https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

Advancing Turbulence Essential Ocean Variable: A Reference Glider-Based Microstructure Dataset from the Western Mediterranean

Florian V. M. Kokoszka[1], Mireno Borghini[2], Katrin Schroeder[3], Jacopo Chiggiato[3], Joaquín Tintoré[4], Nikolaos Zarokanellos[4], Albert Miralles[4], Patricia Rivera Rodríguez[4], Manuel Rubio[4], Miguel Charcos[4], Benjamín Casas[4], and Anneke ten Doeschate[5]

[1]Institute of Marine Science (ISMAR), Department of Earth System Sciences and Technologies for the Environment, National Research Council (CNR), Naples, Italy
[2]Institute of Marine Science (ISMAR), Department of Earth System Sciences and Technologies for the Environment, National Research Council (CNR), Lerici, Italy
[3]Institute of Marine Science (ISMAR), Department of Earth System Sciences and Technologies for the Environment, National Research Council (CNR), Venezia, Italy
[4]SOCIB, Balearic Islands Coastal Observing and Forecasting System, Palma, Spain
[5]Rockland Scientific, Victoria, B.C., Canada

Correspondence to: Florian V. M. Kokoszka (florianvolmermartin.kokoszka@cnr.it)

Abstract. We present a comprehensive dataset of turbulence microstructure measurements collected with a Micro Rider (MR-1000) from Rockland Scientific (RS) mounted on the Slocum Deep Glider "Teresa" across repeated transects between Sardinia and the Balearic Islands (SMART missions, 2015–2024). This dataset constitutes one of the most extensive autonomous glider-based microstructure archives to date for the Western Mediterranean, containing glider sections up to 1000m-depth and delivering quality-controlled vertical profiles of turbulent kinetic energy dissipation rate (ε) and thermal variance dissipation rate (χ) across seasonal cycles and diverse water masses. The data were processed through a rigorous multilevel workflow (L0–L4), following community best practices for processing, quality control, and uncertainty quantification. Final products include estimates of ε from dual shear probes and χ from dual fast thermistor probes, aligned with co-located hydrographic and oxygen measurements. This dataset provides a high-resolution resource for investigating fine-scale mixing, validating parameterizations, improving turbulence representation in models, and modeling physical processes. All data and processing codes are openly provided to support reuse, reproducibility, and integration into global efforts advancing the inclusion of turbulence as an Essential Ocean Variable.

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

1 Introduction

Starting from 2015, CNR-ISMAR in collaboration with SOCIB set up a recurrent Slocum Deep Glider G2 mission along a longitudinal transect between the Sardinia (Italy) and the Balearic Islands (Spain), in the western Mediterranean Sea, called SMART (Sardinia MAllorca Repeated Transect). With the aim of monitoring water masses changes over the recent years and integrating the existing distributed multiplatform observing system in the Western Mediterranean Sea, the transect is also included in the Ocean Glider Program (Testor et al., 2019). Several water masses are present in the study area which allowed us to characterize their temporal and spatial variability. In the upper layer, the Atlantic Water (AW) is present, which interplays with the Mediterranean surface waters, while at intermediate layers there is the presence of the Eastern Intermediate water (EIW), and the operation depth of the glider down to 1000m allows to capture partially the upper part of the Western Mediterranean Deep Water (WMDW) (acronyms follow Schroeder et al. 2024**)**. Such repeated missions are designed to characterize water mass properties and mixing/turbulence levels during different seasons and on the interannual scale.

In addition to the classical "conductivity-temperature-depth" (CTD) package, high-precision turbulence measurements are obtained through shear sensors and high-frequency thermistors installed on the Micro Rider (MR) from Rockland Scientific (RS). Over the past decade, the use of turbulence microstructure sensors mounted on autonomous platforms has significantly expanded the observational capacity of oceanographers to measure small-scale mixing processes, through enabled routine, high-resolution measurements of turbulent kinetic energy dissipation rate (ε, $W\,kg^{-1}$) from shear sensor, and thermal variance dissipation rate (χ, $^{o}C^{2}\,s^{-1}$) from fast response thermistor sensors, over long-duration missions and across dynamic ocean regions (Eriksen et al. 2001, Wolk et al. 2009, Peterson and Fer 2014, St Laurent and Merrifield 2017). While earlier deployments focused on pilot missions or single-process studies, few long-term, multi-season datasets from gliders exist, especially in the Mediterranean Sea environments. The present work builds upon these foundations and significantly extends them by providing one of the most extensive glider-based microstructure datasets to date for the Western Mediterranean. Collected from repeated transects between Sardinia and the Balearic Islands among nearly a decade, this dataset uniquely resolves ε and χ across key water masses and seasons (Kokoszka et al., this dataset), complementing ship-based efforts and contributing to the broader goals of initiatives such as ATOMIX (Fer et al. 2024). The inclusion of processed data together with open-source processing code and rigorous quality control, ensures transparency, reusability, and relevance to multiple disciplines. This dataset thus represents a significant step forward in establishing turbulence from pilot (Le Boyer et al. 2023), to operational Essential Ocean Variable (EOV), addressing a long-standing observational gap and offering a benchmark for future observational and modeling studies.

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

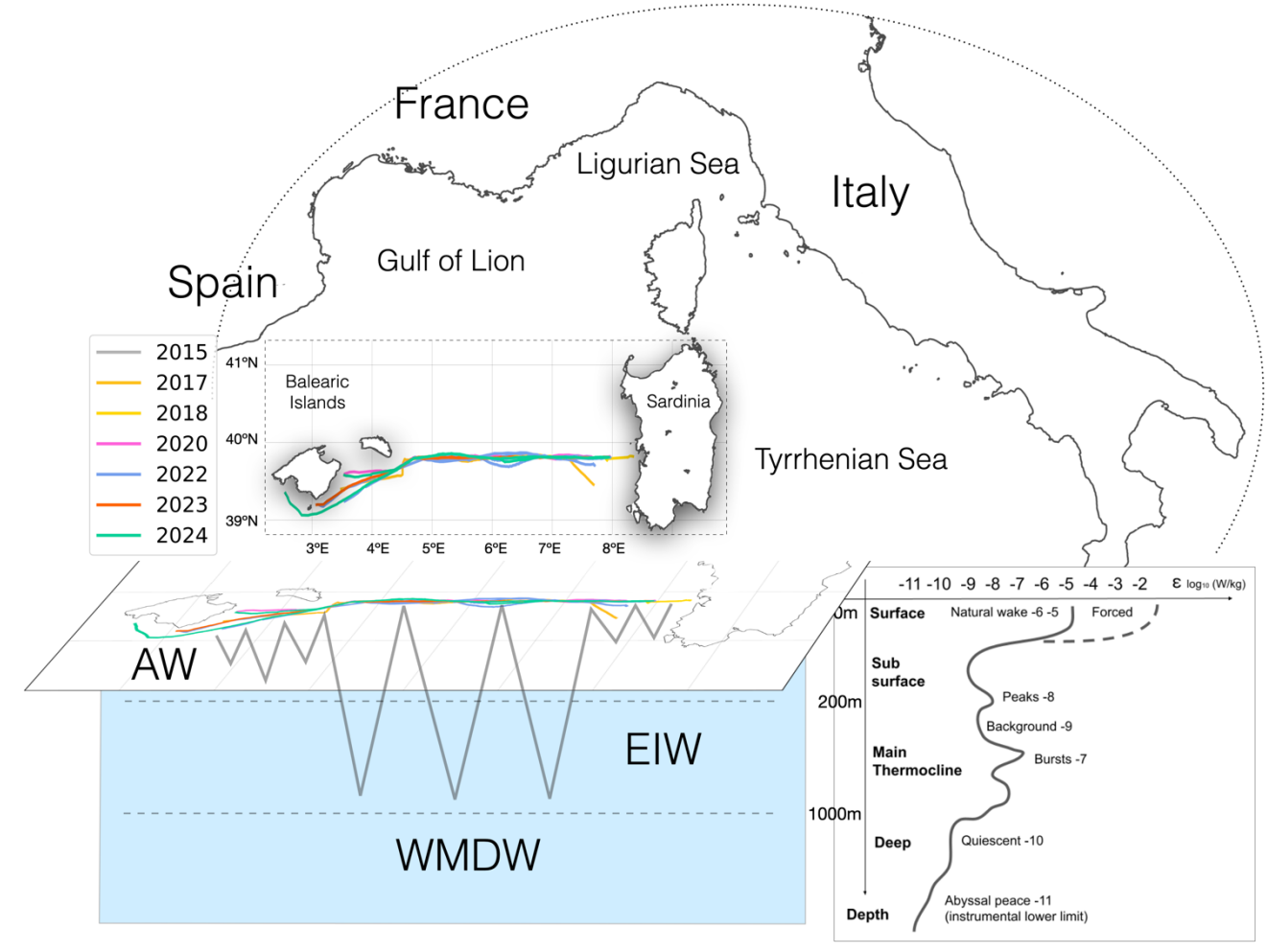


Figure 1. Glider Teresa mission along its longitudinal transect, by years (in colors). The scheme indicates the deployment between Sardinia and Balearic Islands in the Western Mediterranean Sea. Sawing black lines indicate a schematic trajectory from surface up to 1000m-depth, encountering layers of the Atlantic Water (AW) in surface, the Eastern Intermediate Water (EIW) and the Western Mediterranean Deep Water (WMDW). Depth levels associated to these water masses (respectively 0-200m, 200-1000m, 1000m+) are indicative. In bottom right we schematize the levels of dissipation rates of turbulent kinetic energy expected to be observed in depth.

In terms of region of interest, the western Mediterranean Sea serves as a crossroads for oceanic processes that influence regional and basin-scale circulation, water mass transformation, and ecosystem dynamics. The study area, situated between the Balearic Islands and Sardinia (Figure 1), encompasses a complex transitional zone where Atlantic and Mediterranean water masses interact, mesoscale features dominate surface dynamics, and intermediate/deep flows modulate vertical exchanges. This region acts as a nexus for the convergence of multiple circulation systems, including the meandering eastward-flowing Algerian Current, the cyclonic Balearic Current, and the west- and northward propagation of Eastern Intermediate Water (EIW). These interconnected processes make the Balearic-Sardinia corridor

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

a strategic location for investigating the mechanisms governing heat, salt, and biogeochemical fluxes in the Mediterranean. In the study area, the surface layers (0–150 m) are dominated by Atlantic Water (AW), which enters through the Strait of Gibraltar and undergoes progressive salinification as it circulates eastward and cyclonically through the western basin. Below this, the EIW core (200–600 m), characterized by salinities >38.50 PSU and temperatures ~13.3°C, flows northwestward from the Eastern Basin, interacting with locally formed Western Intermediate Water (WIW) in winter.

The region exhibits intense mesoscale variability driven by the instability of the Algerian Current, which generates anticyclonic eddies that propagate into the study area (e.g., Testor et al. 2005, Aulicino et al., 2018). The bathymetry along the transect is predominantly uniform, with depths around 2500 meters, except near the deployment ends where the continental shelves of Sardinia and the Balearic Islands cause shallower topography. Despite its dynamical importance, the Balearic-Sardinia section remained under-sampled due to logistical challenges and the transient nature of its key processes. The MOOSE GE cruises (Testor et al. 2010) that are carried out at annual frequency do not reach the section. Furthermore, these existing hydrographic campaigns provided snapshots but lack the spatial and temporal resolution to capture (i) diurnal-to-seasonal variability in EIW-WIW-WMDW interactions, which may play a role in regulating deep water formation in the Gulf of Lion, (ii) eddy-mediated cross-frontal exchanges that drive subsurface nutrient fluxes to Sardinia's oligotrophic shelf, (iii) responses to climate-driven perturbations, including, e.g., EIW warming and salinification (Schroeder et al., 2016, Testor et al. 2018, Margirier et al. 2020, Chiggiato et al., 2023). A sustained glider transect across this region offers unprecedented capabilities to quantify variations in water mass properties and transport using CTD, dissolved oxygen and microstructure profiles, enabling process-oriented oceanography.

Oceanic turbulent dissipation rates span over nine orders of magnitude with depth ($\log_{10}$ ε in W kg^{-1}, [illegible] shown in Figure; a similar logarithmic range applies to χ, the thermal-variance dissipation rate, which spans its own $\log_{10}$ scale): at the very surface, natural mixed-layer wakes reach $\varepsilon \approx 10^{-6}$–$10^{-5}$; just below, intermittent peaks cluster around $\varepsilon \approx 10^{-8}$; the ambient thermocline supports a background level of $\varepsilon \approx 10^{-9}$ punctuated by sporadic bursts of $\varepsilon \approx 10^{-7}$; in the deep ocean, the flow becomes quiescent at $\varepsilon \approx 10^{-10}$, ultimately approaching instrumental limits near $\varepsilon \approx 10^{-11}$ in the [illegible]vssal pea[illegible]. Autonomous underwater gliders equipped with airfoil shear probes and fast-response thermistors can concurrently resolve ε and χ across this magnitude range during multi-week missions, yielding high-resolution, cross-depth surveys of both mechanical and thermal mixing far beyond the reach of traditional ship-based profilers (Sherman and Davis 1995; Eriksen et al. 2001; Peterson and Fer 2014). Such Turbulent mixing plays a pivotal role in ocean dynamics, influencing heat, salt, nutrient, and carbon fluxes across scales. As such, turbulence (in terms of its related parameters, ε and χ) has gained recognition as emerging (or pilot) Essential Ocean Variable (EOV) (Lindstrom et al. 2012, Le Boyer et al. 2021, https://goosocean.org/what-we-do/framework/essential-ocean-variables/) for sustained ocean observing systems.

Turbulence-related parameters are among the most technically demanding oceanographic variables to measure in situ. Their estimation relies on high-frequency measurements of velocity and temperature fluctuations at centimeter to millimeter scales, requiring fast-response and highly sensitive sensors such as shear probes and FP07 thermistors. These signals must be sampled at high rates and processed into frequency or wavenumber spectra to extract the small-scale turbulent variance from

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

instrument and environmental noise. Spectral methods do not only allow fitting to theoretical turbulence models (e.g., the Nasmyth spectrum for shear; Nasmyth 1970; and the Batchelor spectrum for temperature gradients; Batchelor 1959), but also help isolate turbulent signals from contamination due to platform vibrations and fine-structure variability. Despite the progress made, this indirect estimation process remains technologically challenging and requires careful correction for sensor response, unresolved variance, and motion-induced artifacts.

Historically, the complexity of measuring ε and χ has limited their inclusion in large-scale monitoring efforts. However, recent advances in sensor technology, deployment platforms, and standardized processing protocols have brought these variables to a level of operational maturity. Shear-derived ε (Lueck et al. 2002) and χ estimates are now supported by increasingly robust methodologies for calibration, quality control, and uncertainty quantification (Piccolroaz et al. 2021, Lueck et al. 2024**)**. This progress marks a turning point where both the quantity and quality of turbulence datasets are sufficient to enable their systematic exploitation for improving fine-scale process understanding, parameterizations, and numerical model representation of mixing in the ocean interior. Large turbulence microstructure datasets are increasingly being published and made available through FAIR (Findable, Accessible, Interoperable, Reusable) data practices, supported by community-endorsed metadata standards and controlled vocabularies (e.g., the ATOMIX initiative; Fer et al., 2024), which we aim to align with in the preparation and dissemination of this dataset.

The dataset comprises seven mission-years (2015, 2017, 2018, 2020, 2022, 2023, 2024), ranging from one to over three months of acquisition per year, and covering different seasons (2024: 21st May to 4th July ; 2023: 27th June to 15th August; 2022: 9th September to 12th December; 2020: 2nd March to 5th April; 2018: 23th April to; 2017: 6th to 26th April; 2015: 6th July to 18th August). While the glider missions were routinely conducted and monitored through standard CTD and navigation data, the turbulence microstructure dataset itself remained largely unexplored until this current compilation. As a result, several sensor limitations and data quality issues, affecting early missions went previously undetected. Over time, the acquisition setup and data handling improved significantly, with the period 2020–2024 representing the most consistent and quality-assured segment of the dataset. Earlier missions (2015, 2017) reflect an initial phase of setup and testing, while 2018 data remain excluded due to unrelated technical limitations. The Teresa's dataset consists of a large data ensemble, that once decomposed in continuous sections provides 3446 unique downward or upward gliding profiles across the upper layers of the Western Mediterranean Sea. The general processing choices that we will detail hereafter allowed us to obtain O(10^5) valid estimates of ε and χ after quality control, on a vertical grid of around 1.5m. This provides a rich and multi-purpose data set to be exploited at the crossroad of various important scientific questions from small scales processes to larger-scale variability, in a zone of interest reputed to intercept mesoscale fronts, latitudinal water masses exports, and deep winter convection.

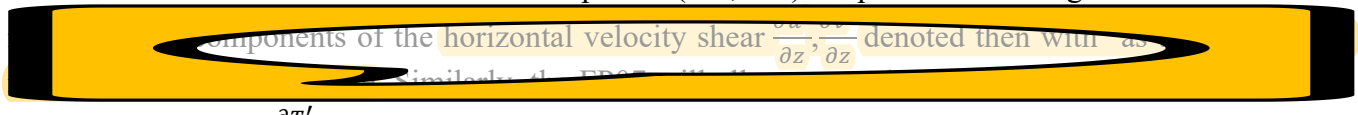
https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

2 Data and Methods

2.1. Microstructure and glider data

Shear and thermistor sensors provide high-frequency time series of data sampled at 512 Hz, which can be transformed into wavenumber or frequency spectra, enabling fits to theoretical turbulence models described thereafter. These high-resolution turbulence data are internally recorded by the MicroRider (model 1000-LP), a microstructure sensor module suitable for integration on a variety of stationary and moving platforms such as gliders, moorings, or wire walkers. The instrument is equipped with sensors located at the front bulkhead, that measure small velocity and temperature fluctuations with respectively two shear probes and two fast-response temperature FP07 sensors. Theses sensors allow to resolve the variance present at small spatial scales where turbulent motions are expected to be significant (on the order of centimeters to decimeters, i.e., ~0.01–0.1 m), and facilitating to transfer energy down to the smallest scales (on the order of millimeters, i.e., ~1 mm) where viscosity will act to finally homogenize the water properties. Such variance is estimated from spectral integration fitted to universal spectral models mentioned thereafter. The two shear probes (sh1, sh2) are positioned orthogonal to each other to



components of the horizontal velocity shear $\frac{\partial u}{\partial z}, \frac{\partial v}{\partial z}$ denoted then with as

temperature through $\frac{\partial T'}{\partial z}$. A pair of piezo-accelerometers serves as two-axis vibration sensor, aside a two-axis inclinometer (pitch and roll angles accurate to 0.1°) to monitor the dynamics of the instrument during the profiling flight. Incident speed should stand within a range from 0.2 to 0.6 m/s, recommended for turbulence measurements using shear probes, being sufficiently fast to satisfy Taylor's frozen turbulence hypothesis, and slow enough to adequately resolve the higher wavenumbers.

While not strictly required for velocity estimation, we exploit the glider's dataset to geolocate the turbulence observations collected by the MR and to compute the instantaneous profiling speed required for converting shear probe signals into physical dissipation units. We apply the Glider Flight Model (GFM) from Merckelbach et al. 2019 to establish the glider incident velocity and angle of attack, used to improve the data conversion (RS Technical Note 039), and the overall turbulent estimates accuracy and further quality control. Processing turbulence data from gliders presents a significant challenge due to the size and complexity of the raw datasets, which must be handled throughout the full processing chain. A single Level 0 file can contain sequences of 2 to 10 or more consecutive upward and downward gliding profiles, each spanning depths from the surface down to 1000 m. These profiles typically represent 10 to 12 hours of continuous acquisition, resulting in file sizes that can reach up to 1 GB per file for deep glides. At the contrary, and detailed further in Section 3 (Processing), the integration of turbulence signals requires spectral averaging using overlapping windows—commonly 4 segments of 3 seconds—yielding a vertical resolution on the order of 1.5 meters. This processing substantially reduces the volume of data, with final profile products typically ranging between 1 and 10 MB per file. Nevertheless, the initial data

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

volume imposes strict constraints on memory handling, processing time, and storage strategy throughout the workflow.

2.2. Turbulent dissipation rates

Microscale turbulence observations enable to estimate of key quantities describing ocean mixing, notably the dissipation rate of turbulent kinetic energy (ε) and the dissipation rate of thermal variance (χ). ε is established from shear fluctuations as in Eq. 1:

$$\varepsilon = \frac{15}{2}\nu\left\langle\left(\frac{\partial u'}{\partial z}\right)^2\right\rangle = \frac{15}{2}\nu\int_0^\infty \psi_s(k)dk \qquad (1)$$

where ν is the kinematic viscosity of water, and $\left\langle\left(\frac{\partial u'}{\partial z}\right)^2\right\rangle$ is the variance of the velocity shear fluctuations, the brackets indicate averaging over a uniform turbulent collection. Here $\psi_s(k)$ is the wavenumber spectrum, k the wavenumber (cpm), related to the frequency $f(s^{-1})$ through the profiling speed $W\ (m\ s^{-1})$as $k = \frac{f}{W}$. Due to non-turbulent variance present in the signal the spectrum can only be used over a certain wavenumber range. A best fit of a reference spectrum, the Nasmyth empirical model (Nasmyth 1970, Osborn and Crawford 1980, Lueck et al. 2002), to the well-resolved part of the spectrum is used to correct for missing variance. We obtain separately ε_1 and ε_2, respectively from sh1 and sh2 probes.

Similarly, χ is determined as in Eq. 2:

$$\chi = 6\kappa_T\left\langle\left(\frac{\partial T'}{\partial z}\right)^2\right\rangle = 6\kappa_T\int_0^\infty \psi_T(k) \qquad (2)$$

where κ_T is the molecular thermal diffusivity and $\left\langle\left(\frac{\partial T'}{\partial z}\right)^2\right\rangle$ the temperature gradient variance resolved by the FP07 thermistor and ψ_T is the temperature gradient (wavenumber) spectrum, for which theoretical models were proposed in Batchelor (1959) and Kraichnan (1968) spectra. Note that once χ is estimated, an associated ε_T can be derived indirectly applying $\varepsilon_T = \nu\kappa_T^2(2\pi k_B)^4$ with k_B is the Batchelor wavenumber established during the spectral fit. Spectral models, wavenumber ranges, and fitting procedures are described in detail in Lueck et al. 2024 and Piccolroaz et al. 2021, respectively, for shears and temperature.

2.3. Processing flow & dataset available

In Figure 2 we illustrate the data flow of our processing chain. We begin with retrieving, archiving, organizing, and listing original data files in directories (Level 0), followed by converting raw data into physical units (Level 1), cleaning and segmenting the time series (Level 2), generating wavenumber spectra from processed sections (Level 3), and finally estimating dissipation rates with quality control metrics (Level 4). These processing levels are designed to standardize the handling of microstructure data

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

and ensure transparency in data processing. Each level builds upon the previous one, adding value and usability to the dataset. By the time data reach L4, they are labeled with a quality control flag, i.e., suitable for addressing complex scientific questions about ocean mixing processes. A general scheme of dataflow is presented thereafter. The dataset published here corresponds to Level 4 data (quality-controlled and validated). Raw and intermediate processing levels (L0–L3) are not included in this publication due to their large volume. Once reached the L4 level, data is exported to a netCDF file (in green on the Figure 2) with an additional list of metadata. The dataset we propose is available here (https://www.seanoe.org/data/00968/107995/, https://doi.org/10.17882/107995). It consists of a unique netCDF "*TERESA_MR_SMART_MISSIONS_2015_2024_L4_QC.nc*" including the missions from the years 2015 to 2024.

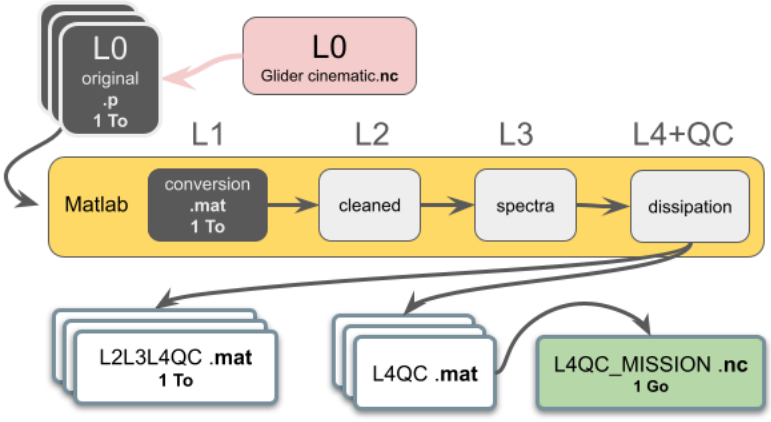


Figure 2: Scheme of the processing steps along the various levels.

3. Processing

All processing routines used in this study are made available (10.5281/zenodo.16541936) and include Python notebooks and MATLAB scripts. The analysis relies on functions from the ODAS v4.51 MATLAB toolbox by RS for shear probe processing, improved with methods outlined in Lueck 2024 et al., and MATLAB routines from Piccolroaz et al. (2021) for thermistor-based estimates. Steps are synthesized in the Table 1, and then detailed in the following sections.

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

Table 1: Summary of the processing routines.

Level 0 **Pre-processing**	**Level 1, 2, 3, 4** **Processing**	**Export**
Glider data Used for GFM model Input Data: **Glider L0 (netCDF)** - L0_flightmodel.ipynb - L0_hotel_file.ipynb. Output: **hotel_file.nc** **Micro Rider data** Input Data: **MR L0 (p-files)** - L0_organize.ipynb Output: **TEMR_L0_organize.csv** (List of all MR files) - L1_overview.m Output: **TEMR_L1_overview.csv** (List of all MR files to be converted and processed + pre-metadata)	Input Data: **MR L0 + hotel_file.nc** Input List: **TEMR_L1_overview.csv** For each file of the list: run_file.m - odas_p2mat.m - L1_make_FID.m - L1_time - L1_glider - L2_sections For each gliding section of each file - run_section.m - L2_section_direction_position - L2_make_section_id - L2_thermistor_source - L2_temperature_source - L2_hipass_lopass - L2_despiking - L3_fft_parameters - L3_hipass_shears - L3_spectra_sh - L3_spectra_th - L4_shears - L4_FP07s - L4_shears_QC - L4_FP07s_QC	**L4 + QC** - make_structure For each gliding section of each file, output: **unique_section.mat** **Final netCDF** netCDF aggregation of all unique_section .mat - L4_make_netcdf.ipynb Output: **"TERESA_MR_SMART_MISSIONS_2015_2024_L4_QC.nc "**

3.0. Pre-processing step — MicroRider data screening and Glider data merging

3.1.1. Glider data

To ensure the quality of turbulence estimates, a correct glider incident velocity must be provided and nested with the original microstructure data, to support the data conversion into physical units and to be used along the general processing. To achieve this, estimates of the glider's speed through water is calculated using the Glider Flight Model (GFM) from Merckelbach et al. (2019). The process employs the Level-0 glider data, available via SOCIB's ERDDAP/THREDDS server (https://thredds.socib.es/thredds/catalog/auv/glider/teresa-cnr_teresa/catalog.html).

In the notebook **L0_flightmodel.ipynb** we load the deployment-specific netCDF file and extract the navigation and physical variables (timestamps, pressure, temperature, conductivity, pitch, roll, oil volume for buoyancy control, and GPS coordinates). The glider's pressure and position data provide a geographical context for the MR profiles, and glider's speed is reconstructed with the flight model using

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.


a physical balance between buoyancy, drag, and pitch (Merckelbach et al. 2019). Code for the model is provided by the authors (https://github.com/smerckel/gliderflight/tree/master https://gliderflight.readthedocs.io/en/latest/index.html). Additionally, glider temperature will be exploited later as a calibrated reference to calculate the kinematic viscosity of seawater (see thereafter in Processing). Once all of these quantities are obtained, a unique hotel file in netCDF format is created with **L0_hotel_file.ipynb**. This is the auxiliary data file that supplies time-synchronized estimates of profiling speed and, and optionally, other dynamic parameters (e.g., angle of attack, pitch) that are not directly available or accurate enough in the raw microstructure data. This external input is particularly useful when the instrument is mounted on platforms such as gliders or AUVs, where the speed through water cannot be reliably estimated from pressure rate alone due to oblique motion or complex vehicle dynamics. An overview of velocities can be consulted in Supplementary Figure 1. Extracted from the hotel, the GFM velocity will override the profiling speed that would be estimated by default from the MR pressure rate change and pitch if no externally computed velocity is provided. The speed is passed to odas_p2mat() via the convert_info.hotel_file argument in input of the function, and it is internally interpolated to the time base of the microstructure data.

3.1.2. MicroRider data screening and pre-load

Files are retrieved and organized in folders, and original raw data files from the MicroRider (.p-files) are listed with **L0_organize.ipynb** in a csv file named TEMR_L0_organize.csv. The following steps are then made using MATLAB scripts. The routine **L1_overview.m** performs a trial check of all listed L0 raw microstructure files, in preparation for L1 data conversion. For each file, it archives the embedded setup_cfg configuration file from the MR (which contains sensor-specific calibration coefficients and acquisition settings), and if needed, replaces it with an updated version to correct inconsistencies in early mission years. It then integrates the external hotel file containing glider data to be nested for data conversion and processing. The script eventually lists general metadata about files (e.g., size, path) and estimates the local time offset to reference glider and MR clocks. TEMR_L0_organize.csv is completed and exported as TEMR_L1_overview.csv that will serve for the processing run. This pre-procedure provides a screening step anticipating the former L1 conversion, ensuring the input data is consistent, patched if necessary, and ready for accurate time-synchronized processing.

3.2. L1 — Converted Data

The master script **run_file.m** calls various sub-scripts (described thereafter) that are used to process a selected L0 raw data file from the screening list and perform L1 conversion, before splitting the data in convenient continuous gliding sections that initiate the L2 step. It starts by identifying the mission year and filename, then extract and archive the internal setup_cfg configuration file. For data collected in or before 2022, it replaces the configuration with an updated version (setup_216_corrected.cfg) to correct known configuration issues. A structure variable (conversion_info) is declared, containing information for data conversion, including a pointer to the external hotel file, which provides synchronized glider

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

speed and angle of attack. A time offset between glider and turbulence timestamps obtained at the pre-processing step is applied there and ensure time alignment. Data is converted from raw signal to physical units through **odas_p2mat().m**, supported by the glider's velocity provided by the hotel file. This glider velocity replaces the MR default speed (inferred from its own sensors alone). Any other variables nested into the hotel file are merged with the microstructure converted data and are made available as additional fields in the MATLAB data structure. A unique filename identifier (FID) is built for each converted file in **L1_make_FID.m**, combining mission metadata such as: glider name (TERESA), sensor type (MR), conversion level (L1_converted), internal file ID and original filename, date and time extracted from the data header. Time vectors are defined in **L1_time.m** for both slow and fast acquisition channels (sampling respectively at 64 and 512 Hz). It combines the starting timestamp (date, hour, minute, second, millisecond) with the elapsed time vectors t_fast and t_slow to produce MATLAB datetime and datenum arrays. If not already interpolated during the conversion, glider variables are interpolated in **L1_glider.m** on both fast and slow time stamp grids to be ready to use alongside the microstructure data. Key variables include angle of attack (AOA), pitch, roll, temperature, conductivity, salinity and density, kinematic viscosity, thermal diffusivity, and position (longitude and latitude).

3.3. L2 — Unique gliding sections

At the end of **run_file.m,** once a data has been fully loaded, the script **L2_sections.m** serves as a transition between L1 and L2 levels. It segments time series into continuous unique gliding sections, suitable and convenient to carry further dissipation estimation. For this, it employs pressure (P_fast) and vertical velocity (W_fast) in the function **get_profile.m** from ODAS. A low-pass Butterworth filter is first applied to W_fast using a cutoff frequency Fc equal to the mean glider speed. This suppresses high-frequency noise at vertical velocity to help detecting the profiling starting and ending indices. A profile is accepted if it satisfies minimum conditions: Depth ≥ Pmin (e.g. 3 dbar); Vertical speed ≥ |Wmin| (e.g. 0.01 m/s); Duration ≥ minDuration (e.g. 60 seconds). The function detects down and upcasts and start and end indexes of each detected section are retained, and a unique integer label is assigned. Direction flags are also set (+1 for downcast, –1 for upcast). Once listed, unique gliding sections will be processed separately (i.e., in loop) and be passed through the sequence of scripts called on **run_sections.m** that we describe thereafter.

3.4. L2 — Cleaning and processing

We identify in **L2_section_direction_position.m** the temporal extent and key navigation attributes, and we extract the section indices of the current gliding section. A unique identifier is created with **L2_make_section_id.m**. The routine considers the original file identification string defined at L1 and adds other strings suffixes from the values of the current section. The core identification is as follows, with XXX being named accordingly from the level of processing to be considered in case of data export (e.g., XXX = 'L1', 'L2', or 'L1L2L3' etc...), if applicable: e.g., *TERESA_MR_**XXX**_converted_file_0062_DAT_063_2024_06_06_23_40_05*. As we separate by unique

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

section, we add: the position as: Lat, Lon *lat_39_7987_lon_07_6116 ;* Navigation: *nav_E ;* Pmin, Pmax: *pmin_0003_pmax_0954* ; Section number: *sec_001;* Section total: *on_004*; Gliding direction: *glid_down.* This convention produces long but robust and comprehensive filenames: *TERESA_MR_**XXX**_converted_file_0062_DAT_063_2024_06_06_23_40_05_lat_39_7987_lon_07_6116_nav_E_pmin_0003_pmax_0954_sec_001_on_004_glid_down.*

The script **L2_thermistor_source.m** identifies the most reliable fast thermistor (FP07) to be used as the reference thermistor signal. It compares the two FP07 time series (T1_fast and T2_fast) against the glider's temperature (T_gl_slow, interpolated on the fast channel) by calculating the Pearson correlation coefficient (cc) between signals. If both sensors correlate significantly with the glider temperature (above a threshold, e.g., cc>0.3), the one with the higher correlation is selected as the master one (flag 11 or 22). If only one meets the threshold, that sensor is chosen (flag 1 or 2). If none correlates significantly, the decision is made based on the variance of each FP07 channel. If both variances are unusually low (below 1e-5), the function flags potential malfunction (flag 0); if only one is below, the other is selected (flag 100 or 200). In case both are above the variance threshold but do not correlate with T_gl_slow, the sensor with the lower variance is preferred (flag 10 or 20). The outcome is stored with a logical value and a string (thermistor_source). This step allows to track potential FP07 malfunctions. Note that a malfunctioning sensor will not pass the quality control applied later in L4.

We set in **L2_temperature_source.m** the glider temperature (T_gl_fast) as the reference signal to calculate kinematic viscosity of seawater, needed to compute ϵ. It bases its decision on the previously assigned Th_source_logic flag from the thermistor comparison step. In most cases (Th_source_logic = 11, 22, 1, 2, 0, 100, 200), the glider temperature is retained as the default temperature input. However, if both thermistors correlate poorly with the glider temperature but exhibit usable variance (flags 10 or 20), glider's T_gl_fast is not trusted and the thermistor with the lower variance is selected (T1_fast or T2_fast). The selected signal is assigned to the temperature_for_dissipation variable along with a descriptive string and logic flag, to be used later in the function get_diss_odas().

The script **L2_hipass_lopass.m** applies sequential high-pass and low-pass Butterworth filters to the shear probe signals (sh1, sh2). First, a high-pass filter with a cutoff frequency of 0.1 Hz is used to remove low-frequency trends and motion-related biases, preserving the turbulent fluctuations of interest. The absolute value of the high-passed signals is computed to obtain envelope-like signals (sh1hpa, sh2hpa). Then, a low-pass filter with a cutoff at 1 Hz is applied to smooth these envelope signals, yielding sh1hpalp and sh2hpalp to be conserved apart for other applications (e.g., visual check).

We perform through **L2_despiking.m** an automated spike detection and removal on the filtered shear signals (sh1hpa, sh2hpa) using a despiking algorithm. It applies the ODAS **despike()** function with a defined amplitude threshold (thresh = 8), a frequency cutoff (fcut = 0.5 Hz), and a smoothing window length (N = 0.04 × Fs, with Fs = 512 Hz) to identify and suppress sharp, non-physical signal excursions. The outputs include the despiked shear signals (sh1hpa_dsp, sh2hpa_dsp), the spike indices, the number of iterations required to converge (pass_count), and the fraction of samples affected (ratio). Additionally, spike indices are conserved with their associated pressure levels (P_spikes_sh1, P_spikes_sh2) to flag and keep track of spike occurrences.

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

3.5. L3-L4 — Spectral computation and turbulent estimates

Wavenumber spectra are calculated from the cleaned gliding sections obtained at the L2 step. We employ there the functions **get_diss_odas()** for shears, and **gradT_dis_spec()** for FP07, that do both spectral computation and integration to obtain ϵ and χ, respectively. Once calculated, the different outputs are organized through L3 and/or L4 products.

Dissipation rates of turbulent kinetic energy (ε) are estimated from shear probe data using spectral integration of the velocity gradient spectra, following the procedures described in Lueck et al. (2024). If ε exceeds ~10^{-5} W kg^{-1}, a transition is made from direct integration in the variance subrange (VSR) to inertial subrange (ISR) fitting, as spectral roll-off and probe resolution limit the reliability of the full-spectrum approach. Following Lueck et al. 2024, quality metrics that will be presented hereafter are calculated and allow to flag quality-controlled passing estimates.

For temperature microstructure, thermal variance dissipation rates (χ) are estimated from FP07 thermistor spectra following Piccolroaz et al. (2021). This includes correction for the sensor's finite time response, which acts as a low-pass filter and attenuates high-frequency content of the temperature gradient spectrum. The correction is based on profiler speed and the thermistor's thermal time constant (typically around 7 ms for FP07 sensors), and is applied through a transfer function modeled after the thermistor's response characteristics. Accurate χ estimation depends critically on this correction, especially in energetic conditions where high wavenumber contributions are significant. As with ε, χ estimates are quality controlled using statistical thresholds and consistency between. Each ε or χ estimate is accompanied by metadata including the wavenumber range of integration, spectral model used, uncertainty metrics (e.g., standard deviations from VSR or ISR methods), and a consolidated QC flag. At the end, if both sensors pass quality assurance, a strict final dissipation or thermal variance estimate is computed as the average across sensors.

We define in **L3_fft_parameters.m** the parameters for the Fast Fourier Transform (FFT). It starts by estimating the glider's mean speed (speed_mean) and vertical speed (w_mean) over the gliding section. The characteristic FFT window duration (tau_fft) is determined by the minimum of two criteria: (i) to avoid signal contamination from the 1.5-m vehicle-scale motions (vehicle_length / speed), and (ii) to resolve a spectral scale of 0.5 cpm. From this duration, the number of FFT points (N_fft) is computed using the sampling frequency, and the corresponding spatial window length (L_fft) is used to define the lowest resolved wavenumber (kl = 1/L_fft). The code also sets the high-pass frequency cutoff (Fhp) and sets the FFT window length of Ntimes=4 times the segments length (N_fft), with a 50% overlap. This set up generally lead to obtain windows of 12 seconds from four FFT segments of 3 seconds.

In **L3_hipass_shears.m, a** 1st-order Butterworth high-pass filter is applied on shears (sh1hpa_dsp, sh2hpa_dsp) with a cutoff frequency (Fhp) derived in the FFT parameters, to remove low-frequency noise. The filtered shear are obtained using zero-phase filtering to prevent phase distortion.

Key parameters for the Fast Fourier Transform (FFT) are used to lead spectral dissipation estimation in **L3_spectra_sh.m.** The ODAS function **get_diss_odas()** computes the dissipation rate ϵ from the shear spectra. Vibration-induced noise from the glider platform and pump is filtered from the raw shear signals through the noise correction implemented in the ODAS v4.5.1 toolbox, which consist

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.


of removing coherent signals between the shears and vibration sensors (Ax, Ay) in the frequency domain, following Goodman et al. 2006. Profiles of dissipation rate (ϵ) are obtained then for each shear sensor.

Spectral analysis of temperature gradient is performed in **L3_spectra_th.m** using the **gradT_dis_spec()** routine from Piccolroaz et al. 2021 to estimate temperature variance dissipation rates (χ). Inputs are prepared with pressure, temperature gradients (dT/dz), and required parameters (e.g., kinematic viscosity of seawater, spectral models). Segments of the temperature gradient signal are analyzed, and theoretical spectra (Batchelor or Kraichnan) are applied to fit the observed spectra. Quality metrics obtained from the routines are: Mean Absolute Deviation (MAD), wavenumber range used, likelihood ratios, and QC flags for spectral fits. Profiles of temperature dissipation rates (χ) are obtained for each FP07 sensor.

Once the estimates are obtained, **L4_shears.m** compiles and organizes the dissipation coefficients. It interpolates glider-derived variables (e.g., speed, temperature, salinity, density, pitch, roll, angle of attack) onto the time base of the estimates. Uncertainty estimates are obtained directly from the spectral fitting routine get_diss_odas() and stored to be employed thereafter. In **L4_FP07s.m** are extracted and organized the dissipation estimates of thermal variance (χ). It retrieves the associated pressure and depth vectors, timestamps, and positions.

3.6. L4 — Quality Control (QC).

A structured set of quality control (QC) is applied in **L4_shears_QC.m** and **L4_FP07s_QC.m** to the individual estimates, respectively (ε_1,ε_2) and (χ_1, χ_2). The QC flag for shear-derived ε is a single cumulative value that encodes the outcome of several individual quality tests, resumed in the Table 2: (1) Figure of Merit (FOM), which fails if the spectral fit exceeds a threshold (FOM > 1.4); (2) Spike Fraction, which flags data with more than 15% of points removed during despiking; (4) Inter-Probe Epsilon Ratio, applied only if both probes have valid FOM, and flags significant disagreement between ε_1 and ε_2 (values out of $2.77 \times \sigma(\ln \varepsilon)$ for VSR; $4.2 \times \sigma(\ln \psi)$ for ISR, see Lueck et al. 2024) ; (8) Spike Iteration Count, which flags segments requiring more than 9 despiking passes; (16) Variance Resolution, which fails if less than 60% of the shear spectrum variance is resolved; (32) Relaxed Spike Fraction, used to flag cases where spike fraction falls between 5% and 15%; (64) Method Mismatch, which flags segments in case of the two probes used different estimation methods (VSR or ISR) on the same data segment; (128) that flags angles of attack out of the range 1.5º-4.5º. The final QC flag is a bitwise sum of failed tests. For example, a QC of (5) means the FOM test (1) and estimates ratio test failed (4).

In the case of thermistor, we employ the QC flag in output of the routine from Piccolroaz et al. 2021, that we conveniently reordered as (0) for good data, (1) if both estimates flags initially (0) but are separated by one order of magnitude in intensity when cross-checked, and (2) for poor estimate. Note that in their routines the poor QC flag combines multiple spectral quality criteria into a single flag that we don't exploit separately, including: (i) the likelihood ratio (LR), requiring that the fit to the Kraichnan model significantly outperforms a power-law fit (LR > 100); (ii) the integration range criterion, ensuring that the spectral peak and roll-off are both resolved; (iii) a signal-to-noise ratio threshold, with SNR > 1.3

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

in the fitted spectral range; and (iv) the effect of the sensor time-response correction to avoid spectral distortion.

Table 2: Summary of QC controls.

QC Flag	Test	Failure Condition	Interpretation
SH 1,2			
0	All	None	Validated, in condition of inter-probes cross check
1	Figure of Merit (FOM)	FOM > 1.4	Poor spectral fit to the model
2	Spike Fraction	Spike fraction > 15%	Possible platform noise or collisions (e.g., zooplankton)
4	Ratio of estimates	To be applied only if FOM is < 1.4 VSR: 2.72 σ(ln ε), ISR: 4.20 σ(ln ψ)	e1-vs-e2 Inter-Probe Ratio If ratio is excessive, the largest is rejected
8	Spike Iteration	Count more than 9 despiking iterations	Possible platform noise or collisions (e.g., zooplankton)
16	Variance Resolution	Resolved variance fraction < 0.6	Incomplete spectrum underestimation of ε, especially in low-dissipation regimes
32	Relaxed Spike Fraction	Spike fraction is 5% < x < 15%	Relaxed version of QC = 2
64	Method mismatch	Not the same method (VSR, ISR) for $\varepsilon_1,\varepsilon_2$	Signal inconsistence
128	Angle of Attack	Out of 1.5°— 4.5°	Angle of attack not suitable to ensure a correct flow sampling
FP07 1,2			
0	All	None	Validated
1	Estimates ratio	Both estimates are ok but differs by one order of magnitude	Rejected
2	Series of spectral test	Kraichnan likelihood < 100, unresolved roll-off region ranges, signal-to-noise < 1.3, sensor time-response correction out of bounds	Rejected

4. Data structure

4.1. Data structure and export

The script **make_structure.m** harvests all relevant metadata, parameters, diagnostics, and turbulence estimates into structured MATLAB objects for each glider profile section. It creates a master structure named "MR" composed of the following subfields: meta, L2_params, L2, L3_params, L3, L4, and QC. The meta structure contains time, location, and mission-specific metadata including start/end timestamps, glider direction, pressure range, and navigation information. L2_params records all settings and results related to shear signal preprocessing, filtering, despiking, and FFT setup. The L2 field holds

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.


the actual filtered and processed shear data. L3_params and L3 retain parameters and outputs of the dissipation estimation routine (e.g., FFT settings). L4 consolidates final dissipation results, including ε estimates from both shear probes and χ estimates from both thermistors, as well as interpolated environmental and glider-derived variables. Finally, the QC field aggregates all quality control metrics, flags, and validated dissipation values.

MR is the full detailed MATLAB structure, including all intermediate signals, spectra, and metadata (100-500 MB.) In contrast, mr is a lighter (1-5 MB). version that keeps only the essential metadata: meta, L2_params, L3_params, L4, and QC. Given the resource limitations in terms of storage and computational capacity, we publish only the lighter mr structure, while retaining the full MR structure internally for traceability, reproducibility, and potential reprocessing if needed. Each individual section file is exported as a .mat structure using the -v7.3 format. Each individual profile-level.mat output is named using a long-form convention that encodes key metadata directly into the filename. For example: *TERESA_MR_QC_converted_file_0062_DAT_063_2024_06_06_23_40_05_lat_39_7987_lon_07_6116_nav_E_pmin_0003_pmax_0954_sec_001_on_004_glid_down.mat*

4.2. netCDF aggregation

We aggregate all the section data and metadata as variables and attributes into a netCDF file. It contains several groups of variables, organized in five different dimensions. Note that suffixes _SHEAR or _THERM serve to distinguish between the variables related to the shears or FP07s sensors, respectively. Here we give a generic example of dimensions in case of a vector of X shear-based estimates for 2 shear sensors, and Y thermistor-based estimates for 2 thermistor sensors, all obtained among Z unique sections among all the mission years:

- SECTION contains scalar values used for each individual section (e.g., processing parameters). Dimension is (Z,1).
- TIME_SPECTRA_SHEAR serves to contain shear-related estimates. Dimension is (X,1)
- TIME_SPECTRA_THERM serves to contain thermistor-related estimates. Dimension is (Y,1)
- N_SHEAR_SENSORS. Dimension is (2, 1).
- N_THERM_SENSORS. Dimension is (2, 1).

Note that TIME_SPECTRA_SHEAR and TIME_SPECTRA_THERM are different given the two different spectral computations leading to slight variations in FFT lengths, and consequently on depth and timestamps associated to their respective estimates. They can be merged later, e.g. on the same depth/time grid, once the QC choices are made to filter out values. Variables and their dimensions are summarized in the Table 3.

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

525

Table 3: Overview of the variables made available in the netCDF file.

Variable name	VarType	Standard Name	Units	Description	Dim1
SECTION	Dimension	number_of_section	—	Dimension	
SECTION_INDEX	Variable	unique_identifier_for_each_section_of_data_from_timeseries	—	Integer index attributed to each section	N_SECTION
SECTION_fs_fast	Variable	fs_fast	Hz	Sampling frequency for fast channels	N_SECTION
SECTION_fs_slow	Variable	fs_slow	Hz	Sampling frequency for slow channels	N_SECTION
SECTION_profiling_direction	Variable	profiling_direction_updown	Logic	Upward (-1) or downward (1) gliding direction	N_SECTION
SECTION_speed_mean	Variable	gliding_speed	m s-1	Average gliding speed over the section	N_SECTION
SECTION_vehicle_length	Variable	vehicle_length	m	Glider length used for advection wavelength	N_SECTION
SECTION_tau_to_avoidadv	Variable	tau_to_avoidadv	s	Duration to avoid the inclusion of advection motions	N_SECTION
SECTION_tau_to_resolve_05cpm	Variable	tau_to_resolve_05cpm	s	Duration to resolve wavelength of 5 cpm	N_SECTION
SECTION_tau_fft	Variable	tau_fft	s	Duration retained for the FFT	N_SECTION
SECTION_Pearson_TglT1	Variable	pearson_TglT1	—	Correlation coefficient bewteen thermistors and glider's temperature	N_SECTION
SECTION_Pearson_TglT2	Variable	pearson_TglT2	—	Correlation coefficient bewteen thermistors and glider's temperature	N_SECTION
SECTION_Therm_source	Variable	therm_source	Flag	Flag identifying the thermistor retained as master	N_SECTION
SECTION_Temp_source	Variable	temp_source	Flag	Flag identifying the temperature retained for kinematic viscosity	N_SECTION
SECTION_f_AA	Variable	f_AA	Hz	Anti-aliasing frequency	N_SECTION
SECTION_SHEAR_fft_length	Variable	SHEAR_fft_length	—	Length of the FFT (in data points)	N_SECTION
SECTION_SHEAR_diss_length	Variable	SHEAR_diss_length	—	Length of the data used for dissipation estimates	N_SECTION
SECTION_SHEAR_overlap	Variable	SHEAR_overlap	—	Length of overlap in FFT (in data points)	N_SECTION
SECTION_SHEAR_goodman	Variable	SHEAR_goodman	Logic	Application of Goodman coherent noise removal: 1-true	N_SECTION
SECTION_SHEAR_HP_cut	Variable	SHEAR_HP_cut	Hz	High-pass filter cut-off frequency	N_SECTION
SECTION_SHEAR_kl	Variable	SHEAR_kmin	cpm	Minimum wavenumber used for estimates	N_SECTION
SECTION_SHEAR_L_fft	Variable	SHEAR_L_fft	m	Length of the FFT (in meters)	N_SECTION
SECTION_SHEAR_N_fft	Variable	SHEAR_N_fft	—	N times the FFT length	N_SECTION
SECTION_SHEAR_FOM_limit	Variable	SHEAR_FOM_limit	—	Figure of Merit limit for quality assurance	N_SECTION
SECTION_SHEAR_diss_ratio_limit_VSR	Variable	SHEAR_diss_ratio_limit	—	Limit of large disagreement between dissipation estimates between shear probes	N_SECTION
SECTION_SHEAR_diss_ratio_limit_ISR	Variable	SHEAR_diss_ratio_limit	—	Limit of large disagreement between dissipation estimates between shear probes	N_SECTION
SECTION_SHEAR_despike_shear_fraction_limit	Variable	SHEAR_despike_shear_fraction_limit	%	Maximum allowed fraction of data removed by de-spiking	N_SECTION
SECTION_SHEAR_despike_shear_fraction_limit_relax	Variable	SHEAR_despike_shear_fraction_limit_relaxed	%	2nd Maximum allowed fraction of data removed by de-spiking	N_SECTION
SECTION_SHEAR_despike_shear_iterations_limit	Variable	SHEAR_despike_shear_iterations_limit	—	The maximum number of iteration allowed for shear de-spiking	N_SECTION
SECTION_SHEAR_variance_resolved_limit	Variable	SHEAR_variance_resolved_limit	—	The minimum fraction of variance resolved for an estimate by spectral integration	N_SECTION
SECTION_SHEAR_f_limit	Variable	SHEAR_f_limit	cpm	The upper limit to exclude frequencies from analysis	N_SECTION
SECTION_SHEAR_fit_2_isr	Variable	SHEAR_fit_2_isr	W kg-1	Threshold for using the method of fitting in the internial subrange. 10e-5 W/kg	N_SECTION
SECTION_SHEAR_spectral_model	Variable	SHEAR_spectral_model	Logic	The model shear spectrum used in dissipation estimates with the integration method	N_SECTION
SECTION_THERM_npoles	Variable	THERM_npoles	—	Transfer function for time response correction 'single' or 'double' pole	N_SECTION
SECTION_THERM_fft_length	Variable	THERM_fft_length	—	Length of the FFT (in data points)	N_SECTION
SECTION_THERM_diss_length	Variable	THERM_diss_length	—	Length of the data used for dissipation estimates	N_SECTION
SECTION_THERM_overlap	Variable	THERM_overlap	—	Length of overlap in FFT (in data points)	N_SECTION
SECTION_THERM_Tdis	Variable	THERM_Tdis	Logic	Type of theoretical spectrum: Kraichnan, Batchelor	N_SECTION
SECTION_THERM_q	Variable	THERM_q	—	Turbulent parameter	N_SECTION
SECTION_THERM_tau_0	Variable	THERM_tau_0	s	Nominal response time	N_SECTION
SECTION_THERM_time_corr	Variable	THERM_time_corr	Logic	Time correction approach: KOC, RSI, NAS, SOM	N_SECTION
SECTION_THERM_int_range	Variable	THERM_int_range	Logic	('L')	N_SECTION
N_SHEAR_SENSORS	Dimension	unique_identifier_for_each_section_of_data_from_timeseries	—	Dimension	
TIME_SPECTRA_SHEAR	Dimension	time_of_shear_estimates	s	Dimension. Seconds since 1970-01-01	
SECTION_NUMBER_SHEAR	Variable	s	—	ID number attributed to each section	TIME_SPECTRA_SHEAR
PSPD_REL	Variable	plateform_speed_wrt_sea_water	m s-1	Glider incident velocity, from GFM (Merckelbach et al. 2019)	TIME_SPECTRA_SHEAR
PSPD_REL_STD	Variable	std_plateform_speed_wrt_sea_water	m s-1	Glider incident velocity std	TIME_SPECTRA_SHEAR
PSPD_AOA	Variable	plateform_angle_of_attack_sea_water	angular degrees	Glider angle of attack, from GFM (Merckelbach et al. 2019)	TIME_SPECTRA_SHEAR
PRES_SHEAR	Variable	sea_water_pressure	dbar	Pressure at dissipation estimates	TIME_SPECTRA_SHEAR
LON_SHEAR	Variable	longitude	decimal degree	Longitude at dissipation estimates	TIME_SPECTRA_SHEAR
LAT_SHEAR	Variable	latitude	decimal degree	Latitude at dissipation estimates	TIME_SPECTRA_SHEAR
KVISC_SHEAR	Variable	kinematic_viscosity_of_sea_water	m2 s-1	KVISC at dissipation estimates	TIME_SPECTRA_SHEAR
Z_SHEAR	Variable	depth	m	Depth at dissipation estimates	TIME_SPECTRA_SHEAR
EPSI_SHEAR	Variable	specific_turbulent_kinetic_dissipation_in_sea_water	W kg-1	Dissipation rate of turbulent kinetic energy, from shear 1 and 2	TIME_SPECTRA_SHEAR N_SHEAR_SENSORS
EPSI_SHEAR_FINAL	Variable	specific_turbulent_kinetic_dissipation_in_sea_water	W kg-1	Dissipation rate of turbulent kinetic energy, averaged from shear 1 and 2 where QC=0	TIME_SPECTRA_SHEAR
EPSI_SHEAR_STD_VSR	Variable	expected_standard_deviation_vsr	W kg-1	2.77 x std(epsilon). See Lueck et al. 2024	TIME_SPECTRA_SHEAR N_SHEAR_SENSORS
EPSI_SHEAR_STD_ISR	Variable	expected_standard_deviation_isr	W kg-1	4.2 x std(psi). See Lueck et al. 2024	TIME_SPECTRA_SHEAR N_SHEAR_SENSORS
EPSI_SHEAR_FLAGS	Variable	dissipation_qc_flags	Integer flag	Quality flags	TIME_SPECTRA_SHEAR N_SHEAR_SENSORS
KMIN_SHEAR	Variable	minimum_wavenumber_used_for_dissipation_estimate	cpm	Wavenumber for spectral integration	TIME_SPECTRA_SHEAR N_SHEAR_SENSORS
KMAX_SHEAR	Variable	maximum_wavenumber_used_for_dissipation_estimate	cpm	Wavenumber for spectral integration	TIME_SPECTRA_SHEAR N_SHEAR_SENSORS
N_S_SHEAR	Variable	number_of_spectral_point_for_dissipation_estimate	—	Spectral integration	TIME_SPECTRA_SHEAR N_SHEAR_SENSORS
FOM_SHEAR	Variable	figure_of_merit	—	Metric for quality control	TIME_SPECTRA_SHEAR N_SHEAR_SENSORS
MAD_SHEAR	Variable	mean_absolute_deviation	—	Metric for quality control	TIME_SPECTRA_SHEAR N_SHEAR_SENSORS
VAR_RESOLVED_SHEAR	Variable	variance_resolved	—	Metric for quality control	TIME_SPECTRA_SHEAR N_SHEAR_SENSORS
DESPIKE_FRACTION_SHEAR	Variable	fraction_of_shear_data_modified_by_despiking	%	Metric for quality control	TIME_SPECTRA_SHEAR N_SHEAR_SENSORS
DESPIKE_PASSCOUNT_SHEAR	Variable	number_of_pass_for_data_despiking	—	Metric for quality control	TIME_SPECTRA_SHEAR N_SHEAR_SENSORS

530

N_THERM_SENSORS	Dimension	number_of_thermistor_sensors	—	Dimension	
TIME_SPECTRA_THERM	Dimension	time_of_thermistor_estimates	s	Dimension. Seconds since 1970-01-01	
SECTION_NUMBER_THERM	Variable	unique_identifier_for_each_section_of_data_from_timeseries	—	ID number attributed to each section	TIME_SPECTRA_THERM
PRES_THERM	Variable	sea_water_pressure	dbar	Pressure at thermistor estimates	TIME_SPECTRA_THERM
LAT_THERM	Variable	latitude	decimal degree	Latitude at dissipation estimates	TIME_SPECTRA_THERM
LON_THERM	Variable	longitude	decimal degree	Longitude at dissipation estimates	TIME_SPECTRA_THERM
KVISC_THERM	Variable	kinematic_viscosity_of_sea_water	m2 s-1	KVISC at dissipation estimates	TIME_SPECTRA_THERM
SW_Diffusivity_THERM	Variable	thermal_diffusivity_of_sea_water	m2 s-1	SWDIFF at dissipation estimates	TIME_SPECTRA_THERM
Z_THERM	Variable	depth	m	Depth at dissipation estimates	TIME_SPECTRA_THERM
XIT_THERM	Variable	specific_turbulent_thermal_dissipation_in_sea_water	°C^2 s-1	Dissipation rate of thermal variance, from thermistor 1 and 2	TIME_SPECTRA_THERM N_THERM_SENSORS
XIT_THERM_FINAL	Variable	specific_turbulent_thermal_dissipation_in_sea_water	°C^2 s-1	Dissipation rate of thermal variance, from thermistor 1 and 2 where QC=0	TIME_SPECTRA_THERM
XIT_THERM_FLAGS	Variable	xit_qc_flags	Integer	Quality flags	TIME_SPECTRA_THERM N_THERM_SENSORS
KB_THERM	Variable	Batchelor_wavenumber_after_MLE_spectral_fitting	cpm	Wavenumber for spectral integration	TIME_SPECTRA_THERM N_THERM_SENSORS
KMIN_THERM	Variable	lower_integration_wavenumber	cpm	Wavenumber for spectral integration	TIME_SPECTRA_THERM N_THERM_SENSORS
KMAX_THERM	Variable	upper_integration_wavenumber	cpm	Wavenumber for spectral integration	TIME_SPECTRA_THERM N_THERM_SENSORS
K_P_THERM	Variable	wavenumber_corresponding_to_fitted_theoretical_spectrum_peak	cpm	Wavenumber for spectral integration	TIME_SPECTRA_THERM N_THERM_SENSORS
LR_THERM	Variable	likelihood_ratio	—	Metric for quality control	TIME_SPECTRA_THERM N_THERM_SENSORS
MAD_T_THERM	Variable	g	—	Metric for quality control	TIME_SPECTRA_THERM N_THERM_SENSORS
MADc _THERM	Variable	threshold_for_mad_between_observed_and_empirical_spectra	—	Metric for quality control	TIME_SPECTRA_THERM N_THERM_SENSORS

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

ROLL_MR	Variable	roll_from_microrider_sensors	Angular degrees	Microrider values at dissipation estimates	TIME_SPECTRA_SHEAR
ROLL_GL	Variable	roll_from_glider_sensors	Angular degrees	Glider values at dissipation estimates, from L0 data	TIME_SPECTRA_SHEAR
PITCH_MR	Variable	pitch_from_microrider_sensors	Angular degrees	Microrider values at dissipation estimates	TIME_SPECTRA_SHEAR
PITCH_GL	Variable	pitch_from_glider_sensors	Angular degrees	Glider values at dissipation estimates, from L0 data	TIME_SPECTRA_SHEAR
P_GL	Variable	sea_water_pressure	decibar	Glider values at dissipation estimates, from L0 data	TIME_SPECTRA_SHEAR
T_GL	Variable	sea_water_temperature	Celsius degree	Glider values at dissipation estimates, from L0 data	TIME_SPECTRA_SHEAR
T1_MR	Variable	thermistor1_microrider	Celsius degree	Values at dissipation estimates	TIME_SPECTRA_SHEAR
T2_MR	Variable	thermistor2_microrider	Celsius degree	Values at dissipation estimates	TIME_SPECTRA_SHEAR
nu_GL	Variable	kinematic_viscosity_of_sea_water	m^2 s-1	Glider values at dissipation estimates, from L0 data	TIME_SPECTRA_SHEAR
C_GL	Variable	conductivity_of_sea_water	S/m	Glider values at dissipation estimates, from L0 data	TIME_SPECTRA_SHEAR
SP_GL	Variable	practical_salinity_of_sea_water	PSU	Glider values at dissipation estimates, from L0 data	TIME_SPECTRA_SHEAR
SA_GL	Variable	absolute_salinity_of_sea_water	g kg-1	Glider values at dissipation estimates, from L0 data	TIME_SPECTRA_SHEAR
RHO_GL	Variable	density_of_sea_water	kg m-3	Glider values at dissipation estimates, from L0 data	TIME_SPECTRA_SHEAR

5. Results

5.1. QC for spikes and Figure of Merit (FOM)

We implemented quality control metrics related to despiking, including a spike fraction test that assigns QC=2 when more than 15% of the clean section points are affected. This flag indicates that the estimate meets all core validation criteria, acceptable figure of merit, agreement between probes, and resolved variance, but eventually shows elevated spiking. These cases are not necessarily invalid, as they may reflect dense biological layers, brief mechanical disturbances, or valid turbulent signals partially masked by benign spikes. A second test based on the number of despiking iterations assigns QC=8 when more than 9 passes are required. There is currently no strong consensus on acceptable iteration limits (Lueck et al., 2024), and ODAS processing caps the count at ten by default, which limits the interpretive value of this flag. Our statistical analysis in Figure 3 reveals a subset of estimates with spike fractions between 5% and 15% and iteration count below the cutoff. To retain these estimates, we introduce a relaxed quality flag, QC=32, marking them as conditionally valid. Due to profiles extending up to 1000 meters, in such long records the likelihood of encountering localized spikes increases, even if their overall effect is minor. Applying a fixed percentage threshold across variable-length records may lead to biased rejection, penalizing longer profiles. We propose to include enventually QC=32 as a compromise, and let the choice to the user, allowing inclusion of these points in secondary analyses while clearly identifying them as non-core data.

The figure of merit (FOM), a key indicator of spectral fit quality, exhibits variability between probes. In our dataset, FOM values from shear probe sh2 distribute consistently around 1 (Figure 3e), while sh1 presents a bimodal distribution centered near 1 and 1.4. Despite its frequent use, there is no universal consensus on the optimal FOM threshold, as it has been shown to depend on probe characteristics, platform type, and environmental context (Lueck et al., 2024). Although we currently apply the FOM threshold of 1.4 recommended by Lueck et al., this limit is acknowledged within the community to be somewhat subjective. As such, in future applications or targeted analyses, a relaxed QC threshold could be considered. This would allow retention of estimates slightly above the nominal cutoff, provided other quality criteria (e.g., agreement between probes) are satisfied, thereby minimizing unnecessary data loss while maintaining traceability of uncertainty.

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

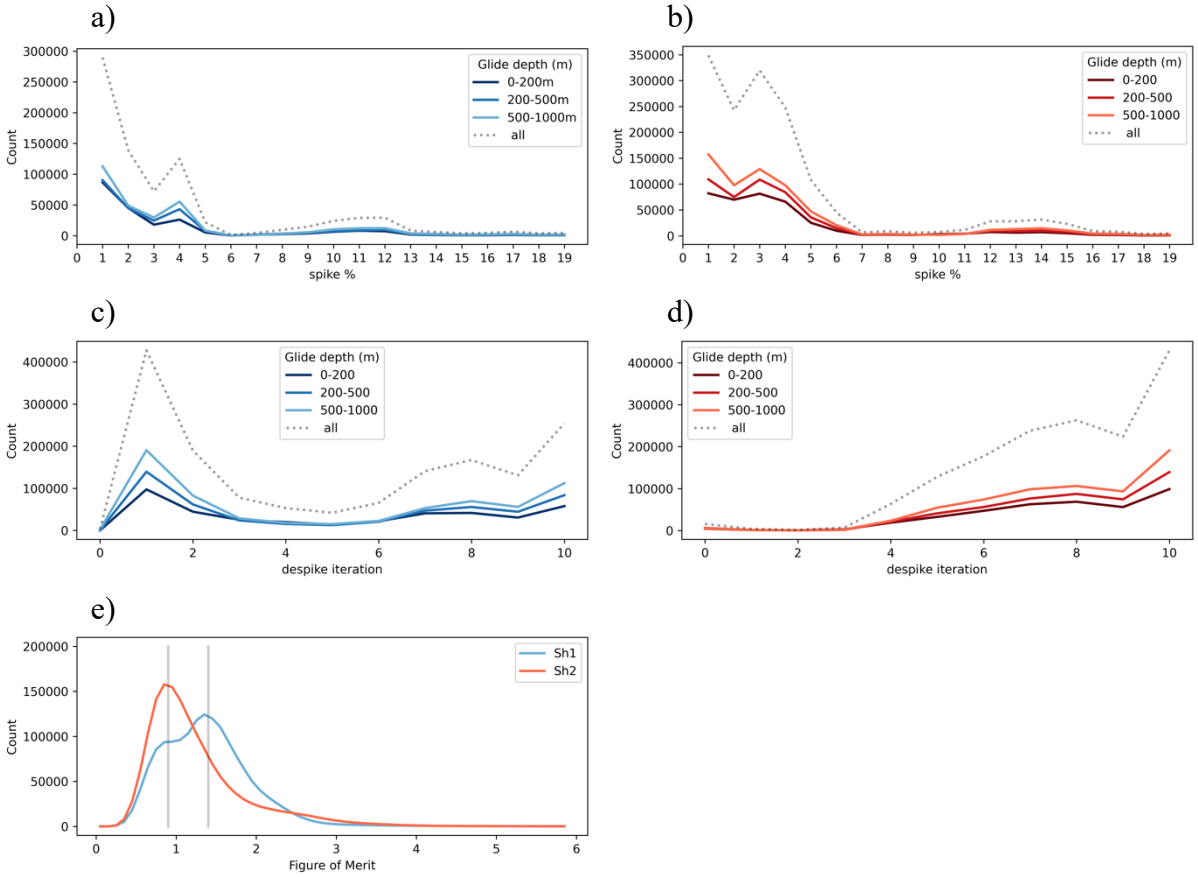


Figure 3: Count of spiking fraction (a,b) and pass count (c,d) for shear 1 (blue) and shear 2 probes (red), by bins of glider depth extension, in meter (i.e. from short to long glide). Count of FOM values (e) for shear 1 (blue) and shear 2 (red).

5.2. QC counting

Figure 4 presents the cross check of shear-based and thermistor-based estimates. We present in Figure 5 the counting of all unique tests we performed on the data (and detailed in Tab. 4). If counting the strict cross-checked cases when both sensors of shear and thermistor pass QC=0, around 10% of good estimates remain for shears, against 16% for thermistor. Employed singularly, sensors reach 28% and 31% for shear 1 and 2, respectively, and 19% and 20% for FP07 1 and 2, respectively. These estimates alone are less affixable then the cross-checked but can be employed with caution by the user in a contextual use.

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

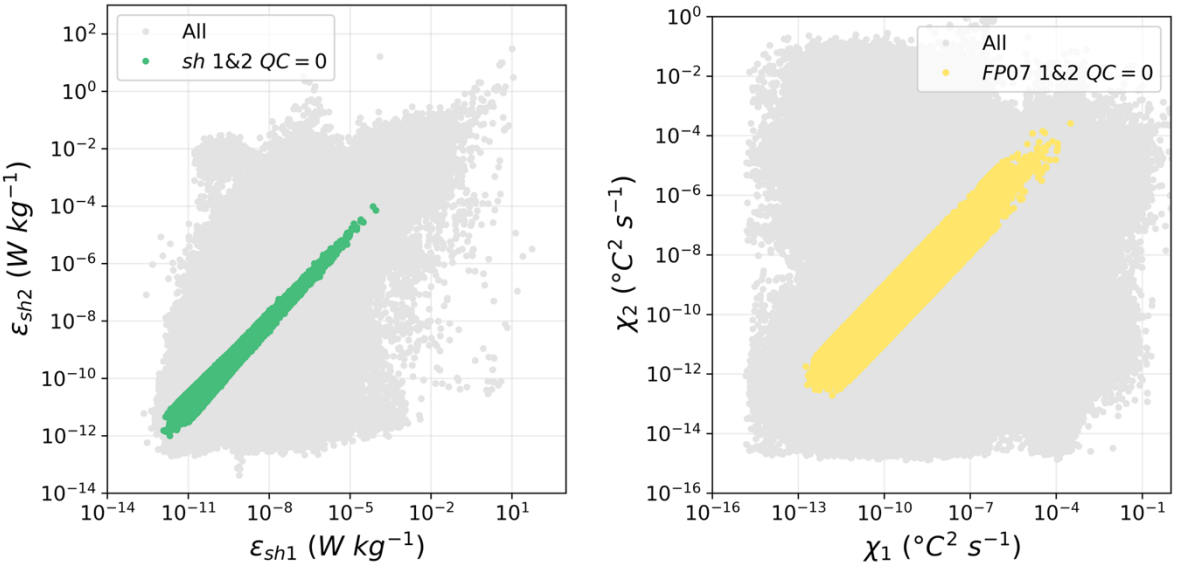


Figure 4 : Scatter plots of cross-checked estimates of (ε_1,ε_2) and (χ_1, χ_2). Gray points indicate all estimates; (Left) green indicates the (best) cross-probe choice; (Right) yellow indicates that both FP07 estimates pass.

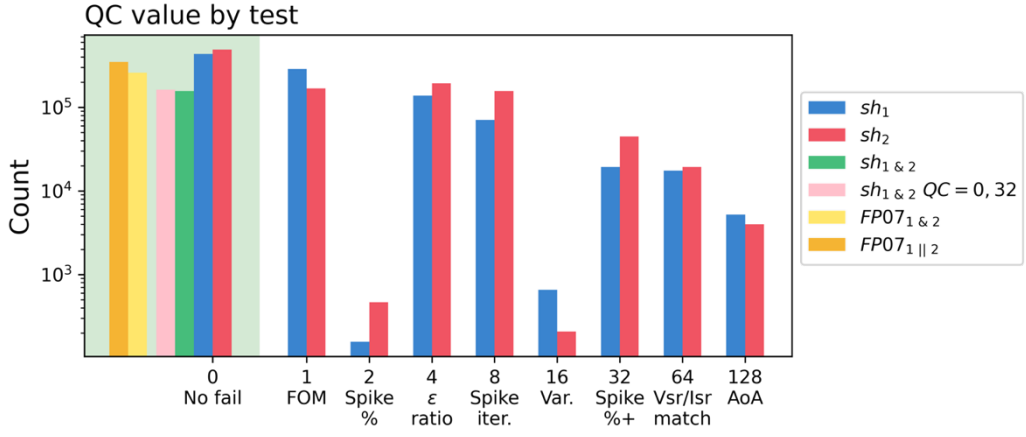


Figure 5: Count of the primary QC flags. Blue and red respectively refer to sh1, sh2 probes. Green indicates the (best) cross-probe choice and pink as a secondary cross-probe choice considering a relaxed despiking fraction criteria. Yellow indicates that both FP07 estimates pass, orange indicates that one on two passes. Cumulative/combination of flags is not shown.

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

600 **Table 4: Count of total and QC-passing estimates.**

Choice	Subset	Count	%
All	sh1,2	1551524	100.00
QC = 0	sh1,2 cross-checked	157039	10.12
QC = 32	sh1,2 cross-checked	5400	0.35
QC = 0,32	sh1,2 cross-checked	162439	10.47
QC = 0	sh1 alone	435071	28.04
QC = 0	sh2 alone	490674	31.63
Choice	**Subset**	**Count**	**%**
QC = 0	fp07 1,2	1548828	100.00
QC = 0	fp07 1,2 combined	259086	16.73
QC = 0	fp07 one among 1,2	350169	22.61
QC = 0	fp07 1 alone	309802	20.00
QC = 0	fp07 2 alone	299453	19.33

5.3. Distributions of QC-passing estimates by temporal and spatial bins

605 Figure 6 shows the distribution of valid estimates at both shear probes, with QC=0 in green and QC=0 or 32 in pink. Estimates with QC=0 for both FP07 thermistors are shown in yellow, while values valid for only one thermistor are shown in orange. The yearly counts reflect progressive improvements in deployment and technical reliability, with 2015 and 2017 representing early testing phases. Seasonal coverage is densest from late spring to mid-autumn, although data from late winter and early spring are 610 also included. In terms of spatial distribution, the core of valid estimates is centered around 500 meters depth, with a higher density on the eastern side of the section.

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

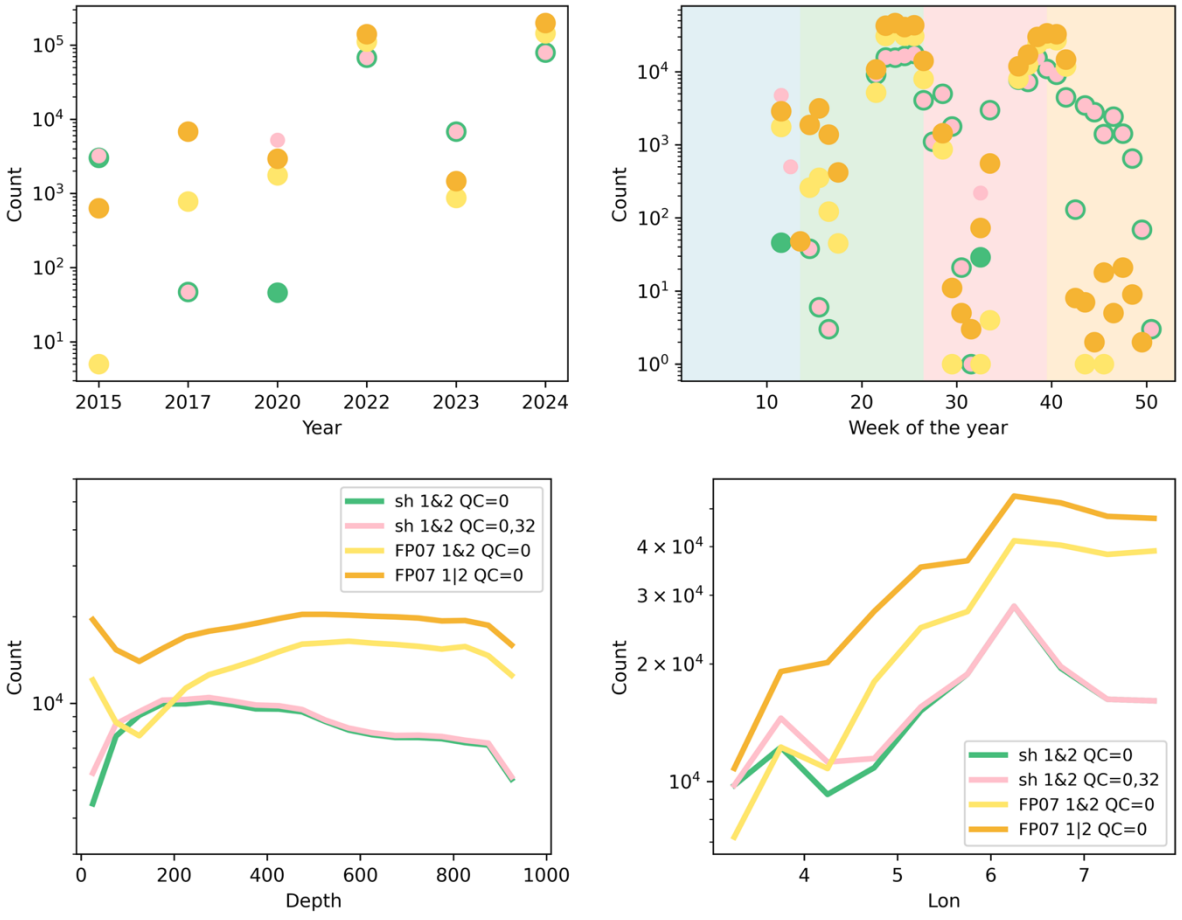


Figure 6: Count of passing data by bins of: Years, week of the year, depth, and longitude. Green indicates the (best) cross-probe choice for shears and pink as a secondary cross-probe choice considering a relaxed despiking fraction criteria. Yellow indicates that both FP07 estimates pass; orange one on two.

5.4. Synthetical averages of estimates across the longitudinal section, and their associated distributions

In Figure 7 we present averaged cross-section of dissipation rate (ε) from shear probes (top) and thermal variance dissipation rate (χ) from FP07 thermistors (bottom), filtered for quality flag QC = 0 (an overview of individual estimates by sensors and years can be consulted in Supplementary Figure 2). Both sections highlight the vertical and horizontal distribution of turbulent mixing across the Balearic–Sardinian transect. The ε field reveals distinct near-surface intensification associated with the Atlantic Water (AW) layer, while subsurface peaks are also visible, notably in the Eastern Intermediate Water (EIW) core and near the top of the Western Mediterranean Deep Water (WMDW). Noteworthy enhancements are observed between 4.5°E and 5°E in the upper 500 m, and between 6°E and 7°E from 500 to 1000 m depth, possibly indicating frontal activity or internal wave breaking. These signals are also

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

evident in χ, which highlights the role of isopycnal exchanges. Around 2.5°E to 3.5°E, the transect crosses the diagonal connecting southern Balearic waters to Mallorca, where turbulence appears locally enhanced. A relatively quiescent band between 3.5°E and 4.5°E separates this region from the central part of the transect, possibly reflecting a dynamical transition zone.

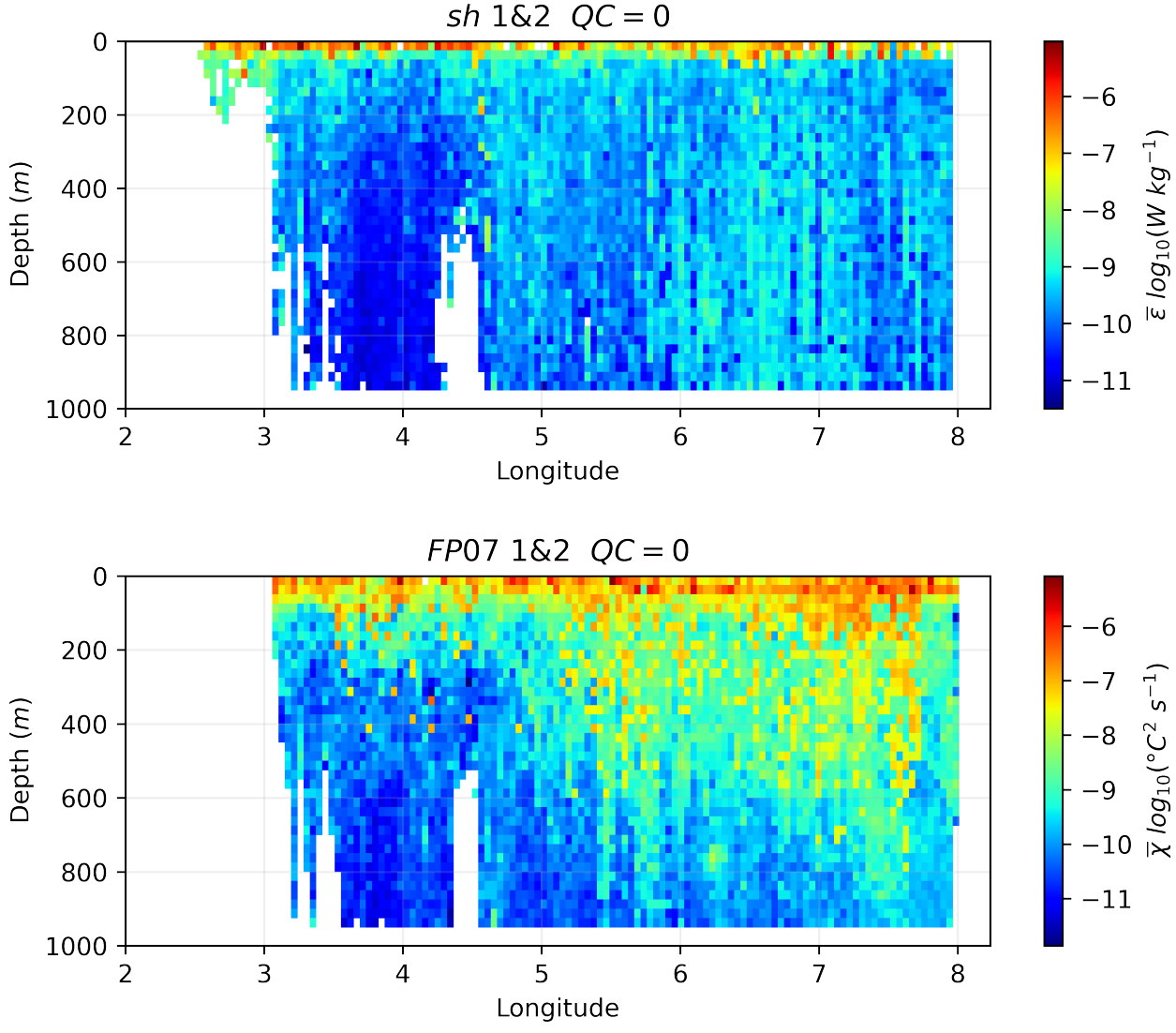
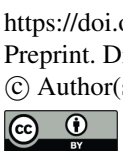


Figure 7: Synthetic sections of all good estimates of ε and χ (all years considered) averaged by depth bins of 25m, and longitudinal bins of 0.045° (circa 5 km).

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

In Figure 8 we present the probability density functions (PDFs) of ε and χ after grouping the estimates by layer, longitude, and season. In the vertical decomposition, upper layers show a progressively broader distribution extending toward more intense ε values. The longitudinal breakdown reveals generally similar distributions across the section, with a tendency for higher intensities on the eastern side. Seasonal decomposition shows distinct patterns: spring is relatively weak, centered between −10.5 and −11; summer displays a bimodal distribution, with one mode near −11 and another near −9.5; autumn shows a similar bimodal structure, with the lower mode shifted toward −10.5; and winter is marked by high ε values, typically between −7 and −8. These patterns support the interpretation that summer and autumn capture intermittent turbulent events under strong stratification, which tends to suppress vertical motion. In contrast, winter conditions favor more frequent and intense turbulence, driven by convection in a well-mixed water column.

For χ, the seasonal pattern is reversed. Spring exhibits the most intense distributions, which progressively shift toward weaker values in summer, autumn, and finally winter. Since χ reflects mixing along isopycnals, it provides a clearer signature of stratification-controlled processes. The decline of χ through the seasons could correspond to the gradual weakening of stratification and the reduced role of isopycnal stirring.

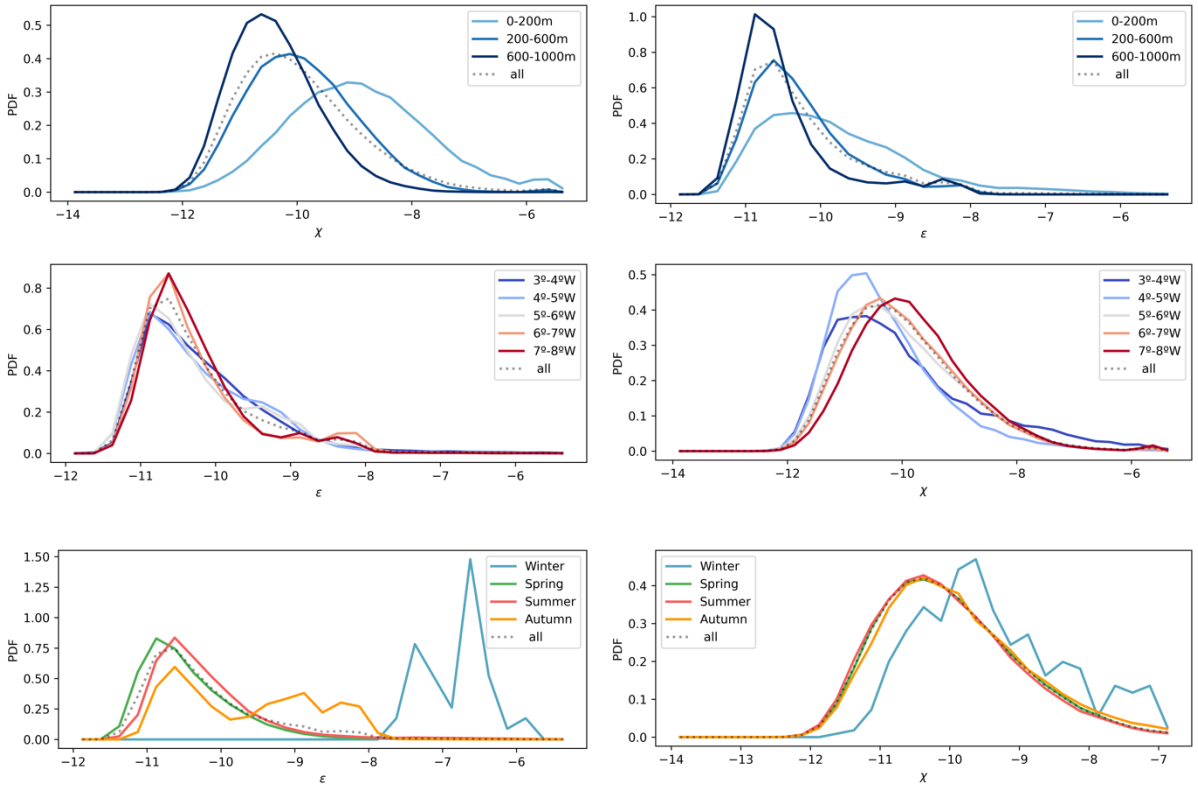


Figure 8: Normalized probability distribution functions (PDF) of cross-checked passing (QC=0) estimates of ε (left) and χ (right) binned by layers, longitudinal fractions and seasons.

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

6. Conclusions

Quality control flags were assigned based on a combination of objective spectral diagnostics and signal characteristics. Estimates flagged as QC=0 represent fully validated data. These spectra passed all core quality criteria, including a good spectral fit (figure of merit ≤ 1.4), minimal spike contamination, inter-probe agreement within expected uncertainty, and sufficient resolved variance. This subset forms the core of the L4 product and can be used with high confidence in scientific analyses.

The broader category of QC ≤ 1 includes both QC=0 and QC=1. While QC=0 estimates are fully valid, QC=1 flags estimates with a figure of merit exceeding 1.4, indicating that the observed spectrum deviates from the theoretical model (e.g., Nasmyth or Lueck), even it is acknowledged by the community that there is no consensual value established for the FOM as it can be subjective to location and platform (Lueck et al. 2024).

Flags with QC > 2 indicate failure of one or more critical quality criteria. Common combinations include significant spike contamination (QC=2 or 8), disagreement between probes (QC=4), or poor resolution of the turbulent variance (QC=16). These estimates are not suitable for general use and should either be discarded or considered only after manual inspection and context-specific evaluation.

A special consideration is given to spike-related quality flags, particularly QC=2 and QC=32. The QC=2 flag is assigned when more than 5 percent of data points are affected by spikes, despite otherwise valid spectral features and probe agreement. Such spiking may arise from biological interference, local mechanical effects, or valid turbulence partially obscured by benign outliers. In addition, the despiking iteration count flag, QC=8, is triggered when more than nine despiking passes are needed. However, the literature offers no consensus on acceptable iteration limits (Lueck et al., 2024), and the ODAS routine caps iteration counts at ten, which limits the interpretive strength of this metric. To recover meaningful data in long profiles where spike accumulation may be more likely, we introduce a relaxed quality flag, QC=32, for cases where the spike fraction falls between 5 and 15 percent and iteration counts remain below the threshold. This intermediate flag acknowledges the potential value of these estimates, especially given that some profiles exceed 1000 meters in length. Applying a fixed percentage criterion across records of varying length may unfairly penalize longer profiles, which are more prone to encounter localized spikes. QC=32 thus represents a compromise: it eventually allows discarded valid data to be included in secondary analyses, without misrepresenting them as core-quality estimates.

In practice, our decision framework recommends using QC=0 as the primary dataset for scientific interpretation. QC=32 may be optionally included in context-aware analyses, particularly where spatial or seasonal completeness is important. When examining data from individual probes (e.g., shear1 or shear2 separately), QC=0 may also be retained under specific conditions, although the absence of cross-validation should be considered. All other QC categories should be excluded from core analyses or subjected to manual review depending on the use case.

This dataset represents one of the most comprehensive glider-based microstructure records collected in the Western Mediterranean to date, spanning nearly a decade from 2015 to 2024. Over this period, deployment protocols and sensor performance progressively improved, resulting in a steady increase in data quality and reliability. The dataset captures multiple seasonal cycles, with denser coverage from late spring through autumn and additional profiles obtained during late winter and early

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

spring. This seasonal breadth allows for the exploration of variability in turbulent mixing under contrasting stratification regimes. By repeatedly surveying a fixed transect between Sardinia and the Balearic Islands, the dataset also provides high-resolution insights into a key hydrographic boundary. This repeated coverage of a dynamic interface aligns with the objectives of the Ocean Gliders Program (Testor et al. 2019), supporting long-term, fine-scale observation of boundary currents, water mass transformation, and vertical mixing in a region of both regional and basin-scale importance.

Despite the strengths of the dataset, several limitations highlight the technical and methodological challenges associated with autonomous turbulence observations. While raw sampling potentially yielded millions of measurement points, strict quality control procedures narrowed the dataset to a validated subset of only tens of thousands of points. This reduction by factors of 100 to 1000 underscores the sensitivity of microstructure measurements to sensor stability, platform dynamics, and environmental conditions. Profiles affected by low incident velocity, strong glider pitch, or localized contamination were routinely excluded. Future improvements could include enhanced real-time monitoring of glider flight and sensor performance, allowing more adaptive sampling strategies. Transmission of diagnostic metadata in near-real time could help identify problematic segments while missions are underway. Additionally, further development of onboard processing and storage, combined with cost-effective and robust sensor solutions, could significantly increase the volume of usable turbulence estimates. The dataset confirms the added value of gliders for observing ocean turbulence, especially when equipped with microstructure payloads such as the Micro Rider. Unlike traditional platforms, gliders deliver continuous, high-resolution vertical sections over hundreds of kilometers without requiring ship support. Their autonomous operation makes them suitable for deployment in remote or challenging environments, offering a sustained observational presence that complements episodic ship-based surveys. Gliders bridge the gap between point measurements from moorings and broad-scale CTD transects and enable four-dimensional views of the ocean when biogeochemical and optical sensors are integrated. Importantly, this dataset demonstrates that key turbulence variables such as ε and χ have reached a level of maturity where they can be regularly retrieved, quality-controlled, and used in scientific and operational contexts. Their inclusion as Essential Ocean Variables is increasingly feasible, with implications for ocean mixing parameterizations, biogeochemical fluxes, and model development (Aydogdu et al. 2025). The approaches documented here can inform broader integration into GOOS, Copernicus, and potentially future Argo extensions, particularly as sensor miniaturization and cost reduction continue to expand the accessibility of turbulence-resolving platforms.

7. Acknowledgments

We thank all the technicians and engineers at SOCIB (Balearic Islands Coastal Observing and Forecasting System, Palma, Spain) and Rockland Scientific for their essential work in the deployment operations and data support throughout the various mission years. F.K. is supported by ITINERIS project, funded by EU - Next Generation EU Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures” - Project IR0000032 – ITINERIS - Italian Integrated

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

Environmental Research Infrastructures System - CUP B53C22002150006. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them. English grammar and phrasing were partially revised using AI-based language tools to improve clarity and readability.

8. Data availability

Turbulence microstructure dataset from Slocum Glider Teresa (Western Mediterranean, 2015–2024) is available from SEANOE at: https://doi.org/10.17882/107995 (Kokoszka et al., 2025a). The glider data supporting the microstructure processing is available at SOCIB's ERDDAP/THREDDS server (https://thredds.socib.es/thredds/catalog/auv/glider/teresa-cnr_teresa/catalog.html).

9. Code availability

The MATLAB and Python code used for data processing and quality control is archived on Zenodo at: https://doi.org/10.5281/zenodo.16541936 (Kokoszka et al., 2025b)
A public notebook to read the data and produce the figures from can be consulted on https://colab.research.google.com/drive/1qsN8n68C3FBiFkfGt32MGPsthtqM-PmU?usp=sharing.

10. Author contributions

The contribution of each author is specified below according to the CRediT taxonomy (Contributor Roles Taxonomy). Conceptualization: FVMK, JT. Data curation: FVMK, MB, AM, PRR, MR, MC, BC. Formal analysis: FVMK. Funding acquisition: MB, KS, JC, JT, NZ. Investigation: FVMK, ATD, . Methodology: FVMK, ATD. Project administration: MB, KS, JC, JT, NZ. Resources: FVMK, JT, NZ, AM, PRR, MR, MC, BC. Software: FVMK, JT, NZ, ATD. Supervision: MB, KS, ATD. Validation: FVMK, ATD. Visualization: FVMK. Writing (original draft): FVMK. Writing (review and editing): FVMK, KS, JC, JT, NZ, ATD.

11. Competing interests

The authors declare that they have no conflict of interest.

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

References

Aulicino, G., Cotroneo, Y., Ruiz, S., Sánchez Román, A., Pascual, A., Fusco, G., Tintoré, J., and Budillon, G.: Monitoring the Algerian Basin through glider observations, satellite altimetry and numerical simulations along a SARAL/AltiKa track, Journal of Marine Systems, 179, 55–71, https://doi.org/10.1016/j.jmarsys.2017.11.006, 2018.

Aydogdu, A., Escudier, R., Hernandez-Lasheras, J., Amadio, C., Pistoia, J., Zarokanellos, N. D., Cossarini, G., Remy, E., and Mourre, B.: Glider observations in the Western Mediterranean Sea: their assimilation and impact assessment using four analysis and forecasting systems, Front. Mar. Sci., 12, 1456463, https://doi.org/10.3389/fmars.2025.1456463, 2025.

Batchelor, G. K.: Small-scale variation of convected quantities like temperature in turbulent fluid Part 1. General discussion and the case of small conductivity, J. Fluid Mech., 5, 113–133, https://doi.org/10.1017/S002211205900009X, 1959.

Chiggiato, J., Artale, V., Durrieu De Madron, X., Schroeder, K., Taupier-Letage, I., Velaoras, D., and Vargas-Yáñez, M.: Recent changes in the Mediterranean Sea, in: Oceanography of the Mediterranean Sea, Elsevier, 289–334, https://doi.org/10.1016/B978-0-12-823692-5.00008-X, 2023.

Eriksen, C. C., Osse, T. J., Light, R. D., Wen, T., Lehman, T. W., Sabin, P. L., Ballard, J. W., and Chiodi, A. M.: Seaglider: a long-range autonomous underwater vehicle for oceanographic research, IEEE J. Oceanic Eng., 26, 424–436, https://doi.org/10.1109/48.972073, 2001.

Fer, I., Dengler, M., Holtermann, P., Le Boyer, A., and Lueck, R.: ATOMIX benchmark datasets for dissipation rate measurements using shear probes, Sci Data, 11, 518, https://doi.org/10.1038/s41597-024-03323-y, 2024.

Goodman, L., Levine, E. R., and Lueck, R. G.: On Measuring the Terms of the Turbulent Kinetic Energy Budget from an AUV, Journal of Atmospheric and Oceanic Technology, 23, 977–990, https://doi.org/10.1175/JTECH1889.1, 2006.

Kokoszka, F. V. M., Borghini, M., Schroeder, K., Chiggiato, J., Tintoré, J., Zarokanellos, N., Miralles, A., Rivera Rodríguez, P., Rubio, M., Charcos, M., Casas, B., and Ten Doeschate, A.: Turbulence microstructure dataset from Slocum Glider Teresa (Western Mediterranean, 2015–2024) (1), https://doi.org/10.17882/107995, 2025a.

Kokoszka, F. V. M.: Teresa MR processing code, , https://doi.org/10.5281/ZENODO.16541936, 2025b.

Kraichnan, R. H.: Small-Scale Structure of a Scalar Field Convected by Turbulence, The Physics of Fluids, 11, 945–953, https://doi.org/10.1063/1.1692063, 1968.

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

Le Boyer, A., Couto, N., Alford, M. H., Drake, H. F., Bluteau, C. E., Hughes, K. G., Naveira Garabato, A. C., Moulin, A. J., Peacock, T., Fine, E. C., Mashayek, A., Cimoli, L., Meredith, M. P., Melet, A., Fer, I., Dengler, M., and Stevens, C. L.: Turbulent diapycnal fluxes as a pilot Essential Ocean Variable, Front. Mar. Sci., 10, 1241023, https://doi.org/10.3389/fmars.2023.1241023, 2023.

Lindstrom, E., Gunn, J., Fischer, A., McCurdy, A., Glover, L. K., and Members, T. T.: A Framework for Ocean Observing, European Space Agency, https://doi.org/10.5270/OceanObs09-FOO, 2012.

Lueck, R., Murowinski, E., and McMillan, J.: RSI Technical Note 039: A Guide To Data Processing; Rockland Scientific International Inc.: Victoria, BC, Canada, 2016, Rockland Scientific International Inc., 2020.

Lueck, R., Fer, I., Bluteau, C., Dengler, M., Holtermann, P., Inoue, R., LeBoyer, A., Nicholson, S.-A., Schulz, K., and Stevens, C.: Best practices recommendations for estimating dissipation rates from shear probes, Front. Mar. Sci., 11, 1334327, https://doi.org/10.3389/fmars.2024.1334327, 2024.

Lueck, R. G., Wolk, F., and Yamazaki, H.: Oceanic Velocity Microstructure Measurements in the 20th Century, Journal of Oceanography, 58, 153–174, https://doi.org/10.1023/A:1015837020019, 2002.

Margirier, F., Testor, P., Heslop, E., Mallil, K., Bosse, A., Houpert, L., Mortier, L., Bouin, M.-N., Coppola, L., D'Ortenzio, F., Durrieu De Madron, X., Mourre, B., Prieur, L., Raimbault, P., and Taillandier, V.: Abrupt warming and salinification of intermediate waters interplays with decline of deep convection in the Northwestern Mediterranean Sea, Sci Rep, 10, https://doi.org/10.1038/s41598-020-77859-5, 2020.

Merckelbach, L., Berger, A., Krahmann, G., Dengler, M., and Carpenter, J. R.: A Dynamic Flight Model for Slocum Gliders and Implications for Turbulence Microstructure Measurements, Journal of Atmospheric and Oceanic Technology, 36, 281–296, https://doi.org/10.1175/JTECH-D-18-0168.1, 2019.

Nasmyth, P. W.: Oceanic turbulence, https://doi.org/10.14288/1.0302459, 2011.

Osborn, T. R. and Crawford, W. R.: An Airfoil Probe for Measuring Turbulent Velocity Fluctuations in Water, in: Air-Sea Interaction, edited by: Dobson, F., Hasse, L., and Davis, R., Springer US, Boston, MA, 369–386, https://doi.org/10.1007/978-1-4615-9182-5_20, 1980.

Peterson, A. K. and Fer, I.: Dissipation measurements using temperature microstructure from an underwater glider, Methods in Oceanography, 10, 44–69, https://doi.org/10.1016/j.mio.2014.05.002, 2014.

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

Piccolroaz, S., Fernández-Castro, B., Toffolon, M., and Dijkstra, H. A.: A multi-site, year-round turbulence microstructure atlas for the deep perialpine Lake Garda, Sci Data, 8, 188, https://doi.org/10.1038/s41597-021-00965-0, 2021.

Schroeder, K., Chiggiato, J., Bryden, H. L., Borghini, M., and Ben Ismail, S.: Abrupt climate shift in the Western Mediterranean Sea, Sci Rep, 6, 23009, https://doi.org/10.1038/srep23009, 2016.

Schroeder, K., Ben Ismail, S., Bensi, M., Bosse, A., Chiggiato, J., Civitarese, G., Falcieri M, F., Fusco, G., Gačić, M., Gertman, I., Kubin, E., Malanotte-Rizzoli, P., Martellucci, R., Menna, M., Ozer, T., Taupier-Letage, I., Vargas-Yáñez, M., Velaoras, D., and Vilibić, I.: A consensus-based, revised and comprehensive catalogue for Mediterranean water masses acronyms, Mediterr. Mar. Sci., 25, 783–791, https://doi.org/10.12681/mms.38736, 2024.

Sherman, J. T. and Davis, R. E.: Observations of Temperature Microstructure in NATRE, J. Phys. Oceanogr., 25, 1913–1929, https://doi.org/10.1175/1520-0485(1995)025<1913:OOTMIN>2.0.CO;2, 1995.

Testor, P., Send, U., Gascard, J. -C., Millot, C., Taupier-Letage, I., and Béranger, K.: The mean circulation of the southwestern Mediterranean Sea: Algerian Gyres, J. Geophys. Res., 110, 2004JC002861, https://doi.org/10.1029/2004JC002861, 2005.

Testor, P., Bosse, A., Houpert, L., Margirier, F., Mortier, L., Legoff, H., Dausse, D., Labaste, M., Karstensen, J., Hayes, D., Olita, A., Ribotti, A., Schroeder, K., Chiggiato, J., Onken, R., Heslop, E., Mourre, B., D'ortenzio, F., Mayot, N., Lavigne, H., De Fommervault, O., Coppola, L., Prieur, L., Taillandier, V., Durrieu De Madron, X., Bourrin, F., Many, G., Damien, P., Estournel, C., Marsaleix, P., Taupier-Letage, I., Raimbault, P., Waldman, R., Bouin, M., Giordani, H., Caniaux, G., Somot, S., Ducrocq, V., and Conan, P.: Multiscale Observations of Deep Convection in the Northwestern Mediterranean Sea During Winter 2012–2013 Using Multiple Platforms, JGR Oceans, 123, 1745–1776, https://doi.org/10.1002/2016JC012671, 2018.

Testor, P., De Young, B., Rudnick, D. L., Glenn, S., Hayes, D., Lee, C. M., Pattiaratchi, C., Hill, K., Heslop, E., Turpin, V., Alenius, P., Barrera, C., Barth, J. A., Beaird, N., Bécu, G., Bosse, A., Bourrin, F., Brearley, J. A., Chao, Y., Chen, S., Chiggiato, J., Coppola, L., Crout, R., Cummings, J., Curry, B., Curry, R., Davis, R., Desai, K., DiMarco, S., Edwards, C., Fielding, S., Fer, I., Frajka-Williams, E., Gildor, H., Goni, G., Gutierrez, D., Haugan, P., Hebert, D., Heiderich, J., Henson, S., Heywood, K., Hogan, P., Houpert, L., Huh, S., E. Inall, M., Ishii, M., Ito, S., Itoh, S., Jan, S., Kaiser, J., Karstensen, J., Kirkpatrick, B., Klymak, J., Kohut, J., Krahmann, G., Krug, M., McClatchie, S., Marin, F., Mauri, E., Mehra, A., P. Meredith, M., Meunier, T., Miles, T., Morell, J. M., Mortier, L., Nicholson, S., O'Callaghan, J., O'Conchubhair, D., Oke, P., Pallàs-Sanz, E., Palmer, M., Park, J., Perivoliotis, L., Poulain, P.-M., Perry, R., Queste, B., Rainville, L., Rehm, E., Roughan, M., Rome, N., Ross, T., Ruiz, S., Saba, G., Schaeffer, A., Schönau, M., Schroeder, K., Shimizu, Y., Sloyan, B. M., Smeed, D., Snowden, D., Song, Y., Swart, S., Tenreiro, M., Thompson, A., Tintore, J., Todd, R. E., Toro, C.,

https://doi.org/10.5194/essd-2025-451
Preprint. Discussion started: 13 August 2025
© Author(s) 2025. CC BY 4.0 License.

Venables, H., Wagawa, T., et al.: OceanGliders: A Component of the Integrated GOOS, Front. Mar. Sci., 6, 422, https://doi.org/10.3389/fmars.2019.00422, 2019.

TESTOR Pierre: MOOSE-GE, https://doi.org/10.18142/235, 2010.

Wolk, F., Lueck, R. G., and St. Laurent, L.: Turbulence measurements from a glider, in: OCEANS 2009, OCEANS 2009, Biloxi, MS, 1–6, https://doi.org/10.23919/OCEANS.2009.5422413, 2009.
Woods Hole Oceanographic Institution, St. Laurent, L., and Merrifield, S.: Measurements of Near-Surface Turbulence and Mixing from Autonomous Ocean Gliders, Oceanog., 30, 116–125, https://doi.org/10.5670/oceanog.2017.231, 2017.