standard array of sensors (Bakulin *et al.* 2020). Accordingly, DAS provides a full picture of the seismic wavefield (Paitz *et al.* 2021), for a variety of sources (such as earthquakes) (Lindsey & Martin 2021). Both novel installations (Fichtner *et al.* 2022; Klaasen *et al.* 2021; Walter *et al.* 2020) and/or commercial telecommunication FOCs (Biondi *et al.* 2021) can be equally used. Hence, DAS can be exploited to monitor both natural (Biondi *et al.* 2017; Lindsey *et al.* 2017; Lellouch *et al.* 2019; Hudson *et al.* 2021; Nayak & Ajo-Franklin 2021; Ugalde *et al.* 2022) and induced (Li & Zhan 2018; Karrenbach *et al.* 2019; Obermann *et al.* 2022) seismicity.

Despite the advantages posed by measurement density and application in poorly monitored environments, DAS is usually affected by lower signal-to-noise ratios (SNRs) compared to standard seismic sensors (Li & Zhan 2018; Walter *et al.* 2020) and could show coherence loss even for short interchannel distances (van den Ende & Ampuero 2021). Indeed, DAS suffers from signal directivity, that is the strain/strain rate is measured uniaxially for their longitudinal components along the fibre direction; strong amplitude variations related to FOC's coupling with the ground, and higher susceptibility to local changes in rock properties (Ajo-Franklin *et al.* 2019; van den Ende & Ampuero 2021; Trabattoni *et al.* 2022; Yang *et al.* 2022). These issues are even more crucial for superficial FOCs, due to marked variability of elastic parameters in shallow sediments. Furthermore, as with most modern seismic data sets, manual analysis of the abundant DAS channels (typically >1000) is usually unfeasible. Hence, automatic techniques are required to exploit DAS data for common seismological tasks, such as picking of arrival times for event location.

Arrival times at multiple seismological sensors are traditionally exploited for the estimation of the event locations (Lay & Wallace 1995). This procedure can also be extended to the numerous DAS channels, providing densely spaced information. However, the estimates of arrival times, especially when using automatic methods relying on amplitude ratio between background noise and the incoming signal, are significantly affected by the local SNR therefore, by the above-mentioned DAS features. Hence, the cable geometry, local geological conditions, and the pool of natural and anthropogenic noise sources, provide diversified noise environments throughout the FOC. Although the DAS data redundancy may help in the detection stage (Hudson *et al.* 2024, Li *et al.* 2021, Porras *et al.* 2024), the inversion of DAS arrival times for event locations may be adversely influenced from the aforementioned elements. Nevertheless, various successful case studies have already demonstrated DAS capability in detecting and/or locating seismic events (Lindsey *et al.* 2017; Jousset *et al.* 2018; Lellouch *et al.* 2020; Walter *et al.* 2020; Nishimura *et al.* 2021; van den Ende & Ampuero 2021; Zhu *et al.* 2021; Klaasen *et al.* 2021; Fichtner *et al.* 2022; Flóvenz *et al.*

2022; Trabattoni *et al.* 2022; Currenti *et al.* 2023; Klaasen *et al.* 2023; Biagioli *et al.* 2024). However, a study assessing the influence of specific noise sources on epicentral parameter uncertainties and covering diversified DAS deployments is still lacking.

The present study focuses on the quantitative evaluation of four potential noise distributions in the estimated arrival times for a seismic wave impinging at a FOC. We consider 15 real-world cases of DAS deployments, to have a better coverage of the possible interactions between geometries and noise sources. For each DAS array, we use earthquake recordings in which independent locations were already available. Four synthetic *P*-wave arrival-times data sets, contaminated following four noise assumptions, and one observed data set, estimated using a standard automatic picking procedure, are prepared and then inverted using a probabilistic approach that permits the estimate of the *a posteriori* probability density function for source location. Synthetic and experimental locations are then mutually compared to evaluate the similarity of both shape and location.

Our results indicate that modelling noise sources strictly dependent on the geometrical features of the DAS array fails in fully reproducing the experimental location uncertainties. Thus, more complex and difficult-to-model *a priori* features, such as signal directivity, site effects or cable coupling, are likely key contributors. These complexities must be properly considered when designing the optimal geometry and extent of DAS projects for earthquake monitoring at the local and regional scales.

2 DATA AND METHODS

2.1 DAS arrays: 15 real word cases

This study considers data from 15 DAS installations (references in the following paragraphs as RHONEGLETSCHER, MOUNT MEAGER, POROTOMO, GRÍMSVÖTN, STANFORD-1, STANFORD-2, FORESEE, AZUMA VOLCANO, HENGILL-NORSAR, HENGILL-GFZ, HCMR, MONTEREY, NESTOR, MEUST and CANARY), obtained either from open-access repositories (Feigl *et al.* 2016; Lindsey *et al.* 2020; Lior 2020; Villasenor *et al.* 2020; Zhu *et al.* 2021; Klaasen 2021; Spica *et al.* 2023) or restricted databases. Three main categories are identified to describe the installation contexts of each experiment, that is (1) scientific 'fit-for-purpose' cables (Fig. 2), (2) superficial telecommunication cables (Fig. 3) and (3) submarine telecommunication cables (Fig. 4). 'Fit-for-purpose' DAS arrays refer to installations designed from scratch for research purposes, without utilizing previously existing cables. Typically such arrays exhibit good azimuthal coverage of the retrieved source regions (e.g. POROTOMO and GRÍMSVÖTN deployments). Superficial telecommunication arrays are generally constrained by the urban street network, on which all the main infrastructures are installed (e.g. STANFORD-1, STANFORD-2 and FORESEE). Submarine telecommunication arrays have less articulated geometries (e.g. MONTEREY and NESTOR deployments), due to the lower number of infrastructural obstacles. Hence, among the collected data sets they have lower azimuthal coverage.

For each DAS array, we select recordings from local events (purely tectonic, volcano-tectonic and ice quakes) located within a distance of less than 100 km from the DAS channel closest to the interrogator. The known location is obtained either from official seismological catalogues or from the institution responsible for the
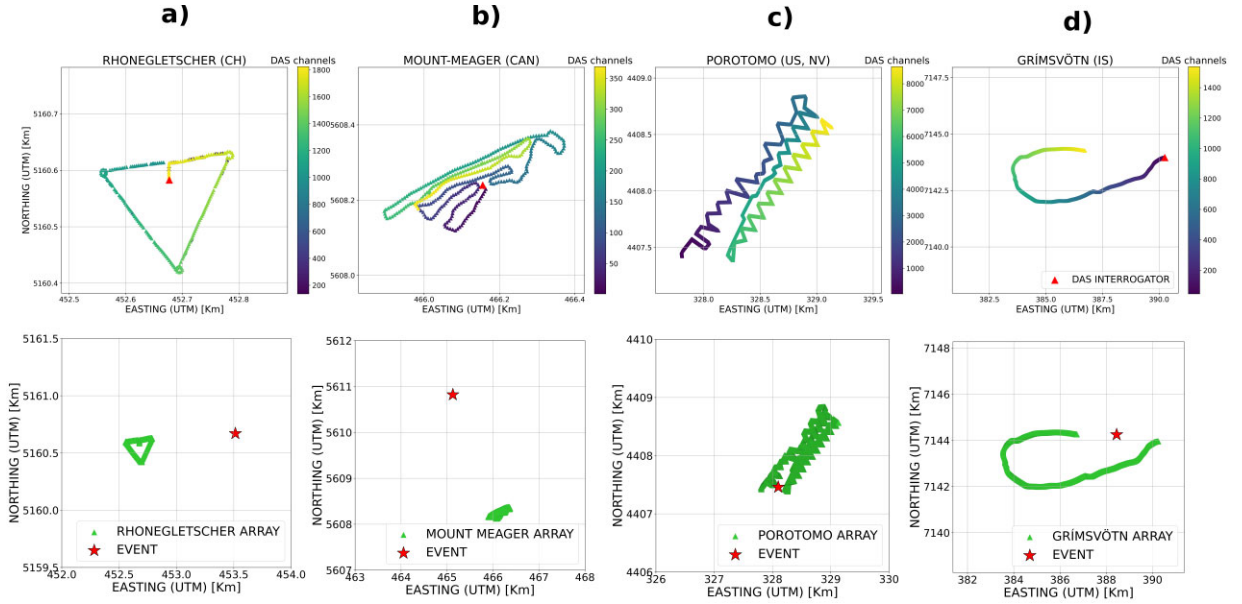
Figure 2. Upper row: 'fit-for-purpose' DAS array geometries (red triangle showing the location of the DAS interrogator and colours representing the number of channels). Bottom row: geometrical relations between the array and the known event location (red star). (a) RHONEGLETSCHER data set, (b) MOUNT-MEAGER data set, (c) POROTOMO data set and (d) GRÍMSVÖTN data set.
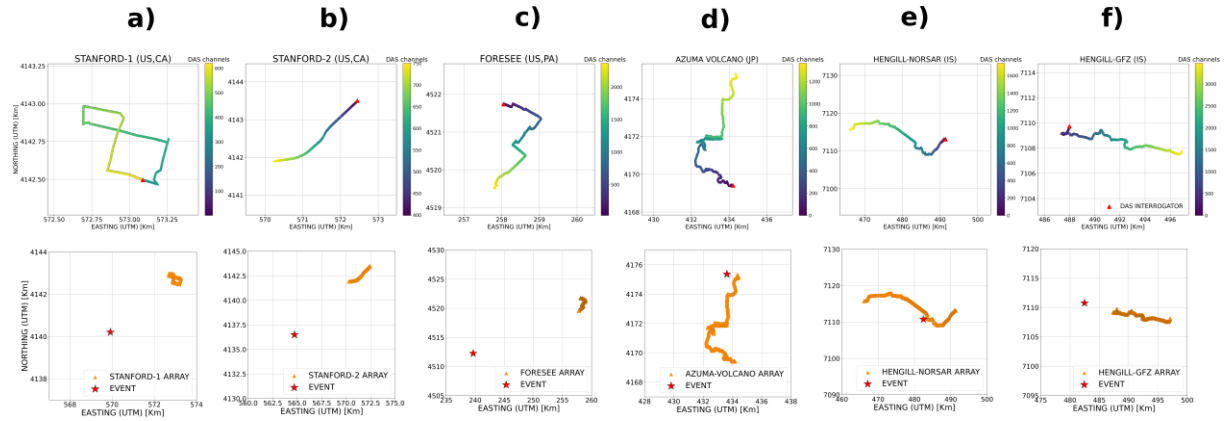

Figure 3. Upper row: 'superficial telecommunication' DAS array geometries. Bottom row: geometrical relations between the array and the known event location. The meanings of the symbols are the same as in (Fig 2). (a) STANFORD-1 data set, (b) STANFORD-2 data set, (c) FORESEE data set, (d) AZUMA-VOLCANO data set, (e) HENGILL-NORSAR data set and (f) HENGILL-GFZ data set.

specific data acquisition (see Data Availability statement). Alternatively, where an *a priori* location was not available, this information is estimated from the inversion of selected arrival time data (manually checked and picked) on available colocated geophones or DAS channels (MOUNT MEAGER, RHONEGLETSCHER and GRÍMSVÖTN). Table 1 summarizes the major features of the FOCs used in the study and the selected events.

2.2 Automated picking of *P*-wave arrival times

The recorded events are pre-processed using de-trending (linear and constant trends removal) and bandpass filtering. The corner frequencies are chosen specifically for each event, inspecting the ratio of the frequency spectra before and during the event. Spatial subsampling (sum of neighbouring DAS channels) is performed under the condition that the gauge length is at least twice the channel spacing. This method is useful in reducing incoherent noise and allows for an improvement of the SNR, especially in poorly coupled

array portions (Piana Agostinetti *et al.* 2022). While more sophisticated processing techniques for DAS data have been suggested (Isken *et al.* 2022), they were not tested in this study. The emphasis was placed on maintaining a straightforward data preparation procedure and focusing on subsequent modelling of the noise sources. A standard automatic picking procedure (Baer & Kradolfer 1987) is adopted to measure the first onsets of each pre-processed event. *S*-waves are voluntarily not picked since the noise modelling procedure (see Section 1.3.3) considers mis-picks of *S* waves classified as *P* arrivals. Table 2 summarizes these pre-processing steps for each event-DAS array pair alongside the parameters used for the automatic picker. Untriggered DAS channels are not included as data for the successive inversion. Noteworthy, other automatic pickers might perform better for DAS data (e.g. Zhu *et al.* 2023). However, a comprehensive assessment of the performances of different picking algorithms for DAS data is beyond the scope of this work, and it is left for future studies.
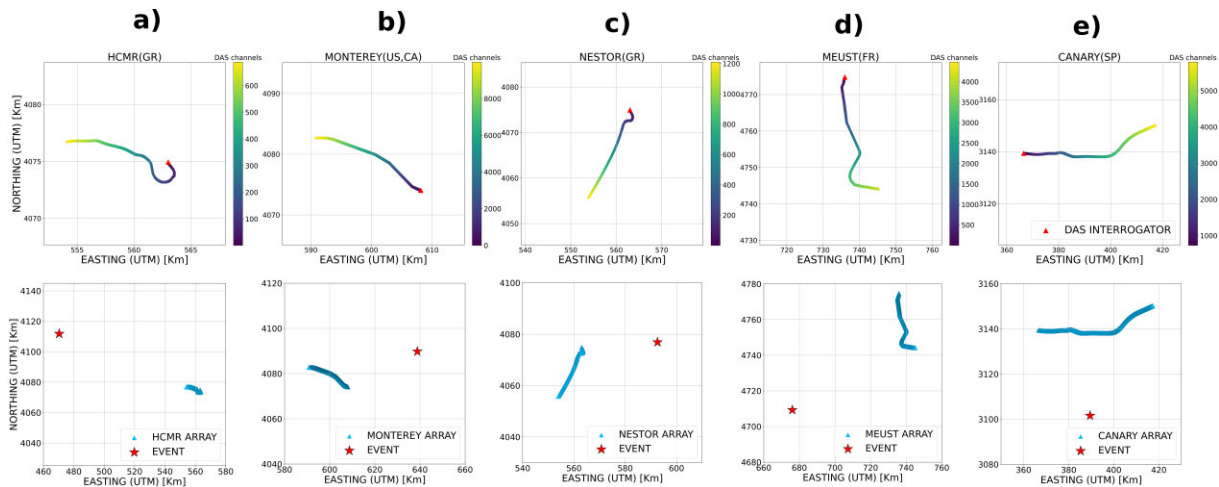
Figure 4. Upper row: 'Submarine telecommunication' DAS array geometries. Bottom row: geometrical relations between the array and the known event location. The meanings of the symbols are the same as in (Fig 2). (a) HCMR data set, (b) MONTEREY data set, (c) NESTOR data set, (d) MEUST data set and (e) CANARY data set.
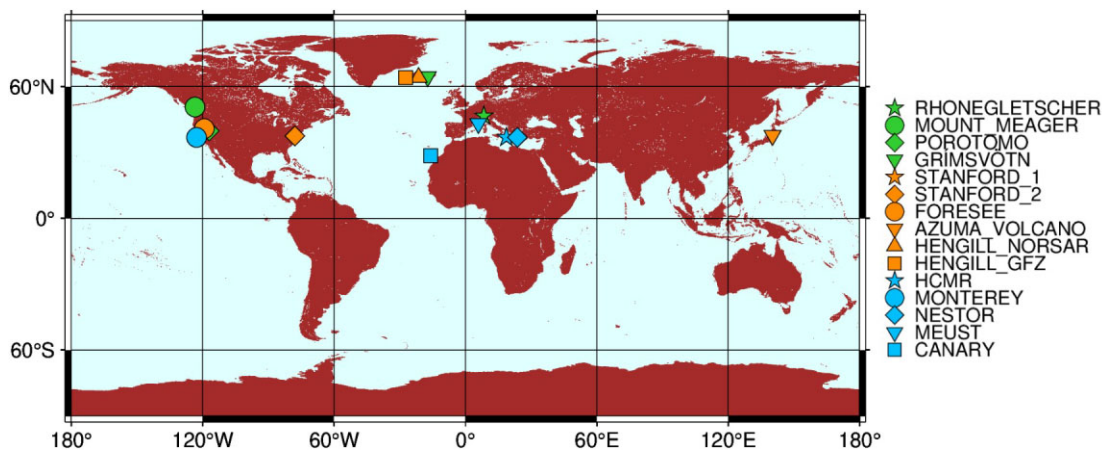

Figure 1. Geographical distribution of the data sets collected for this study. Three categories are identified: (1) 'Fit-for-purpose' fiber optic cables (green), which represent FOCs installed for scientific studies (MOUNT-MEAGER, RHONEGLETSCHER, POROTOMO and GRÍMSVÖTN), (2) 'superficial' telecommunication fibre optic cables (orange), which represent FOCs installed in superficial environments ($STANFORD-1$, $STANFORD-2$, FORESEE, $AZUMA-VOLCANO$, $HENGILL-NORSAR$ and $HENGILL-GFZ$), (3) 'Oceanic telecommunication' fibre optic cables (blue), which represent FOCs installed in continental shelf or continental shelf-oceanic floor transition environments (HCMR, MONTEREY, NESTOR, MEUST and CANARY).

Table 1. Data sets collected in this study.

Context	DAS ID	Location	Length [km]	N° channels (gauge length, channel spacing [m])	Event dist. [km]
Fit-for-purpose	RHONEGLETSCHER	Switzerland	1.7	422 (8,4)	0.8
Fit-for-purpose	MOUNT-MEAGER	Canada	3	380 (8,8)	2.8
Fit-for-purpose	POROTOMO	Nevada [USA]	8.6	8620 (10,1)	0.3
Fit-for-purpose	GRÍMSVÖTN	Iceland	14.1	1728 (8.16,8.16)	1.8
superficial telecom	STANFORD-1	California [USA]	2.6	626 (8.16,4.08)	3.9
superficial telecom	STANFORD-2	California [USA]	2.8	353 (16,8)	10.4
superficial telecom	FORESEE	Pennsylvania [USA]	4.9	2432 (10,2)	20.7
superficial telecom	AZUMA VOLCANO	Japan	14.3	1404 (40.8,10.2)	6.0
superficial telecom	HENGILL NORSAR	Iceland	34.8	1742 (20,10)	9.3
superficial telecom	HENGILL GFZ	Iceland	14.6	3648 (6,3)	5.6
Submarine telecom	HCMR	Greece	13.2	688 (19.2,19.2)	99.7
Submarine telecom	MONTEREY	California [USA]	19.9	9993 (10,2)	34.5
Submarine telecom	NESTOR	Greece	26.2	1365 (19.2,19.2)	29.6
Submarine telecom	MEUST	France	44.8	4480 (19.2,19.2)	88.8
Submarine telecom	CANARY	Canary islands [SP]	59.8	5983 (10,10)	44.2

Note: The epicentral distance is computed from the DAS channel closest to the interrogator.

Table 2. Pre-processing and picker parameters.

DAS ID	Bandpass corners [Hz]	Stacking [N° of stacks]	N° of resulting channels	Automatic picker main parameters
RHONEGLETSCHER	[10–80]	yes [2]	211	[15,30,20,60,100,100]
MOUNT-MEAGER	[10–100]	no	380	[10,20,20,60,100,100]
POROTOMO	[1–20]	yes [10]	860	[10,20,20,60,100,100]
GRÍMSVÖTN	[1–50]	no	1728	[20,40,20,60,100,100]
STANFORD-1	[1–20]	yes [2]	313	[10,20,20,60,100,100]
STANFORD-2	[1–25]	yes [2]	176	[10,20,20,60,100,100]
FORESEE	[1–20]	yes [5]	486	[30,60,20,60,100,100]
AZUMA VOLCANO	[1–50]	yes [4]	351	[20,40,20,60,100,100]
HENGILL NORSAR	[1–20]	yes [2]	579	[10,20,20,60,100,100]
HENGILL GFZ	[1–20]	yes [2]	1824	[20,40,20,60,100,100]
HCMR	[5–20]	no	688	[30,60,20,60,100,100]
MONTEREY	[1–25]	yes [5]	1998	[30,60,20,60,100,100]
NESTOR	[5–25]	no	1365	[20,40,20,60,100,100]
MEUST	[5–25]	no	4480	[20,40,20,60,100,100]
CANARY	[5–15]	no	5983	[30,60,20,60,100,100]

Note: The automatic picker parameters (Baer & Kradolfer 1987) (thr1, thr2, tupevent, tdownmax, present-len, p-dur) are provided as described in the documentation of the implemented python package (Beyreuther *et al.* 2010).

2.3 Noise models

Four synthetic tests (i.e. SYNTH-01, SYNTH-02, SYNTH-03 and SYNTH-04) are implemented to model specific perturbations of *P*-wave traveltimes along the FOCs. Synthetic traveltimes are computed on a homogenous model from the known locations. While this could seem, at first glance, a coarse approximation, our goal is not to retrieve accurate source locations; rather, we aim to investigate the effects of different noise models on the pattern of location uncertainties. Therefore, we do not make any attempt to use more accurate, site-specific velocity models, since this would add additional complexities. Moreover, although every installation context might have specific noise patterns for the estimated arrival times, this study focuses on simulating simpler noise distributions, common to all the analysed data sets. Nevertheless, for the FORESEE deployment, we conducted a multi-event analysis (Figs S7-1, S7-2 and S7-3 of the supplementary material). Modelling these distributions is based on features that ar often observed in estimated arrival times, such as: (a) an increase in the dispersion of arrival times for increasing epicentral distance, (b) irregular grouping of arrival times in 'standard' and 'time-delayed' ensembles and (c) complex delayed arrival times depending on SNR (i.e. positive outliers). While the first noise assumption is related to the geometric relationships between the event and the array, the remaining distributions are independent of geometry. Fig. 5 provides an overview of the modelling procedure for an analysed case study (AZUMA VOLCANO).

We quantitatively describe the computation of noise-contaminated synthetic *P* traveltimes $st_{P,i}$ for the ith DAS channel, starting from noise-free traveltimes $tt_{P,i}$ in the following subsections.

2.3.1 SYNTH-01: white Gaussian noise with constant variance along the DAS array

In this test Gaussian noise with zero mean, $\mu = 0$, and constant variance σ_1 for all the DAS channels is modelled (eq. 1). This test reproduces an idealized DAS array characterized by identical arrival time uncertainties along its entire length. This assumption proves useful in evaluating the inherent uncertainty arising from the problem's geometry;

$$st_{P,i} = tt_{P,i} + \mathcal{N}(\mu, \sigma_1). \tag{1}$$

2.3.2 SYNTH-02: distance-dependent white Gaussian noise

In this test Gaussian noise with variance linearly dependent on the distance (d_i) from the event ($\mu = 0$ and σ_2) is modelled (eq. 2). This approach aims at reproducing the increase of the uncertainty in arrival times estimates for increasing epicentral distance, following the decrease of signal amplitude due to anelastic attenuation and geometrical spreading (Klein 2002):

$$st_{P,i} = tt_{P,i} + \mathcal{N}(\mu, \sigma_2), \tag{2}$$

where

$$\sigma_2 = \sigma_1 \left(\alpha + \beta \left(\frac{d_i - d_{min}}{d_{max} - d_{min}} \right) \right)$$

$$\alpha = 0.5$$

$$\beta = 2$$

d_{max} = maximum distance from the known event location

d_{min} = minimum distance from the known event location

2.3.3 SYNTH-03: mis-picked P wave

This test builds upon the statistical assumptions used in SYNTH-01 and extends its application to a time-delayed seismic phase, specifically the 'S' phase. SYNTH-03 seeks to simulate a specific noise pattern observed in arrival times estimated on DAS data, that is a 'mis-pick' between the first onset and a delayed stronger seismic phase. Since both local directivity and cable coupling inhomogeneities can explain this error source, an arbitrary threshold (α) is set to establish the ratio between simulated *P* (70 per cent) and *S* phases (30 per cent, eq. 3). The validity of this test is limited to events close to the DAS array, since an ambiguity between *P* and *S* phases might exist only for very-short moveout;

$$st_{P,i} = \begin{cases} tt_{P,i} + \mathcal{N}(\mu, \sigma_1) & \text{if } \alpha < 0.7, \\ tt_{S,i} + \mathcal{N}(\mu, \sigma_1) & \text{if } \alpha \geq 0.7 \end{cases} \quad \alpha \sim U(0, 1). \tag{3}$$

2.3.4 SYNTH-04: time-delayed P-wave picking

In this test the median SNR ($\tilde{x}$) is considered as the threshold for categorizing channels into two groups: (1) channels to be contaminated with SYNTH-01-like (Gaussian) noise statistics (mean = μ,
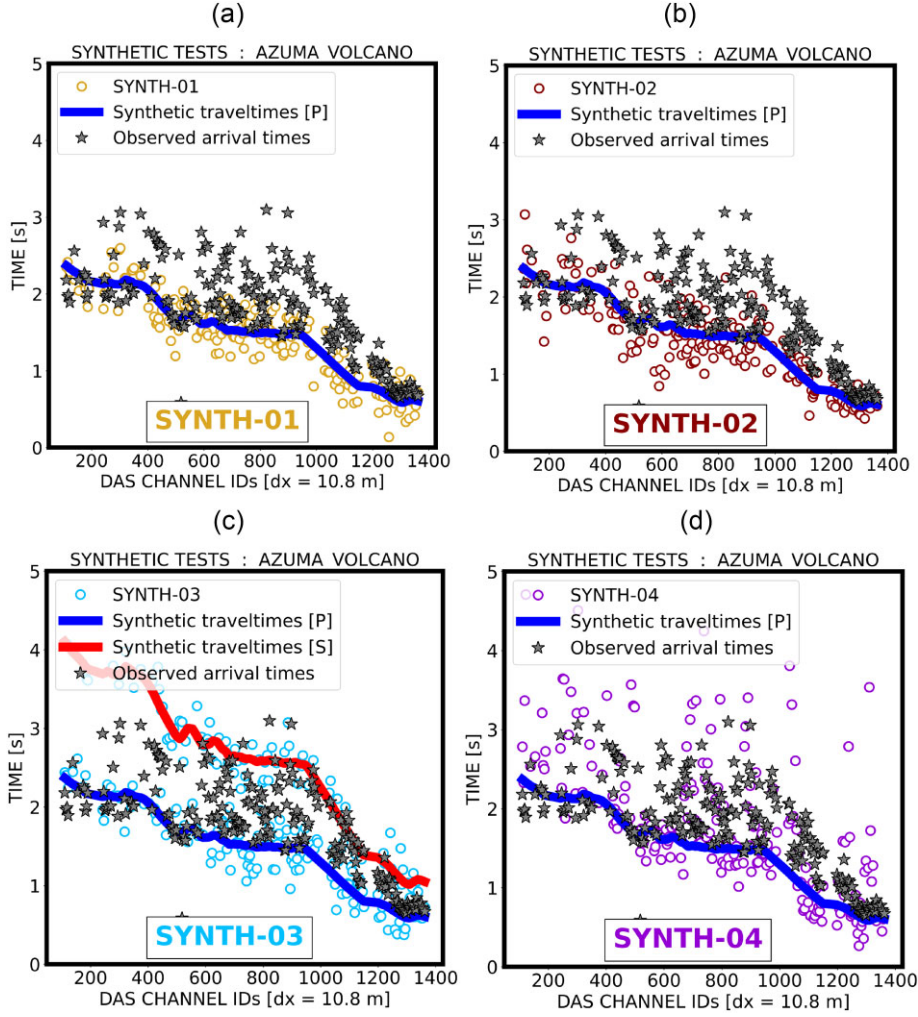
Figure 5. Observed arrival times (grey stars), synthetic traveltimes (solid lines, blue and red for *P* and *S* waves, respectively) and noise-contaminated arrival times (different colors depending on the synthetic test). (a) SYNTH-01 test, (b) SYNTH-02 test, (c) SYNTH-03 test and (d) SYNTH-04 test for the AZUMA-VOLCANO data set. Observed arrival times and SYNTH- data are then inverted for estimating hypocentral location parameters.

variance $= \sigma_1$), (2) channels to be contaminated with the absolute value of a broader Gaussian noise source (mean $= \mu$, variance $= 4 * \sigma_1$, eq. 4). This test is considered the first tentative to examine a frequent observation in automatically retrieved DAS arrival times, that is: delayed arrival times whose distribution does not exhibit any evident pattern. Several mechanisms may concur in generating this phenomenon, for instance local site conditions or signal-generated noise. Although this noise model is similar to what has been modelled already by SYNTH-03, in this case, the error source is not derived from a specific phase mis-pick, but more generally related to the SNR.

$$st_{P,i} = \begin{cases} tt_{P,i} + \mathcal{N}(\mu, \sigma_1) & \text{if } SNRi \geq \tilde{x}, \\ tt_{P,i} + |\mathcal{N}(\mu, 4\sigma_1)| & \text{if } SNRi \leq \tilde{x} \end{cases} \qquad \tilde{x} = \text{SNR median.}$$

(4)

2.4 McMC location of seismic events

Data uncertainties affect model parameter estimation in geophysical inverse problems, especially when non-linearity and non-uniqueness exist (Mosegaard & Tarantola 2002). The earthquake location problem displays these characteristics. Casting a geophysical inverse problem in a Bayesian framework (Bayes & Price 1763)

means working with probability distributions on model parameters, based on: (i) prior information, (i) updated information coming from data and (ii) combined posterior information (Posterior Probability Density, PPD). This method is fundamental to assess how the data uncertainty affects the reliability of the estimated subsurface parameters. Here, *P*-wave arrival times, both synthetic and observed, are exploited for the estimation of the hypocentral parameters. A Hierarchical Markov chain Monte Carlo (McMC) approach is adopted for Bayesian inference of the event and elastic model parameters (longitude, latitude, depth, origin time, *Vp*). Details about the specific McMC algorithm can be found in (Riva *et al.* 2024). In particular, we focus on the PPDs expressing longitude and latitude uncertainties, derived from the inversion of *P*-wave arrivals only. For both observed and synthetic data, the solutions are represented as a set of samples from the PPD, coloured according to their spatial density. The results from SYNTH-01 inversions are exploited for assessing a 'geometrical uncertainty' of the problem (see Section 2.5.1), that is the PPDs for longitude and latitude estimated with SYNTH-01 traveltimes are examined with four geometrical descriptors, to look for their possible correlations with such parameters. More generally, comparing the results from the synthetic data sets and the observed one, the focus is on the relative change of the epicentral location

PPDs (sample clouds) as a result of the adoption of different noise models (see Section 2.5.2).

2.5 Description of the epicentral location PPDs

2.5.1 Geometrical descriptors

Due to the uniaxial sensitivity of the FOC, DAS geometry is generally assumed to be one of the most critical parameters for detecting seismic phases. Its influence is therefore investigated here in detail, using results from the inversion of SYNTH-01 data sets. The average standard deviations for the hypocentre solutions of SYNTH-01 test are scaled to the area encompassed by the array, to account for their relative dimensions. The geometrical constrain of the problem is evaluated by associating specific descriptors to the computed standard deviations. Specifically, four geometrical parameters are evaluated in this work to describe the relations between the known event and the DAS arrays (i.e. AREA-PAR-1, AREA-PAR-2, aspect ratio, and azimuthal gap). Among these, two are explicitly designed for DAS applications, AREA-PAR-1 and AREA-PAR-2, while the rest are conventional geometric descriptors in seismology (Fig. 6):

(i) AREA-PAR-1: the ratio between the distance between the event and the closest channel.

(ii) AREA-PAR-2: the ratio between the area covered by the array and the area defined by the vectors connecting event and DAS channels.

(iii) Aspect ratio: the ratio between the differences in latitude and longitude extension of the DAS array.

(iv) Azimuthal gap: the largest angle between two vectors connecting DAS channels and the event.

2.5.2 Location uncertainties comparison

Both synthetic and observed traveltimes are inverted, and the PPDs for the epicentral location are compared considering their shape and spatial location. We make some assumptions about the four noise models that could influence the retrieved PPDs from synthetic tests. For example, the noise variance, denoted as σ_1, assumed for modelling the synthetic traveltimes, controls the dispersion of the solution samples (i.e. the size of the cloud of samples). Nevertheless, here the primary focus is on the *distribution* of the solutions, that is the shape and location of the PPDs. The ultimate goal of this procedure is to compare the PPDs derived from the inversion of observed and synthetic traveltimes and to evaluate the performance of the noise models in (partially) reproducing the experimental location uncertainties. For this goal, we (a) cross-correlated the normalized PPDs, thus considering the shape similarity and (b) calculated the Cartesian distances of the PPD maximum values, thus accounting for the differences in their spatial location. As a result, two scores (Score 1 and Score 2), one for shape similarity and one for spatial separation, are defined. Operationally, we compute 2-D grids (here we consider a dimension of 200 × 200) out of McMC samples, for the computation of Score 1 and Score 2. After normalization of the scores on their maximum values, an average value is eventually obtained for each synthetic test (Total Score, Fig. 7). The Total Score identifies a representative synthetic test for each case study, thus the one that 'mimics at best' the observed PPDs (Fig. 8).

3 RESULTS

Fig. 8 illustrates an example of location PPDs derived by inverting observed and synthetic arrival times contaminated with four noise distributions. In general, we note that both the shape and location of the PPDs depend markedly on the assumptions made to contaminate the noise-free traveltimes. For a general overview, we collected the results of all the synthetic tests in the supplementary material (Figs S1–S15).

Following the procedure introduced in Section 2.5.2, we evaluated the similarity in shape (Score 1) and spatial location (Score 2) of the experimental and synthetic PPDs and eventually identified, for each case study, the test mimicking at best the observed location uncertainties (Total Score, Fig. 9). Figs S16 and S17 in the supporting information collect Score 1, Score 2 and the Total Score for each case study. The statistics of Score 1 and Score 2 indicate SYNTH-02 and SYNTH-04 as the noise models 'best mimicking' the shape (Fig. 9a) and location (Fig. 9b) of observed PPDs, respectively. Thus, a not-purely Gaussian noise source (SYNTH-04) helps reproduce the misplacement of the observed locations compared to the known locations. Overall, the Total Score shows similar results, with two relative maxima: SYNTH-02 and SYNTH-04 (Fig. 9c). Therefore, synthetic (a) noise sources related to the geometry and (b) independent, more complex distributions, have similar performance in mimicking the observed location uncertainties.

We now proceed to a more careful examination of the different case studies:

(i) SYNTH-01 shows the most similar PPDs compared to the experimental ones (Score 1) for POROTOMO, HENGILL-GFZ and MEUST case studies, while the closest (Score 2) for MOUNT-MEAGER, HENGILL-GFZ, MEUST and CANARY. Overall (Total Score), SYNTH-01 is the 'best mimicking' PPD for POROTOMO, HENGILL-GFZ (Fig. 11a), NESTOR (Fig. 11b) and MEUST. SYNTH-01 is correlated with three of the analysed geometrical descriptors (see Section 2.5.1). To highlight this, we adopted a logarithmic scale for the normalized standard deviation of SYNTH-01 solutions (normalization based on the area spanned by the array), thus mitigating the significant influence of low values. Specifically, a positive correlation is observed for AREA-PAR-1, AREA-PAR-2, and the azimuthal gap. Noteworthy the azimuthal gap values are generally high, with approximately 70 per cent of arrays falling within the range of 300° to 350°. Notable exceptions are POROTOMO, GRÍMSVÖTN, AZUMA VOLCANO and HENGILL-NORSAR deployments, in the range of 140° to 250°. The aspect ratio does not show an evident correlation (Fig. 10). Moreover, clustering depending on the installation context is not evident.

(ii) SYNTH-02 has the highest Score 1 for MOUNT-MEAGER, STANFORD-2, HCMR, NESTOR and CANARY, and the highest Score 2 for POROTOMO, FORESEE and HCMR. Overall, SYNTH-02 has the highest Total Score for MOUNT-MEAGER (Fig. 11c), STANFORD-2, FORESEE and HCMR (Fig. 11d).

(iii) SYNTH-03 has the highest Score 1 for GRÍMSVÖTN, AZUMA-VOLCANO and HENGILL-NORSAR and the highest Score 2 for AZUMA-VOLCANO and HENGILL-NORSAR. Overall, it shows the highest Total Score for GRÍMSVÖTN, AZUMA-VOLCANO (Fig. 12a) and HENGILL-NORSAR (Fig. 12b). Interestingly, for the AZUMA-VOLCANO case study, if compared to the outcomes from SYNTH-01, SYNTH-02 and SYNTH-04 (Fig. S8), SYNTH-03 PPD is shifted toward the north–northeast, coherently to the observed location.

(iv) SYNTH-04 has the highest Score 1 for RHONE-GLETSCHER, STANFORD-1, FORESEE and MONTEREY and
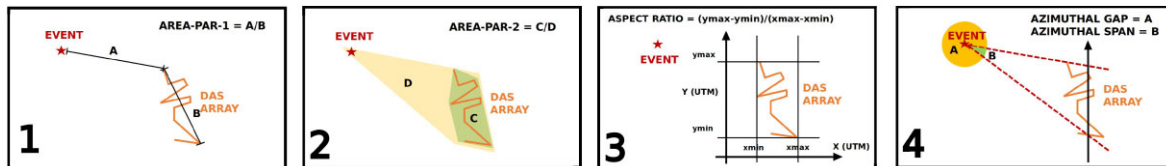
Figure 6. Geometrical parameters to describe the relations between the DAS arrays and the recorded event. Simple sketches describing the geometrical parameters: AREA-PAR-1, AREA-PAR-2, ASPECT RATIO and AZIMUTHAL GAP.
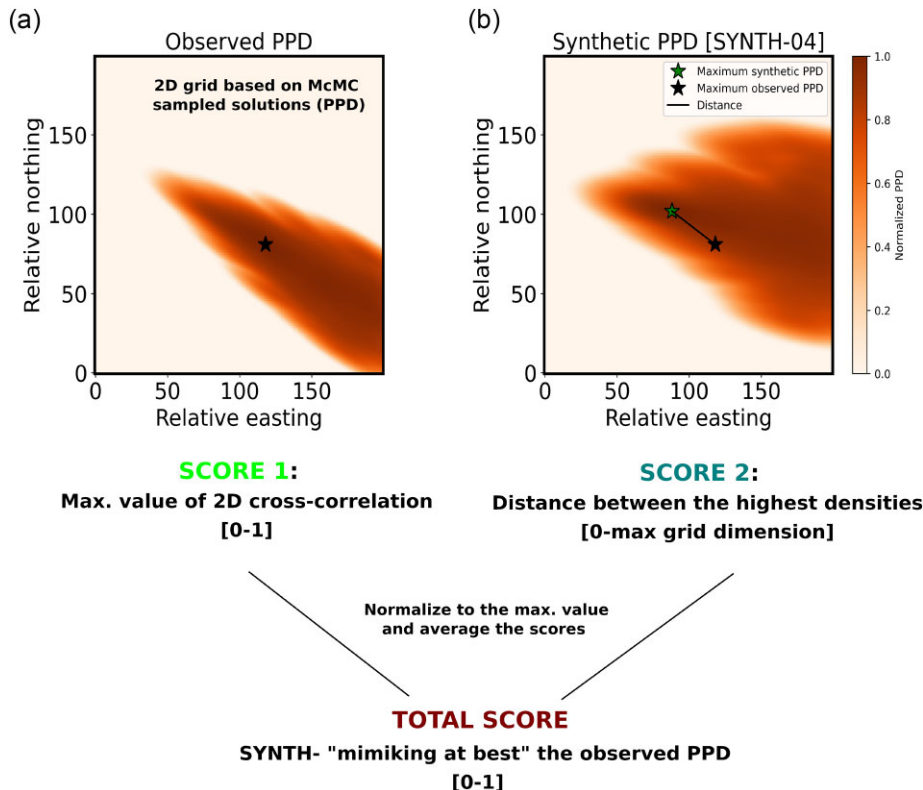

Figure 7. The adopted methodology to compare observed and synthetic PPDs based on 'Scores'. The top section of the figure shows 2-D grids representing observed and synthetic PPDs for a case study (RHONEGLETSCHER). The bottom section describes how Score 1 and Score 2 are computed: Score 1 is derived from the 2D cross-correlation of the PPDs, and Score 2 is based on the distance between the highest PPD values. An average and normalized (0–1) value (Total Score) is obtained for each synthetic test. The highest value among the synthetic tests identifies the one 'mimicking at best' the observed PPDs in terms of both shape and location. (a) 2-D grid representing the observed PPD, where the black star marks the highest PPD value. (b) 2-D grid representing the synthetic PPD, with the green star indicating the highest PPD value, and the thick black line showing the distance between the observed and synthetic highest PPD values (Score 2).

the highest Score 2 for RHONEGLETSCHER, GRÍMSVÖTN, STANFORD-1, MONTEREY and NESTOR. Overall, it has the highest Total Score for RHONEGLETSCHER (Fig. 12c), GRÍMSVÖTN, STANFORD-1, MONTEREY (Fig. 12d) and CANARY. Noteworthy, SYNTH-03 and SYNTH-04 have *ex-aequo* the highest Total Score for GRÍMSVÖTN. Moreover, for RHONEGLETSCHER, GRÍMSVÖTN and MONTEREY case studies, SYNTH-04 reproduces the directionality shift of the possible solutions toward the observed location.

Although the adopted 'Score' method, for how it is conceptualized, always provides a 'best mimicking' synthetic test, few case studies visually exhibit poor similarity to the experimental PPDs (POROTOMO, STANFORD-2, HENGILL-GFZ and MEUST). Figs S3, S6, S10 and S14 provide a full coverage of them. Either a strong influence of the cable geometry or peculiar disturbances in the observed data, not accounted in the synthetic tests, are possible

explanations for these observations. Moreover, the modelled noise sources might interact collectively, rather than individually.

4 DISCUSSION

The noise models we defined to contaminate noise-free *P*-wave DAS travel times, helped better understand the complexity of the observed location uncertainties (Figs 11 and 12). In fact, synthetic tests indicate how altering noise assumptions on arrival times significantly affects the shape and azimuth of the cloud of possible event locations (Fig. 8). However, the noise models we consider in this analysis account only for a limited set of possible disturbances that affect DAS traveltimes. Indeed, other, here not explored, noise models might exist. Moreover, besides unexplored noise sources, discrepancies between the known and observed locations may arise due to 3-D heterogeneities along the paths from the known event to DAS channels, which are not considered in the forward model.
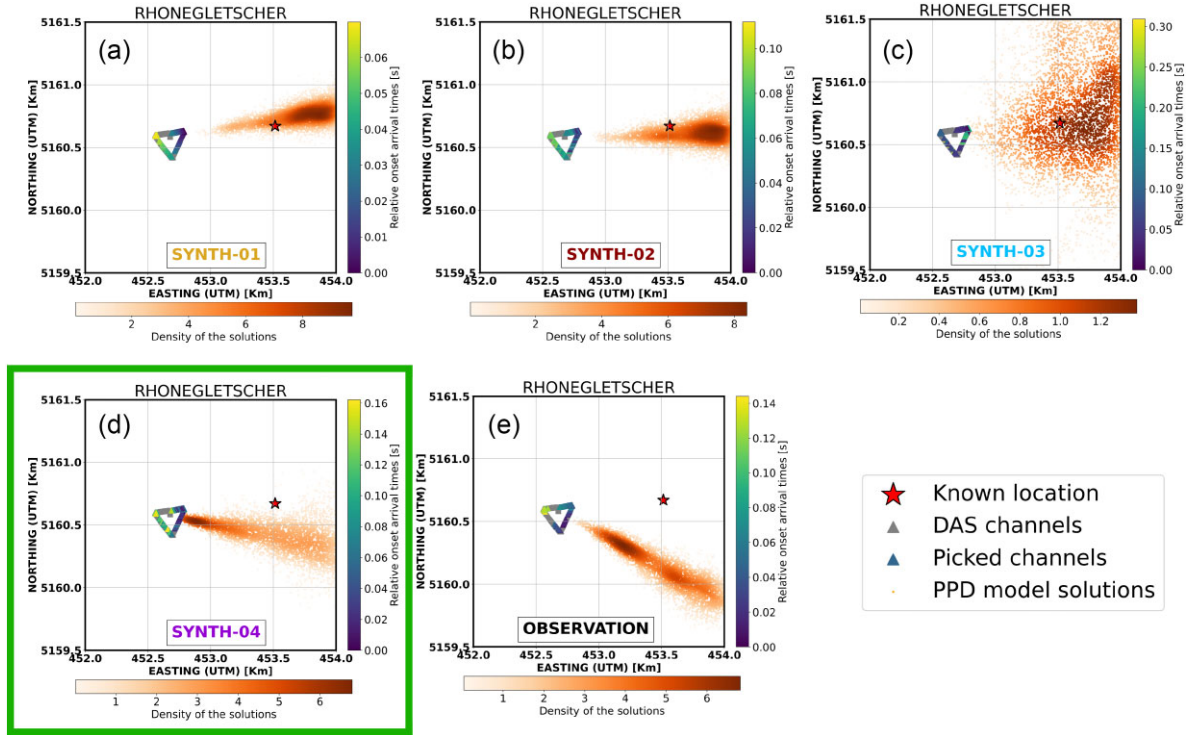
Figure 8. A complete example for the RHONEGLETSCHER DAS array, from location uncertainties with the four synthetic tests to the ones obtained with the observed data. The location uncertainty is illustrated by a set of samples approximating the PPDs (orange points coloured by density). The known location is marked as a red star. The DAS geometry is provided as grey triangles, while the triggered channels and the corresponding locations with colours depend on the relative P-wave arrival time. The green rectangle identifies the synthetic test (SYNTH-04) with the highest Total Score. (a) Event location uncertainties with inverted SYNTH-01 synthetic arrival times, (b) event location uncertainties with inverted SYNTH-02 synthetic arrival times, (c) event location uncertainties with inverted SYNTH-03 synthetic arrival times, (d) event location uncertainties with inverted SYNTH-04 synthetic arrival times and (e) event location uncertainties with inverted observed arrival times.
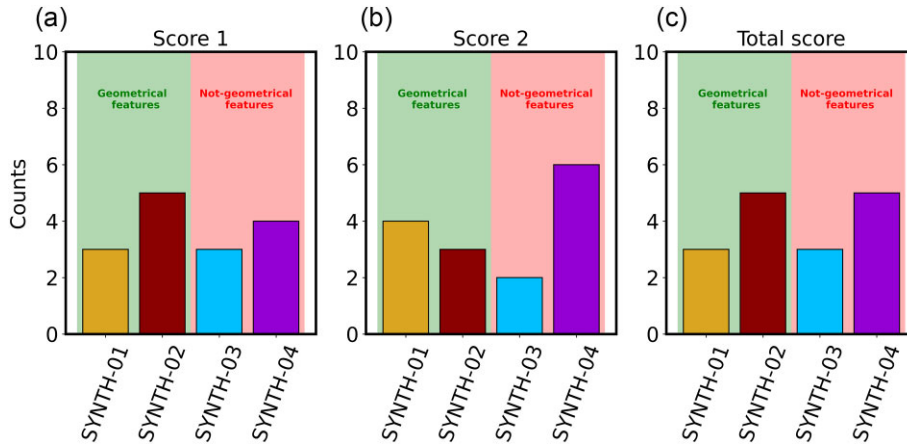

Figure 9. Statistics on the synthetic tests mimicking the observed PPDs. Green areas identify SYNTH-01 and SYNTH-02, which are based on noise sources related to the geometry of the problem (array + event). Red areas identify SYNTH-03 and SYNTH-04, which are based on noise sources unrelated to the geometry of the problem. (a) Score 1 histograms represent the similarity of synthetic PPDs to observed PPDs. (b) Score 2 histograms represent the distance between the highest values of synthetic and observed PPDs. (c) The total Score is the average of the normalized (0–1) Score 1 and Score 2.

Uncertainties in the known location cannot be excluded *a priori*, potentially accounting for the unresolved case studies.

Geometrical parameters, such as large azimuthal gap values in the analysed case studies, might partially explain the observed sensitivity to noise: when the azimuthal gap is lower (i.e. the geometry provides a higher constraint on the event location), PPDs vary less markedly (Figs S4 and S9). Results show how traditional geometrical indices or specially designed parameters, like AREA-PAR-1

and AREA-PAR-2, are correlated with the location uncertainties (SYNTH-01). Therefore, if a coherent ideal Gaussian noise is assumed as the only noise source in the DAS data, thus avoiding further complexities, the geometry of the problem remains a fundamental constraint that drives the distribution of possible solutions. (Fig. 10). This observation suggests continuing to evaluate (as in
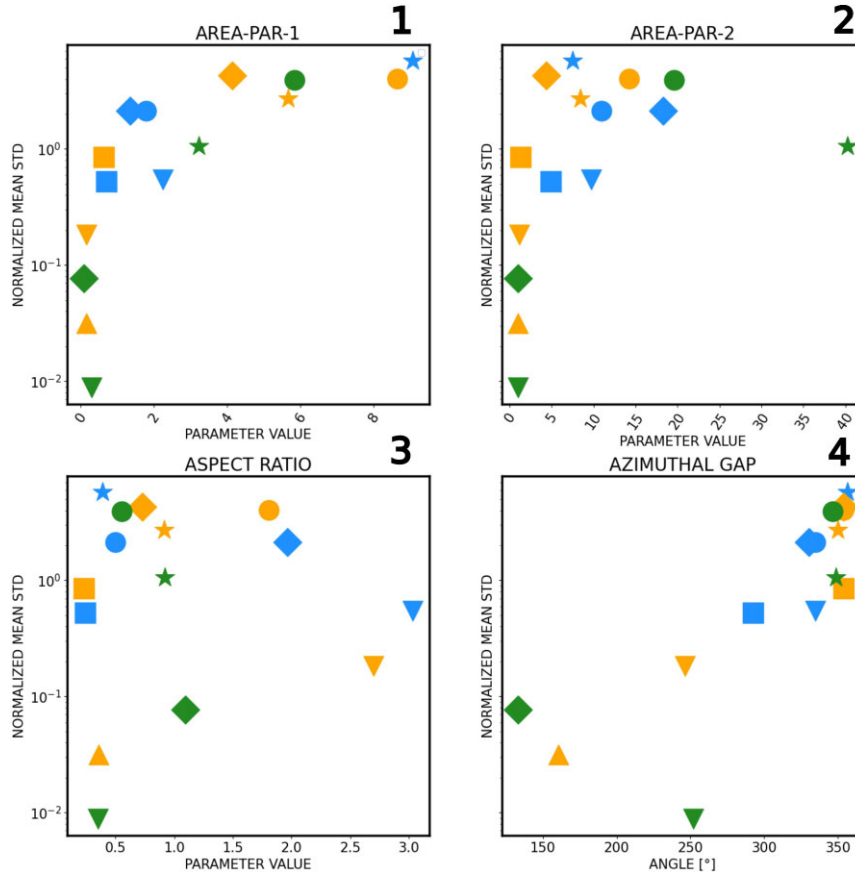
Figure 10. The geometrical parameters for each DAS array are compared with the normalized standard deviation from SYNTH-01. Colours and markers meanings (triangles, squares, etc.) are the same of Fig. 1. Positive correlations are present for AREA-PAR-1, AREA-PAR-2 and the azimuthal gap.

traditional networks) the geometry of the array before data acquisition, when possible, by considering the target source regions to be monitored (Toledo *et al.* 2020).

Although geometry matters, especially when arrival times can be assumed Gaussian-distributed, synthetic tests show that in real DAS systems, the uncertainty in the location process is strongly influenced by more complex traveltime statistics. Possible sources of noise, not directly correlated to the local geometry of the cable, are (i) axial directivity (Kennett 2022), (ii) site effects (Trabattoni *et al.* 2022) and (iii) cable coupling (Celli *et al.* 2024). Even though directivity is related to the angle of incidence of the wave front, near-surface distortions make its modelling extremely difficult, and not strictly related to cable azimuth (Bozzi *et al.* 2024). In fact, results here show how SYNTH-01 and SYNTH-02, thus the synthetic tests reproducing noise sources directly related to the geometry of the problem, are far from being the overall best-performing synthetic tests. Therefore, the error statistics associated with the automatic arrival times estimated in this work are often not adequately reproduced by assuming standard energy decay with distance and/or Gaussian-distributed noise with similar properties throughout the cable. Instead, SYNTH-03 and SYNTH-04, which model disturbances not strictly related to geometry, are able to reproduce observed location uncertainties in a significant number of case studies (Fig. 9). Specifically, the performance of SYNTH-03 suggests how phase mis-pick for the onset waveforms can be relevant in DAS real data, while SYNTH-04 suggests that the noise in P-wave arrival times is not purely Gaussian-distributed. Apart from the physical mechanisms related to the intrinsic DAS features (directivity, site

effects, and cable coupling), additional processes may explain the origin of SYNTH-04 noise distribution; optical noise (Zhong *et al.* 2021) or converted and guided waves (Lellouch *et al.* 2022) in near-cable structures can indeed be possible sources.

5 CONCLUSIONS

In this study, we model four noise sources on *P*-wave arrival times in DAS data using recordings from 15 different deployments and assess their influence on source location for events at local/regional distances. We show how, similar to traditional observational seismology, location uncertainties are influenced by the deployment's geometry and its position relative to the source region. Indeed, larger azimuthal gaps result in less constrained solutions. However, our results suggest the presence of other effects, more complex and not strictly dependent on the geometry of the problem (energy decay along the cable), affecting the noise statistics of the P-wave arrival times. We indeed show a significant impact of synthetic noise sources, especially the one mimicking phase mis-picks and not-purely Gaussian positive residuals, on event locations.

Synthetic tests suggest an *a priori* evaluation of noise statistics, together with the problem's geometry, for the quantification of precision/resolution of a novel DAS monitoring system. Typically, the design of earthquake monitoring infrastructures is based on the *a priori* evaluation of the geometry that minimizes the localization error for a given focal volume. In this process, when no experimental observations are available, noise is usually modelled using simplified assumptions. Our work demonstrates how applying a
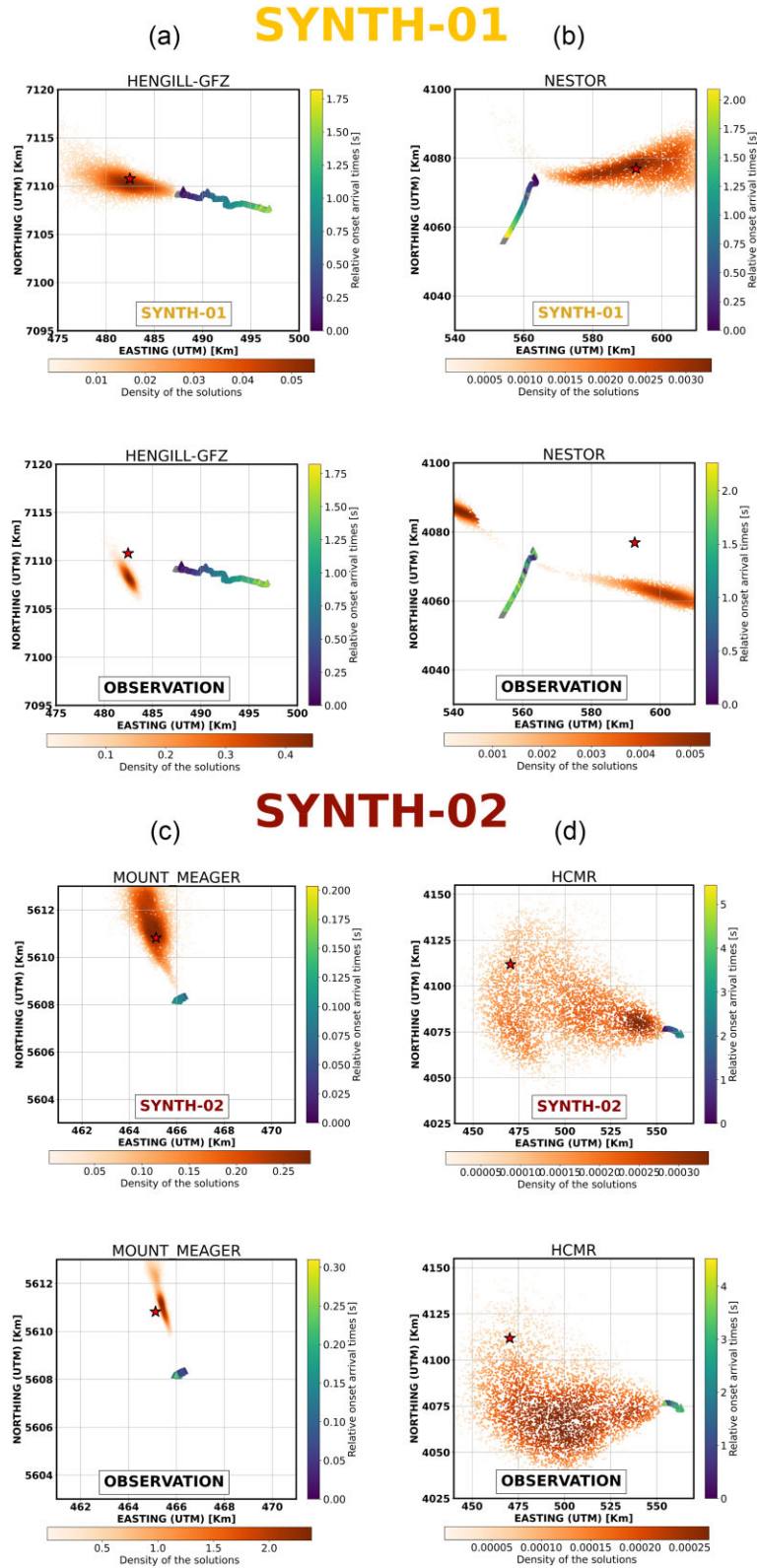
Figure 11. SYNTH-01 and SYNTH-02 (upper rows) are the synthetic tests that best mimic (highest Total Score) the ones retrieved from observed arrival times (bottom rows). (a) HENGILL-GFZ, (b) NESTOR, (c) MOUNT-MEAGER and (d) HCMR case studies.

similar evaluation to a DAS installation oversimplifies the problem.

Indeed, much more sophisticated noise models, accounting for intrinsic and peculiar elements of the DAS arrays not strictly related to geometrical factors, are required:
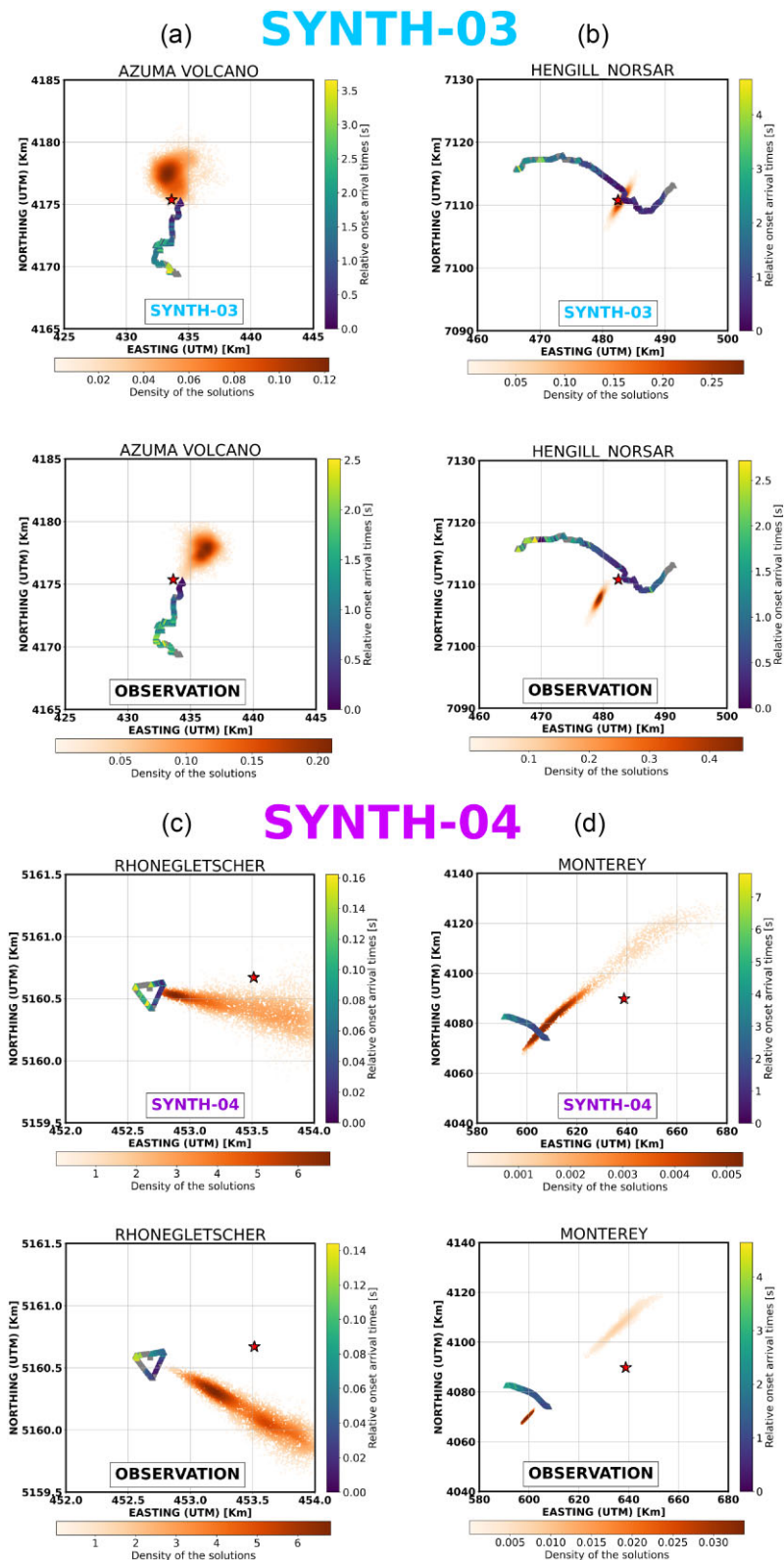
Figure 12. SYNTH-03 and SYNTH-04 (upper row) as the synthetic test mimicking at best (highest Total Score) the one retrieved from observed arrival times (bottom row). (a) AZUMA VOLCANO, (b) HENGILL-NORSAR, (c) RHONEGLETSCHER and (d) MONTEREY case studies.

(i) The first factor is the directional DAS sensitivity, for which the cable is only sensitive to the component of the ground motion polarized along the cable. Although this effect is theoretically related to geometry, surface geology introduces additional complications, limiting the capacity to model at the local scale the geometry of the wave front.

(ii) The second element is the marked sensitivity of the DAS measurements to the shallow geology. In fact, the amplitude of the strain rate scales with the slowness of the propagation medium, so the SNR of DAS recordings can vary significantly even over short distances.

(iii) Finally, the coupling of the FOC to the ground is a factor playing a primary role in controlling the quality of the recordings.

Such three attributes (and optical noise or guided waves along the cable) can individually or collectively lead to local changes in signal amplitudes, and, thus, impact measured phase arrival times, leading to complex noise statistics.

Overall, we show how event location based on P-wave arrival times estimated with DAS is affected by complex sources of uncertainty, not strictly related to geometry. These elements indicate the intrinsic difficulty of an *a priori* optimization of the geometry of DAS networks for earthquake monitoring and suggest an integrated study on noise statistics.

DATA AVAILABILITY STATEMENT

The following public repositories and a specific Zenodo repository store the analysed seismic events and/or associated metadata (last access 10-08-2023):

Link to the Zenodo repository: https://doi.org/10.5281/zenodo.8177033

(i) RHONEGLETSCHER
The event and geometry are provided in the Zenodo repository. The known location is obtained from manual picking of *P*-wave onsets at local seismometers, close to the DAS array.

(ii) MOUNT-MEAGER
The link to the repository is https://doi.org/10.5281/zenodo.4728303 and specifically navigate to `stack_20191001_090012.h5` for the event. The geometry is provided in the Zenodo repository. The known location is obtained from manually selected DAS channels.

(iii) POROTOMO
For the specific event navigate to DASH data, then 20160314 and :
`PoroTomo_iDAS16043_160314104148.h5,`
`PoroTomo_iDAS16043_160314104218.h5,`
`PoroTomo_iDAS16043_160314104248.h5)`
(https://gdr.openei.org/submissions/980.   The   geometry is   available   at   https://gdr.openei.org/submissions/980. The   known   location   is   reported   in   (Li & Zhan 2018),   specifically   in   the   supporting   material: that   is `M~ -0.074 catalog earthquake (event ID: 2201050).`

(iv) GRÍMSVÖTN
The event and geometry are provided in the Zenodo repository. The known location is obtained from manually selected DAS channels.

(v) STANFORD-1 and STANFORD-2
The events and geometries are provided by the authors under request.
The known locations can be found in the USGS catalogue, respectively: (https://earthquake.usgs.gov/earthquakes/eventpage/nc72751160/executive) (https://earthquake.usgs.gov/earthquakes/eventpage/nc73355340/executive).

(vi) FORESEE
For the specific event navigate to August data and then to   `PSUDAS_UTC_20190827_075804.511.tdms`:   https://www.datacommons.psu.edu/commonswizard/MetadataDisplay.aspx?Dataset=6290. The cable geometry is provided in the

Zenodo repository. An additional repository for the geometry: https://doi.org/10.1785/0220220279.
The known location can be found in the PASEIS network catalogue http://paseis.geosc.psu.edu/.

(vii) AZUMA-VOLCANO
The event and geometry are provided in the Zenodo repository. The known location is reported in (Nishimura *et al.* 2021).

(viii) HENGILL-NORSAR/GFZ
The events and geometries are provided in the Zenodo repository.
The   known   location   is   provided   by   the   authors: `event id, 23/07/2021 09:56:16; 204371.4668, 2585043.963, 3483.657209, 1.52, (EPSG:8088).`

(ix) HCMR-NESTOR-MEUST
For the events and the geometry navigate to earthquake waveforms in the following link: https://osf.io/4bjph/.
The known locations are reported in (Lior 2020), for the events   corresponding   to   this   origin   times,   respectively: `HCMR, 19/04/2019 03:30:19; NESTOR, 23/04/2019 19:25:51, MEUST, 21/07/2019 23:01:58.`

(x) MONTEREY
Navigate to the event and geometry using the following link: https://github.com/njlindsey/Photonic-seismology-in-Monterey-Bay-Dark-fiber1DAS-illuminates-offshore-faults-and-coastal-ocean.
The known location can be found at the same link.

(xi) CANARY
Navigate to `2020/08/01 03:40:44.03` for the event recording at the following link: https://www.fdsn.org/networks/detail/ZI_2020/. The known location is reported here: https://www.ign.es/web/resources/sismologia/www/dir_images_terremotos/fases/2020/es2020oznpu.dat.
The geometry of the cable is provided by the authors under request.

Details about the McMC algorithm can be found in (Riva *et al.* 2024).

The automatic picker is available as a Python package at this link: https://docs.obspy.org/master/packages/autogen/obspy.signal.trigger.pk_baer.html.

Figures have been prepared using PyGMT (https://www.pygmt.org/latest) and MatPlotLib (https://matplotlib.org/stable/index.html) packages.

SUPPORTING INFORMATION

Supplementary data are available at *GJIRAS* online.

suppl_data

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

REFERENCES

Ajo-Franklin, J.B. *et al.*, 2019. Distributed acoustic sensing using dark fiber for near-surface characterization and broadband seismic event detection, *Sci. Rep.,* **9**(1328), 1–14.

Baer, M. & Kradolfer, U., 1987. An automatic phase picker for local and teleseismic events, *Bull. seism. Soc. Am.,* **77**(4), 1437–1445.

Bakulin, A., Silvestrov, I. & Pevzner, R., 2020. Surface seismics with DAS: an emerging alternative to modern point-sensor acquisition, *Leading Edge,* **39**(11), 808–818.

Bayes, T. & Price, 1763. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S, *Phil. Trans. R. Soc. Lond.,* **53,** 370–418.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y. & Wassermann, J., 2010. Obspy: a python toolbox for seismology, *Seismol. Res. Lett.,* **81**(3), 530–533.

Biagioli, F., Métaxian, J.-P., Stutzmann, E., Ripepe, M., Bernard, P., Trabattoni, A., Longo, R. & Bouin, M.-P., 2024. Array analysis of seismo-volcanic activity with distributed acoustic sensing, *Geophys. J. Int.,* **236**(1), 607–620.

Biondi, B., Martin, E., Cole, S., Karrenbach, M. & Lindsey, N., 2017. Earthquakes analysis using data recorded by the Stanford DAS array, in *2017 SEG International Exposition and Annual Meeting,* OnePetro.

Biondi, B., Yuan, S., Martin, E.R., Huot, F. & Clapp, R.G., 2021. Using telecommunication fiber infrastructure for earthquake monitoring and near-surface characterization, in *Distributed Acoustic Sensing in Geophysics: Methods and Applications,* pp. 131–148, eds Li, Y., Karrenbach, M. & Ajo-Franklin, J. B., Geophysical Monograph Series, American Geophysical Union.

Bozzi, E. *et al.*, 2024. Complex spatial distribution of onset amplitude and waveform correlation: case studies from different DAS experiments, *Bull. Geophys. Oceanogr.,* **65**(2), 271–290.

Celli, N.L., Bean, C.J. & O'Brien, G.S., 2024. Full-waveform simulation of DAS records, response and cable-ground coupling, *Geophys. J. Int.,* **236**(1), 659–674.

Currenti, G. *et al.*, 2023. Distributed dynamic strain sensing of very long period and long period events on telecom fiber-optic cables at Vulcano, Italy, *Scientific Reports,* **13**(1), 4641.

Feigl, K., Taverna, N. & Rossol, M., 2016. Underwater DAS detection [dataset] .

Fichtner, A., Klaasen, S., Thrastarson, S., Çubuk-Sabuncu, Y., Paitz, P. & Jónsdóttir, K., 2022. Fiber-optic observation of volcanic tremor through floating ice sheet resonance, *Sismic Rec.,* **2**(3), 148–155.

Flóvenz, Ó.G. *et al.*, 2022. Cyclical geothermal unrest as a precursor to Iceland's 2021 Fagradalsfjall eruption, *Nat. Geosci.,* **15**(5), 397–404.

Hill, D., 2015. Distributed acoustic sensing (DAS): theory and applications, in *Frontiers in Optics ,* OSA Technical Digest (online) (Optica Publishing Group), Optica Publishing Group.

Hudson, T. S., Klaasen, S., Fontaine, O., Bacon, C., Jonsdottir, K. & Fichtner, A. (2024). Towards a widely applicable earthquake detection algorithm for fibreoptic and hybrid fibreoptic-seismometer networks, ArXiv preprint,.

Hudson, T.S. *et al.*, 2021. Distributed acoustic sensing (DAS) for natural microseismicity studies: a case study from Antarctica, *J. geophys. Res.,* **126**(7), e2020JB021493.

Isken, M.P., Vasyura-Bathke, H., Dahm, T. & Heimann, S., 2022. Denoising distributed acoustic sensing data using an adaptive frequency–wavenumber filter, *Geophys. J. Int.,* **231**(2), 944–949.

Jousset, P. *et al.*, 2018. Dynamic strain determination using fibre-optic cables allows imaging of seismological and structural features, *Nat. Commun.,* **9**(1), 1–11.

Karrenbach, M., Cole, S., Ridge, A., Boone, K., Kahn, D., Rich, J., Silver, K. & Langton, D., 2019. Fiber-optic distributed acoustic sensing of microseismicity, strain and temperature during hydraulic fracturing, *Geophysics,* **84**(1), D11–D23.

Kennett, B.L., 2022. The seismic wavefield as seen by distributed acoustic sensing arrays: local, regional and teleseismic sources, *Proc. R. Soc. A,* **478**(2258), 20210812.

Klaasen, S., 2021. Data repository for: distributed acoustic sensing in volcano-glacial environments—Mount Meager, British Columbia, *J. geophys. Res.,* **126**(11), doi:10.1029/2021JB022358.

Klaasen, S., Paitz, P., Lindner, N., Dettmer, J. & Fichtner, A., 2021. Distributed acoustic sensing in volcano-glacial environments: Mount meager, British Columbia, *J. geophys. Res.,* **126**(11).

Klaasen, S., Thrastarson, S., Çubuk-Sabuncu, Y., Jónsdóttir, K., Gebraad, L., Paitz, P. & Fichtner, A., 2023. Subglacial volcano monitoring with fibre-optic sensing: Grímsvötn, iceland, *Volcanica,* **6**(2), 301–311.

Klein, F.W., 2002. User's Guide to HYPOINVERSE-2000, a Fortran Program to Solve for Earthquake Locations and Magnitudes, 4/2002 version, Tech. rep., US Geological Survey.

Lay, T. & Wallace, T.C., 1995. *Modern Global Seismology,* Elsevier.

Lellouch, A. *et al.*, 2022. Microseismic analysis over a single horizontal distributed acoustic sensing fiber using guided waves, *Geophysics,* **87**(3), KS83–KS95.

Lellouch, A., Lindsey, N.J., Ellsworth, W.L. & Biondi, B.L., 2020. Comparison between distributed acoustic sensing and geophones: downhole microseismic monitoring of the forge geothermal experiment, *Seismol. Soc. Am.,* **91**(6), 3256–3268.

Lellouch, A., Yuan, S., Ellsworth, W.L. & Biondi, B., 2019. Velocity-based earthquake detection using downhole distributed acoustic sensing-examples from the San Andreas Fault Observatory at depth velocity-based earthquake detection using downhole distributed acoustic sensing, *Bull. seism. Soc. Am.,* **109**(6), 2491–2500.

Li, Z. & Zhan, Z., 2018. Pushing the limit of earthquake detection with distributed acoustic sensing and template matching: a case study at the Brady geothermal field, *Geophys. J. Int.,* **215**(3), 1583–1593.

Li, Z., Shen, Z., Yang, Y., Williams, E., Wang, X. & Zhan, Z., 2021. Rapid response to the 2019 Ridgecrest earthquake with distributed acoustic sensing, *AGU Adv.,* **2**(2), e2021AV000395.

Lindsey, N.J. & Martin, E.R., 2021. Fiber-optic seismology, *Annu. Rev. Earth planet. Sci.,* **49,** 309–336.

Lindsey, N.J., Martin, E.R., Dreger, D.S., Freifeld, B., Cole, S., James, S.R., Biondi, B.L. & Ajo-Franklin, J.B., 2017. Fiber-optic network observations of earthquake wavefields, *Geophys. Res. Lett.,* **44**(23), 11–792.

Lindsey, N.J., Rademacher, H. & Ajo-Franklin, J.B., 2020. On the broadband instrument response of fiber-optic DAS arrays, *J. geophys. Re.,* **125**(2), e2019JB018145.

Lior, I., 2020. Underwater DAS detection [dataset] .

Mestayer, J. *et al.*, 2011. Field trials of distributed acoustic sensing for geophysical monitoring, in *SEG Technical Program Expanded Abstracts 2011,* pp. 4253–4257, Society of Exploration Geophysicists.

Molenaar, M.M., Hill, D., Webster, P., Fidan, E. & Birch, B., 2011. First downhole application of distributed acoustic sensing (DAS) for hydraulic fracturing monitoring and diagnostics, in *SPE Hydraulic Fracturing Technology Conference and Exhibition,* SPE–140561, SPE.

Mosegaard, K. & Tarantola, A., 2002. Probabilistic approach to inverse problems, *Int. Geophys. Ser.,* **81**(A), 237–268.

Nayak, A. & Ajo-Franklin, J., 2021. Distributed acoustic sensing using dark fiber for array detection of regional earthquakes, *Seismol. Res. Lett.,* **92**(4), 2441–2452.

Nishimura, T., Emoto, K., Nakahara, H., Miura, S., Yamamoto, M., Sugimura, S., Ishikawa, A. & Kimura, T., 2021. Source location of volcanic earthquakes and subsurface characterization using fiber-optic cable and distributed acoustic sensing system, *Sci. Rep.,* **11**(1), 1–12.

Obermann, A. *et al.*, 2022. Combined large-n seismic arrays and DAS fiber optic cables across the Hengill Geothermal Field, Iceland, *Seismol. Soc. Am.,* **93**(5), 2498–2514.

Paitz, P., Edme, P., Gräff, D., Walter, F., Doetsch, J., Chalari, A., Schmelzbach, C. & Fichtner, A., 2021. Empirical investigations of the instrument response for distributed acoustic sensing (DAS) across 17 octaves, *Bull. seism. Soc. Am.,* **111**(1), 1–10.

Parker, T., Shatalin, S. & Farhadiroushan, M., 2014. Distributed acoustic sensing–a new tool for seismic applications, *First Break,* **32**(2).

Piana Agostinetti, N., Villa, A. & Saccorotti, G., 2022. Distributed acoustic sensing as a tool for subsurface mapping and seismic event monitoring: a proof of concept, *Solid Earth,* **13**(2), 449–468.

Porras, J., Pecci, D., Bocchini, G. M., Gaviano, S., De Solda, M., Tuinstra, K., Lanza, F., Tognarelli, A., Stucchi, E. & Grigoli, F., 2024. A semblance-based microseismic event detector for DAS data, Geophysical Journal International, **236**, 3 1716–1727

Riva, F., Agostinetti, N.P., Marzorati, S. & Horan, C., 2024. The microseismicity of Co. Donegal (Ireland): defining baseline seismicity in a region of slow lithospheric deformation, *Terra Nova,* **36**(1), 62–76.

Spica, Z.J. *et al.*, 2023. Pubdas: a public distributed acoustic sensing datasets repository for geosciences, *Seismol. Soc. Am.,* **94**(2A), 983–998.

Toledo, T., Jousset, P., Maurer, H. & Krawczyk, C., 2020. Optimized experimental network design for earthquake location problems: applications to geothermal and volcanic field seismic networks, *J. Volc. Geotherm. Res.,* **391**, 106433.

Trabattoni, A., Festa, G., Longo, R., Bernard, P., Plantier, G., Zollo, A. & Strollo, A., 2022. Microseismicity monitoring and site characterization with distributed acoustic sensing (DAS): the case of the Irpinia Fault System (Southern Italy), *J. geophys. Res.,* **127**(9), e2022JB024529.

Ugalde, A., Becerril, C., Villasenor, A., Ranero, C.R., Fernández-Ruiz, M.R., Martin-Lopez, S., González-Herráez, M. & Martins, H.F., 2022. Noise levels and signals observed on submarine fibers in the Canary Islands using DAS, *Seismol. Soc. Am.,* **93**(1), 351–363.

van den Ende, M. & Ampuero, J.-P., 2021. Evaluating seismic beamforming capabilities of distributed acoustic sensing arrays, *Solid Earth,* **12**(4), 915–934.

Villasenor, A., Ugalde, A., Gonzalez Herraez, M., Martins, H.F. & Cosin, M., 2020. Canary Islands seismic monitoring with Distributed Acoustic Sensing [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/ZI_2020.

Walter, F., Gräff, D., Lindner, F., Paitz, P., Köpfli, M., Chmiel, M. & Fichtner, A., 2020. Distributed acoustic sensing of microseismic sources and wave propagation in glaciated terrain, *Nat. Commun.,* **11**(1), 1–10.

Yang, Y., Atterholt, J.W., Shen, Z., Muir, J.B., Williams, E.F. & Zhan, Z., 2022. Sub-kilometer correlation between near-surface structure and ground motion measured with distributed acoustic sensing, *Geophys. Res. Lett.,* **49**(1), e2021GL096503.

Zhan, Z., 2020. Distributed acoustic sensing turns fiber-optic cables into sensitive seismic antennas, *Seismol. Res. Lett.,* **91**(1), 1–15.

Zhong, T., Cheng, M., Lu, S., Dong, X. & Li, Y., 2021. RCEN: a deep-learning-based background noise suppression method for DAS-Vsp records, *IEEE Geosci. Remote Sens. Lett.,* **19**, 1–5.

Zhu, T., Shen, J. & Martin, E.R., 2021. Sensing earth and environment dynamics by telecommunication fiber-optic sensors: an urban experiment in Pennsylvania, USA, *Solid Earth,* **12**(1), 219–235.

Zhu, W., Biondi, E., Li, J., Yin, J., Ross, Z.E. & Zhan, Z., 2023. Seismic arrival-time picking on distributed acoustic sensing data using semi-supervised learning, *Nat. Commun.,* **14**(8192).