Supporting Information for "High resolution seafloor thermometry and internal wave monitoring using Distributed Acoustic Sensing"

Julián Pelaez Quiñones[1], Anthony Sladen[1], Aurelien Ponte[2], Itzhak Lior[3],

Jean-Paul Ampuero[1], Diane Rivet[1], Samuel Meulé[4], Frédéric Bouchette[4],

Ivane Pairaud[2], Paschal Coyle[5]

[1]Université Côte d'Azur, CNRS, Observatoire de la Côte d'Azur, IRD, Géoazur, Sophia Antipolis, *250* rue Albert Einstein, *06560*, Valbonne, France

[2]IFREMER, Université de Brest, CNRS, IRD, Laboratoire d'Océanographie Physique et Spatiale, IUEM, Brest, France

[3]Institute of Earth Sciences, The Hebrew University, Jerusalem, Israel

[4]Geosciences-M/GLADYS, Université de Montpellier, CNRS, Montpellier, France

[5]Aix-Marseille Université, CNRS/IN2P3, CPPM, Marseille, France

Contents of this file

Introduction

This file contains complementary information to our main manuscript, principally details about the sensing instruments, methods, processing and some additional figures.

Text S1. Principle of Distributed Acoustic Sensing

Distributed Acoustic Sensing (DAS) systems make use of single optic fibers cased inside (un)armored cables, for instance existing Telecommunication cables, to sense the environment. So far, DAS systems require a dark fiber to operate. Coherent laser pulses are regularly sent along the fiber and their Rayleigh back-scattered signature is used as a proxy for temperature and strain perturbations affecting the optical path length (due to local elongations and refractive index variations of the fiber) over specific sections of the cable, which can be localized (López-Higuera, 2002; Hartog, 2017). These perturbations are traced-back along the fiber by converting the two-way travel time of light to distances with the known speed of light in silica. Measurements are averaged along a few meters of cable (gauge length) at a defined distance step (spatial sampling). In contrast to DAS, Distributed Temperature Sensing (DTS) is based on Raman-scattering, while Distributed Temperature and Strain Sensing (DSTS) is based on Brillouin-scattering.

The DAS interrogator unit used for our main analysis is an ϕ-OTDR hDAS (High fidelity distributed acoustic sensor) designed by Aragón Photonics, which provides measurements in strain units. One specificity of the hDAS system is the fact that it sends a chirped light signal. Details can be found in (Pastor-Graells et al., 2016; Fernández-Ruiz et al., 2019). The sampling frequency was 100 Hz in the first couple days of the campaign and then switched to 500 Hz.

The DSTS system used to validate the simultaneous LF-DAS (indirect) measurements was a Febus Optics G1-C set to record with a gauge length of 10 m and sampling resolution of 2.0 m over 30 km. The temporal sampling was set to 15 min to keep the data noise level at a reasonable level. The DAS system in this case was a Febus A1-R DAS interrogator with gauge length of 10 m and sampling resolution of 4.8 m over 40 km of cable.

Text S2. Extracting the low-frequency component of DAS data

Because of the high sampling rates and large DAS data volumes acquired, a conventional low-pass filtering approach was not possible to isolate the low-frequency content of the raw data. Thus, a parallel-computing approach with a moving average was instead implemented for efficiency in the reduction of the thousands of channels

We implemented a moving average windows of 5 minutes with 60% overlap independently to each channel. This implies an output sampling frequency of $\sim$8.33 mHz and a maximum resolvable frequency of $\sim$1.66 mHz (the latter is the inverse of twice the averaging window and does not necessarily match the Nyquist-criterion frequency that would be expected from the data point sampling rate). Our experience with different windows showed this combination to be a good compromise between a smoothing that is not excessive as to preserve the LF content, while being enough to remove spikes, high frequency noise, and to reduce the data size by a considerable proportion

The original data acquired is automatically segmented in sections of several days due to a laser refreshing procedure of the interrogator. Each segment has different trends, large value offsets and most of the times gaps in between. We demean the first segment and adjust the remaining segments with respect to the last value of the previous ones to

ensure continuity between them and to smooth-out large data breaks. This is performed for each channel separately. Although some of the consecutive segments show different trends which are likely related to instrumental drift, we did not correct these to avoid distorting and losing true signal, since an objective instrumental drift correction function is unknown to us. The data gaps in the signal were filled using cubic interpolation between segments. This allows for processing routines that require continuous time series (spectral decomposition and filtering). In this exploratory stage, we do not filter out "bad quality" channels, given that a criteria to define their "usefulness" (which may or may not be related to ground-seabed coupling) is not yet completely understood. A last pre-processing step is to remove the along channel mean amplitude temporal fluctuation from each sample of the data (DAS temporal response or common-noise correction) using a band of channels around a central channel to find each average. This procedure provides smoother time series, while the effect of the laser time fluctuations and strong amplitude spikes/steps is minimized. The data was highpass-filtered at 0.009 mHz prior to frequency-wavenumber transformation using a 2D Direct Fourier Transform.

Text S3. Conversion of strain to temperature

As outlined in Ide, Araki, and Matsumoto (2021), at long time scales (low frequencies), the apparent strain differences are expected to be caused by refractive index variations of the fiber due to temperature changes in the environment, instead of being caused by LF strain-related elongations on the fiber, since such LF strains could hardly couple energy into the fiber and their effect is much smaller in magnitude than the temperature effect. The formula that approximately describes this variations is:

$$\frac{\mathrm{d}\epsilon}{\mathrm{d}T} = n\alpha + \frac{\mathrm{d}n}{\mathrm{d}T}$$

where ϵ, T, n and α represent the observed (apparent) strain, the environment's temperature, silica's refractive index (typically around $7 \cdot 10^{-6}$ K^{-1} at room temperature) and its thermal expansion coefficient, respectively. The authors explain that a typical value for $\mathrm{d}n/\mathrm{d}T$ is 10^{-5} (constant) while the $n\alpha$ term is expected to be much smaller, in the order of 10^{-7}. Under these assumptions, a $\Delta s = 1$ nanostrain difference is approximately equivalent to $\Delta T = 0.1$ mK.

An absolute difference-normalization of each separate LF-DAS channel, i.e. between zero and the maximum value of each channel, is applied before conversion to temperature differences.

For the comparison of LF-DAS with the thermistor chain in Fig. 2, the best-matching cable channel was found via cross-correlation maxima search. The maximum correlations were found with the deepest, 50 m deep, temperature sensor of Cap Vieux, which is almost touching the seafloor and better replicates the FO cable configuration. We note, however, that the maximum normalized correlations have spread maxima at roughly 60%, i.e. the highest correlations near zero-lag were similar over a range of a few tens of channels; this result is expected given that both sensors are not collocated but separated by a few kilometers. The best-matching LF-DAS channel is located $\sim$4 km away from the thermistor chain.

As outlined in the main text, the FO inside the Toulon cable is relatively loose and can creep inside the cable when deformed slowly, at very low-frequencies. When rapidly

deformed by e.g. high-frequency seismic or acoustic waves, it responds proportionally to the stresses without creeping. This further contributes to explain why at LF, the effect of temperature is dominant whereas strains appears negligible.

Text S4. Inertial variability

The inertial period T_c at a given latitude θ reflects the variability time scale of important mesoscale to large scale oceanographic processes. This period is determined by the Coriolis effect due to the angular momentum conservation for traveling objects that are subjected to the earth's rotation centrifugal force and can be estimated via:

$$\mathrm{T}_c = {f_c}^{-1} = (2\Omega \sin\theta)^{-1}$$

where f_c is known as the Coriolis frequency and Ω is the rotation rate of the earth ($\sim 7.29 \times 10^{-5}$ rad/s). This translates into an inertial period of $\sim$17.5h at the mean latitude of the Toulon cable (43°N).

Text S5. Empirical Mode Decomposition and Hilbert-Huang Transform Parameters

EMD and HHT analyses (Huang et al., 1998) were performed by using the EMD Python package developed by Quinn, Lopes-dos Santos, Dupret, Nobre, and Woolrich (2021). Several of the examples and built-in functions of the package were implemented in our study. The mask sifting (mode separation) scheme (Deering & Kaiser, 2005) produced the best results for the LF-DAS data. This approach allows us to obtain a set of well-behaved Intrinsic Mode Functions (IMFs) that represent generalized spectral components or empirical modes of the input signal. A proper sifting leads to IMFs that are purely oscillatory functions with zero reference levels from which instantaneous amplitude and frequency

attributes are obtained by means of a Hilbert-Huang Transform (HHT) (Huang & Wu, 2008). The masks are monochromatic signals introduced into the Intrinsic Mode Function (IMF) under consideration to avoid mixing of modes with very different frequencies: as the high frequency components are always captured and separated first during the sifting, a mask signal with a frequency higher than a long period oscillation in the signal contributes to separate the latter correctly from the other higher frequency components. Most of the default mask sifting parameters of the package were the basis of our processing. The amplitude of these masks were uniformly computed as ratios of the standard deviation of the input for all IMFs; their frequency successively increasing at factors of 2. Four masks were applied to each IMF and the sift threshold was set to 10^{-8}. Eight IMFs were calculated in total.

The instantaneous attributes (amplitude and frequency) of each IMF were found via amplitude-normalized Hilbert transform (NHT) as in (Huang et al., 2009). Channels with anomalous extrema were muted under a 3-standard deviation outlier criterium. We applied a logarithmic binning of 1000 grid points between 0.001 and 1.0 mHz to ensure enough spectral resolution. Amplitudes were stacked to obtain the binned HHT. The HHT spectra were normalized as power spectral density (divided by $f_{sampling} \cdot N_{samples}$). To obtain the HHT spectra, we averaged all the instantaneous attributes of each IMF over a selected range of channels. This results in a stacked spectrogram-like output representing the dominant spectral power spectral density over a section of cable. The LF-DAS time series were pre-filtered with a highpass at 0.0007 mHz (equivalent to nearly 16 days - the total duration of the deployment) and pre-averaged every two consecutive channels to

increase their SNR. The final images were smoothed using a Gaussian kernel convolution filter with one standard deviation. For Figs 3b,c, each IMF is weighted by its instantaneous amplitude, so to obtain an image analogous to a spectrogram that captures the time-evolution of the spectral components.

Care was taken to select a timespan for analysis with no large data breaks and to reject channels with anomalously uniform or large values or spikes (as seen from Fig. 2a,b), as these artifacts can largely affect the EMD (Stallone et al., 2020). Furthermore, the averaging of the instantaneous attributes of each IMFs across a sufficiently long cable range helps to balance out such undesired effects, in case that artifacts may remain at some channels. Supplementary Figure S1 shows an example of such decomposition for a selected channel using the EMD Python package.

References

Deering, R., & Kaiser, J. (2005). The use of a masking signal to improve empirical mode decomposition. In *Proceedings. (icassp '05). ieee international conference on acoustics, speech, and signal processing, 2005.* (Vol. 4, p. iv/485-iv/488 Vol. 4). doi: 10.1109/ICASSP.2005.1416051

Fernández-Ruiz, M. R., Costa, L., & Martins, H. F. (2019). Distributed acoustic sensing using chirped-pulse phase-sensitive otdr technology. *Sensors*, *19*(20), 4368. doi: 10.3390/s19204368

Hartog, A. (2017). *An Introduction to Distributed Optical Fibre Sensors* (1st ed.). CRC Press. doi: 10.1201/9781315119014

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., ... et al.

(1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A.*, *454*, 903–995. doi: 10.1038/s41598-020-72193-2

Huang, N. E., & Wu, Z. (2008). A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, *46*(2). doi: 10.1029/2007RG000228

Huang, N. E., Wu, Z., Long, S. R., Arnold, K. C., Chen, X., & Blank, K. (2009). On instantaneous frequency. *Advances in Adaptive Data Analysis*, *01*(02), 177-229. doi: 10.1142/S1793536909000096

Ide, S., Araki, E., & Matsumoto, H. (2021). Very broadband strain-rate measurements along a submarine fiber-optic cable off cape muroto, nankai subduction zone, japan. *Earth Planets Space*, *73*(63). doi: 10.1186/s40623-021-01385-5

López-Higuera, J. M. (Ed.). (2002). *Handbook of Optical Fibre Sensing Technology* (1st ed.). John Wiley & Sons, ltd.

Pastor-Graells, J., Martins, H., Garcia-Ruiz, A., Martin-Lopez, S., & Gonzalez-Herraez, M. (2016). Single-shot distributed temperature and strain tracking using direct detection phase-sensitive otdr with chirped pulses. *Optics express*, *24*(12), 13121–13133. doi: 10.1364/OE.24.013121

Quinn, A. J., Lopes-dos Santos, V., Dupret, D., Nobre, A. C., & Woolrich, M. W. (2021). EMD: Empirical Mode Decomposition and Hilbert-Huang Spectral Analyses in Python. *Journal of Open Source Software*, *6*(59), 2977. doi: 10.21105/joss.02977

Stallone, A., Cicone, A., & Materassi, M. (2020). New insights and best practices for

the successful use of empirical mode decomposition, iterative filtering and derived algorithms. *Sci Rep*, *10*(15161). doi: 10.1038/s41598-020-72193-2

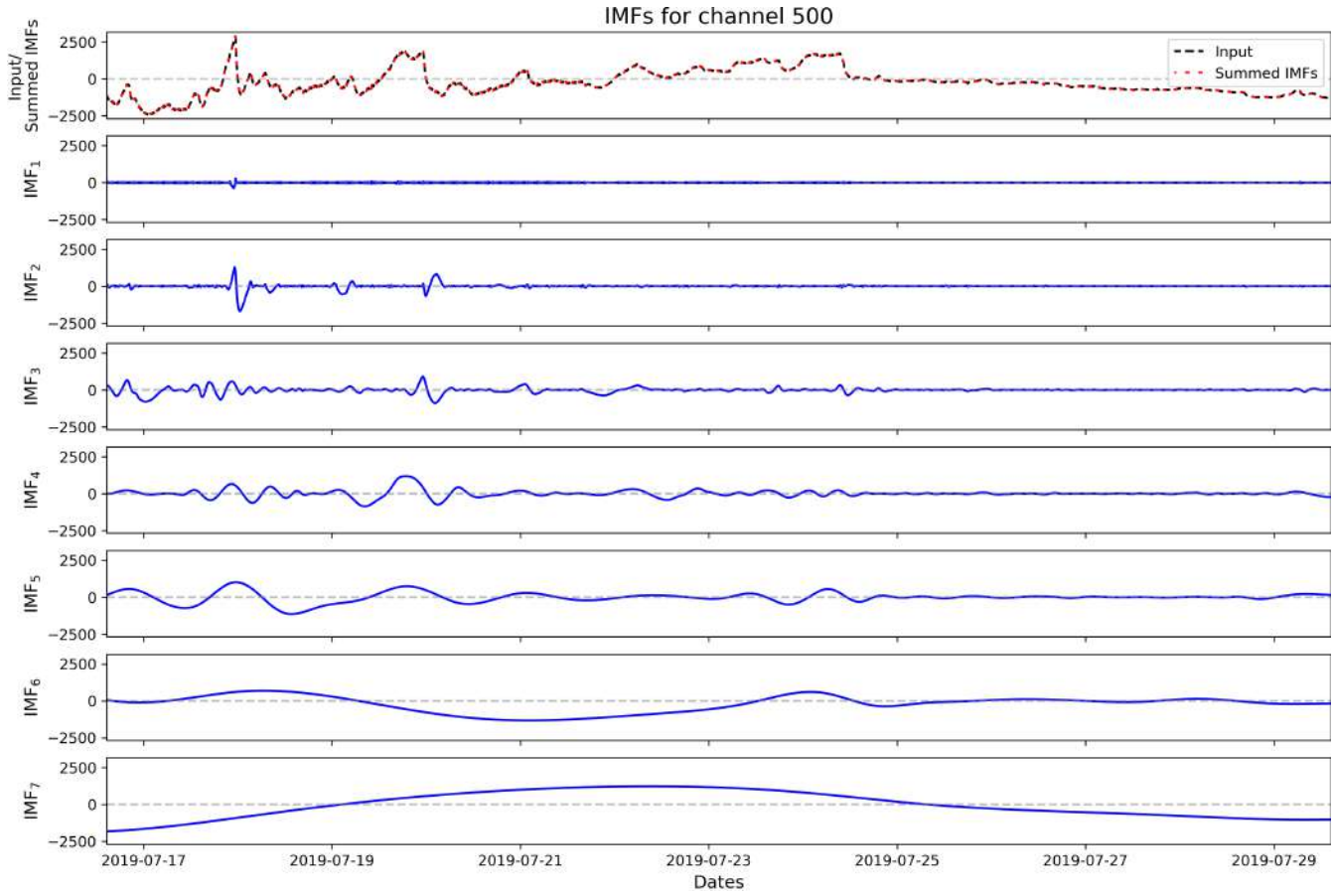


Figure S1. Sample Intrinsic Mode Functions (IMFs) for a selected LF-DAS channel.

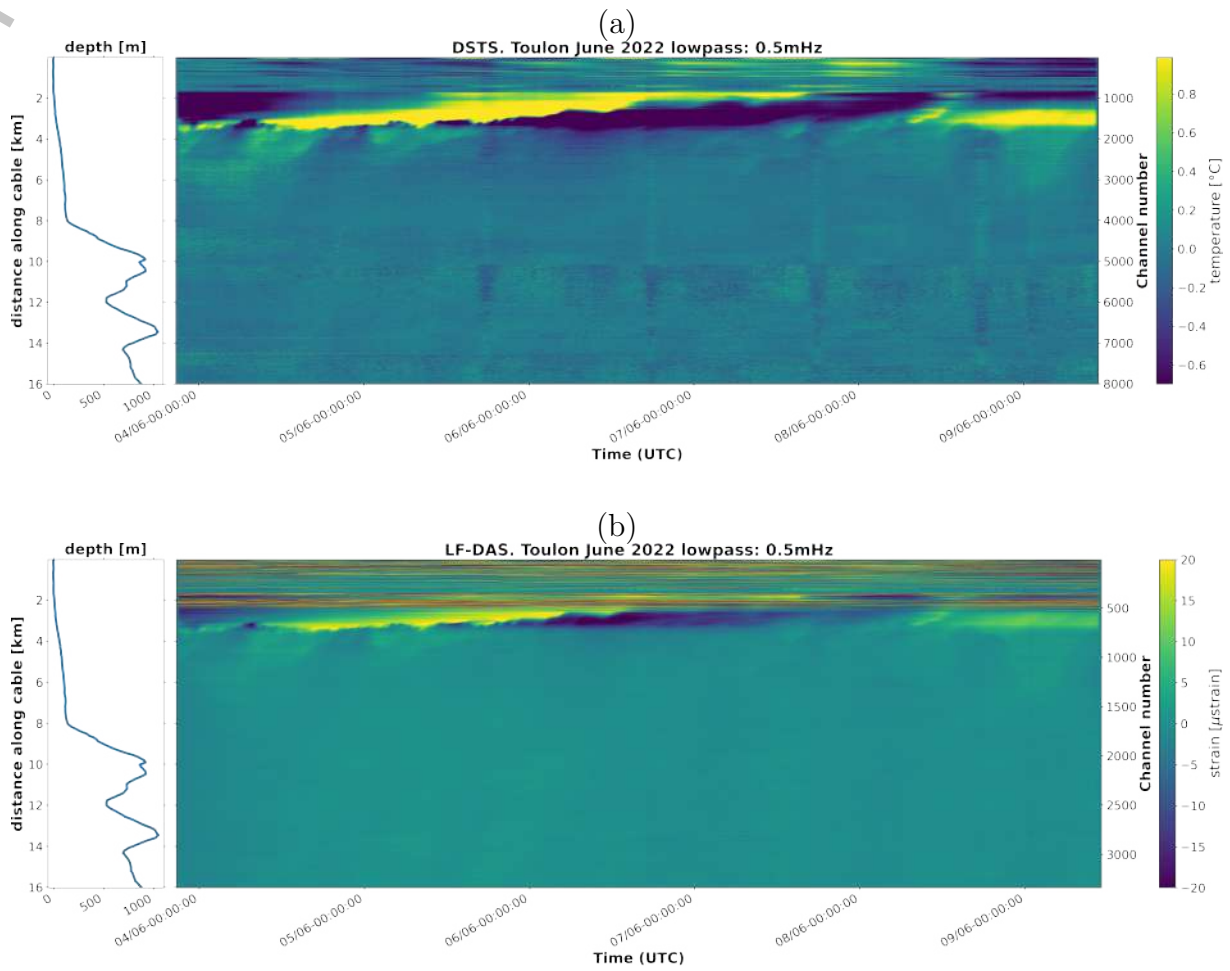
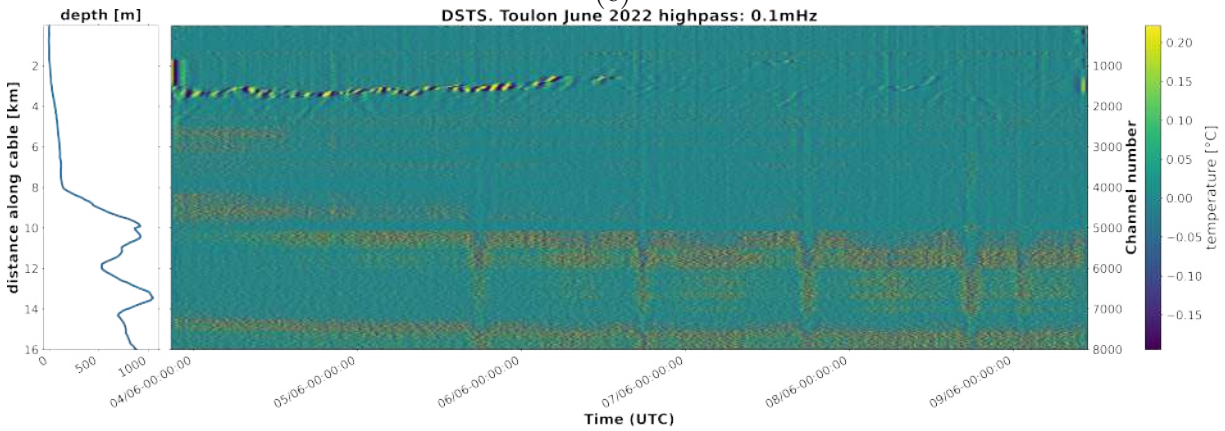


Figure S2. Collocated DSTS and LF-DAS measurements in Toulon, June 2022 - Filtered ensemble comparison. Lowpassed DSTS (a) and LF-DAS (b)

(c)



(d)

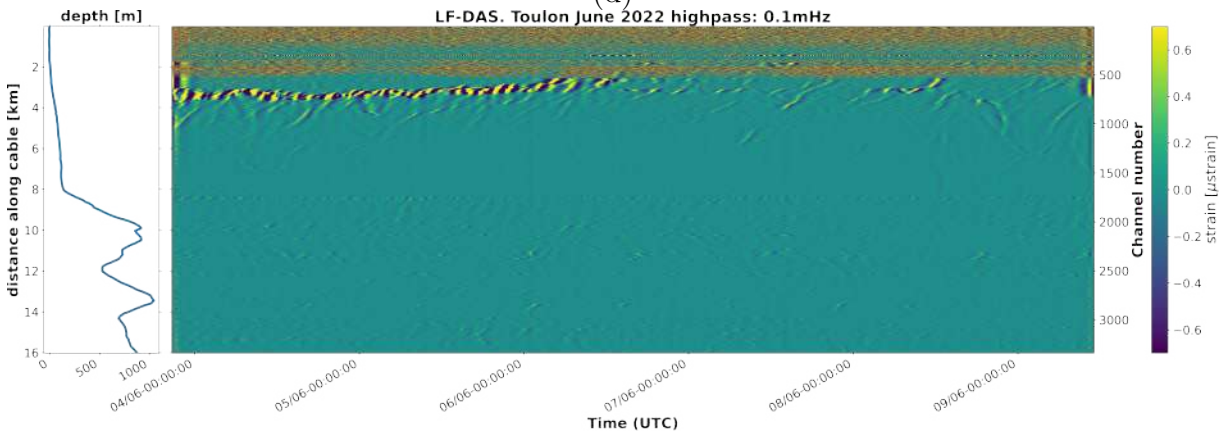


Figure S2. (cont.) collocated DSTS and LF-DAS measurements in Toulon, June 2022 - Filtered ensemble comparison. Highpassed DSTS (c) and LF-DAS (d).