Applied Acoustics 246 (2026) 111239

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

The DeuteroNoise dataset: An open, calibrated, multi-basin resource for vessel noise and natural soundscapes in European coastal waters

Ignasi Nou-Plana [a], Adrian Teaca [b], Giovanni Zambon [c], Rosa Ma Alsina-Pagès [a,*]

[a] Human Environment Research (HER), La Salle - Universitat Ramon Llull, Carrer de Sant Joan de la Salle, 42, Barcelona, 08022, Catalunya, Spain
[b] National Institute for Research and Development on Marine Geology and Geoecology - GeoEcoMar, Bulevardul Mamaia 304, Constanţa, 900581, Romania
[c] Department of Earth and Environmental Sciences, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, Milan, 20126, Italy

HIGHLIGHTS

- Open-access dataset of vessel noise and natural soundscapes in three European basins.
- Nearly 700 h of calibrated hydrophone recordings paired with AIS vessel metadata.
- Enables robust cross-regional comparisons and ecological assessments of noise impacts.
- Scalable PostgreSQL-FastAPI platform with interactive retrieval and exploration tools.
- Supports AI-based classification models for vessel-type identification and soundscape analysis.

ARTICLE INFO

Keywords:
Underwater acoustics
Vessel noise
Anthropogenic noise pollution
Underwater soundscape monitoring
AIS data integration
Machine learning classification

ABSTRACT

Understanding underwater soundscapes is essential for assessing the impacts of maritime noise pollution on marine environments. Characterizing these soundscapes is crucial both for evaluating current noise impacts and for developing future mitigation strategies. Several datasets have been published focusing on the acoustic signatures of common vessel types; however, most remain restricted to a single location, are not fully open to the public, and lack scalable storage or dissemination tools. The DeuteroNoise Dataset addresses these gaps; an open-access corpus that pairs long-duration, calibrated hydrophone recordings with time-synchronised Automatic Identification System data to document coastal vessel noise and contrasting natural soundscapes at three European sites: the Catalan coast, the Venice Lagoon, and the western Black Sea. Six short-term fixed-station campaigns conducted since December 2023 have produced nearly 700 h of publicly available continuous audio with more than 11 h of labelled audio. Each recorded and identified event is correlated with the ship's identity, position, and speed metadata; the dataset therefore spans cargo vessel traffic, workboats, leisure craft, and non-anthropogenic background sounds. Vessel types are categorized and linked to their acoustic signatures, facilitating analyses of soundscape dynamics and ecological impact. Built on PostgreSQL with a FastAPI backend, and served through an interactive web interface, the dataset offers a scalable platform for large-scale retrieval and exploration. By integrating calibrated, multi-basin recordings with vessel metadata in an openly accessible, scalable framework. The DeuteroNoise Dataset represents the first resource of its kind in Europe, enabling robust cross-regional comparisons, supporting the development of AI-based classification models, advancing ecological research on anthropogenic noise, and setting a new benchmark for underwater soundscape monitoring worldwide.

1. Introduction

Underwater environments are never silent. Marine animals rely on sound for communication, navigation, feeding, and predator avoidance, and even small changes in the acoustic background can alter their behaviour or stress levels [1,2]. Over recent decades, coastal waters have become increasingly noisy due to growing vessel activity, yet many regions remain poorly documented at the soundscape level, even as new European directives on underwater noise are being established. For research fields such as ecological monitoring, bioacoustics, and impact assessment, the lack of calibrated and comparable recordings makes it

* Corresponding Author.
Email addresses: ignasi.nou@salle.url.edu (I. Nou-Plana), rosamaria.alsina@salle.url.edu (R.M. Alsina-Pagès).

https://doi.org/10.1016/j.apacoust.2026.111239
Received 20 October 2025; Received in revised form 9 December 2025; Accepted 16 January 2026
Available online 28 January 2026
0003-682X/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

difficult to understand, for example, how noise levels vary across locations, seasons, and habitat types. Interest in underwater acoustics from anthropogenic sources has increased since the early 20th century, coinciding with the introduction of active sonar technology [3]. Initially tailored for military applications, such as anti-submarine warfare and fleet navigation, sonar systems have progressively broadened their scope to address civilian applications, including maritime traffic management, fisheries monitoring, and marine environmental protection [4,5]. This has been paralleled by the marked increase in global shipping and density of coastal vessel traffic [6], intensifying both scientific and regulatory attention to underwater noise generated by vessels as a significant marine pollutant. Ecological studies over recent decades have extensively documented how both continuous and impulsive underwater noise might affect marine vertebrates, particularly whales and dolphins [7]. Invertebrates, on the other hand, such as marine deuterostomes like tunicates, remain underrepresented in acoustic ecological research, and initiatives such as the DeuteroNoise project [4] have recently begun to address this gap.

Despite substantial progress in understanding the biological effects of underwater noise, critical gaps persist in the characterisation of marine soundscapes at regional and continental scales. First, it remains unclear how baseline acoustic conditions vary across distinct hydrogeographic settings, since most existing datasets are restricted to a single site and cannot be compared due to differences in equipment, calibration, or deployment protocols. Second, the ways in which vessel traffic patterns, including vessel type, density, and movement dynamics, shape low frequency sound levels are not well quantified outside a few intensively monitored regions. Third, there is limited knowledge of how natural and anthropogenic contributions to the soundscape fluctuate across field and seasonal cycles, particularly in shallow coastal habitats where many ecologically important organisms reside. Finally, without long duration recordings paired with reliable vessel metadata, publicly available, it is difficult to derive baseline indicators required for ecological exposure experiments and for regulatory assessments.

Sometimes, data might be constrained by military purposes or geographically limited. This research aims to deliver a comprehensive, open-access resource to support robust comparative analyses, enhance regional acoustic modelling efforts, and promote strategies to mitigate underwater noise pollution, aligning with international regulatory frameworks such as the Marine Strategy Framework Directive (MSFD) Descriptor 11 [5].

The study builds upon previous work that characterised vessel noise and soundscapes in several European basins, within the framework of the DeuteroNoise JPI Oceans project (NoiseInTheSea- 2022–0011), which aims to establish a shared European infrastructure for characterizing and mitigating underwater noise pollution. The dataset and analyses presented here build upon our previous contributions that have explored regional case studies on the Catalan coast and Port of Barcelona [8], introduced the first releases of the DeuteroNoise Dataset [9], and demonstrated its potential for open-access soundscape exploration across multiple European basins [10]. Moreover, the project has emphasized the ecological implications of vessel noise on marine invertebrate deuterostomes [4], thus broadening the scope of bioacoustics impact assessments beyond traditionally studied vertebrates. By consolidating these efforts, our goal is to provide a robust scientific foundation and open-access tools that support regulatory needs and foster collaborative research on anthropogenic noise in European seas.

This study makes four scientific contributions. First, it provides the first calibrated multi basin underwater acoustic dataset covering contrasting European coastal environments, enabling cross regional soundscape comparison under a unified measurement protocol. Second, it introduces a standardized and fully reproducible acquisition methodology, integrating hydrophone calibration, controlled mooring deployment, and synchronised AIS tracking, which addresses long standing limitations in comparability across existing datasets [11]. Third, it quantifies basin specific differences in vessel traffic composition, passage characteristics, and low frequency SPL patterns, demonstrating how regional maritime conditions can shape underwater soundscapes. Fourth, it delivers an open data infrastructure and API that allow reproducible retrieval, analysis, and extension of the dataset, supporting ecological modelling, regulatory assessment, and machine learning research on underwater noise.

The subsequent sections of this paper provide a detailed overview of the state of the art in underwater acoustic monitoring (Section 2), description of the recording system and methodology employed (Section 3), information about the field campaigns and study sites (Section 4), comprehensive details about the DeuteroNoise Dataset structure and contents (Section 5), key insights derived from the dataset (Section 6), example analyses including Sound Pressure Levels (SPL) profiles and site comparisons (Section 7), and concluding remarks.

2. State of the art

When it comes to vessels, research efforts concentrate on reducing acoustic emissions by examining vessel-radiated noise: enhancing vessel design by optimizing hull structures and propeller geometries to minimize acoustic output [12]; following regulatory requirements for environmental impact under international conventions [13]; and refining predictive acoustic propagation models crucial for risk assessment and ecological management of sensitive marine habitats [14,15]. Advances in computational methodologies, particularly deep learning techniques such as convolutional and residual neural networks [16,17], have significantly enhanced the capacity to analyze and interpret these mentioned vessel acoustic signatures. But these, depend on robust, well-annotated acoustic datasets that represent scenarios from different parts of the world.

Current datasets lack broad ecological relevance and geographical diversity, with examples including region-specific studies from the Santa Barbara Channel [18], vessel recordings at Atlantic Undersea Test and Evaluation Center (AUTEC) in the Bahamas [19], jet-ski activity monitoring [20], and Arctic icebreaker assessments [21]. To address these limitations, recent initiatives have provided publicly accessible datasets such as ShipsEar, comprising around 3 h of recordings across 11 vessel types in the Ría de Vigo [22], DeepShip, which offers roughly 47 h of recordings from a single ocean observatory near Vancouver [23], and VTUAD, consisting of approximately 49 h of short clips [24]. While these resources represent important progress, they remain constrained in duration, ecological representativeness, and geographical coverage. In contrast, the DeuteroNoise Dataset offers nearly 700 h of calibrated underwater recordings collected across three hydrogeographically distinct European basins (Catalan coast, Venice Lagoon, and the Black Sea), paired with synchronised Automatic Identification System (AIS) metadata for more than 500 vessel passages.

2.1. ShipsEar dataset

The ShipsEar dataset is one of the earliest publicly released, annotated collections of underwater vessel noise. It contains 90 recordings with a total duration of approximately three hours and covers eleven vessel types ranging from small boats to large ships. The recordings were gathered under real marine conditions in the Ría de Vigo, northwestern Spain, during autumn 2012 and summer 2013. In addition to the target vessels' acoustic signatures, the dataset includes port background noise, natural ambient sounds, and occasional marine mammal calls, providing a realistic soundscape. In the original ShipsEar study, a Gaussian Mixture Model classifier was introduced to demonstrate the potential of the dataset, and subsequent research employing deep learning methods has achieved high recognition rates, with reported accuracies between 94% and 97%. Despite these contributions, the dataset is geographically restricted to a single ria and has a relatively small duration, which can limit the generalization of models trained on it to other marine environments.

2.2. DeepShip dataset

The DeepShip dataset represents a more recent benchmark (2021) with larger scale and greater vessel diversity. It comprises 613 underwater recordings with a total of about 47 h of audio, covering 265 individual ships that were grouped into four categories: cargo, tanker, passenger, and tug. Recordings were obtained between May 2016 and October 2018 at the Strait of Georgia's Delta Node, near Vancouver, through Ocean Networks Canada observatories. The dataset spans multiple seasons and sea states, capturing a variety of real-world noise conditions, including additional anthropogenic activities and marine life. The initial DeepShip study introduced a separable convolution-based autoencoder model that achieved 77.5% classification accuracy. Subsequent works employing multi-scale feature fusion and attention mechanisms improved recognition performance to around 87.5%. Although DeepShip significantly expands the scale of open vessel-noise data, its recordings are limited to a single observatory node, restricting the geographic diversity of the dataset. Moreover, the available recordings consist only of vessel excerpts rather than the full continuous acoustic data.

2.3. VTUAD dataset

The Vessel Type Underwater Acoustic Dataset (VTUAD) is a very recent open benchmark (2025) that integrates vessel acoustics with environmental data. It contains approximately 175,000 one-second audio clips, equivalent to around 49 h of recordings. The dataset includes four vessel categories — tug, tanker, cargo, and passenger ships — together with an ambient noise class. Data were collected using a distance-stratified design with defined inclusion and exclusion radii. A unique feature of VTUAD is that each audio clip is paired with co-located CTD (conductivity, temperature, depth, salinity) sensor snapshots, which enrich the dataset with environmental propagation context. Baseline convolutional neural networks based on spectrogram inputs achieved around 95% accuracy in within-scenario vessel identification, but performance declined to about 84% across distance bins due to lower signal-to-noise ratios at greater ranges. More recent transformer-based approaches with metadata fusion improved robustness, reaching up to 96% accuracy. Nevertheless, the dataset remains geographically limited to a single site and exhibits sensitivity to propagation effects over distance.

Despite these advances, current resources remain limited in two critical aspects. First, each corpus is limited to one geographical area: ShipsEar samples only the Port of Vigo, DeepShip uses a single Delta node, and VTUAD relies on one shelf mooring. Moreover, the available recordings consist mainly of vessel excerpts rather than continuous acoustic data, limiting their ability to represent the full soundscape. Second, because acquisition hardware and calibration procedures differ across projects, absolute SPL statistics are not directly comparable; researchers must apply normalisations that might modify the real spatial contrasts.

The DeuteroNoise Dataset has been conceived to remove these barriers. Using an identical URec-384k + Aquarian AS-1 front-end, we recorded three hydrogeographically contrasting basins.

3. Recording system, methodology, and study sites

To ensure scientific reproducibility across basins, all recording campaigns followed a standardized methodology consisting of a fixed hardware chain, a calibrated hydrophone system, an identical acquisition script, a controlled mooring configuration, and a synchronised AIS-acoustic workflow. This protocol, applied uniformly at every site, constitutes a repeatable data-collection framework designed to generate comparable soundscape measurements suitable for quantitative acoustic analyses.

3.1. The DeuteroNoise project

The DeuteroNoise project, funded under the JPI Oceans framework, was established to address a central gap in marine acoustic ecology: the limited understanding of how underwater noise pollution from shipping affects invertebrate deuterostomes. These taxa are phylogenetically close to vertebrates and play important ecological roles in marine food webs as suspension feeders and energy carriers from plankton to higher trophic levels [8]. Their relatively tractable biology, both in the laboratory and in situ, makes them ideal model organisms for investigating noise effects on physiology, sensory systems, and behavior.

The project pursues a dual approach. First, it characterizes the soundscapes of multiple European basins affected by maritime traffic, with deployments in the North Adriatic Sea, the Lagoon of Venice, the Black Sea, the North Sea, and the Catalan coast of the western Mediterranean. In each basin, paired sites were selected to represent contrasting conditions: one heavily impacted by vessel traffic and one comparatively quiet, allowing direct comparisons of anthropogenic and natural soundscapes. Second, the project investigates biological responses by exposing selected invertebrates to both field-recorded and laboratory-simulated noise. By integrating calibrated hydrophone recordings, real-time AIS vessel tracking, and laboratory assays, DeuteroNoise aims to produce the dataset presented in this study, which not only quantifies noise levels across European coastal environments but also links them to ecological impact.

The following section summarises the workflow followed in every DeuteroNoise campaign, related to the hydrophone deployment and the AIS acquisition of the data. The protocol was already reported in earlier conference papers [10] and was kept identical across all basins to ensure data comparability.

3.2. Hydrophone

- Recorder-hydrophone unit: All sites were monitored with autonomous uRec-384k recorders (Nauta-RCS, Italy) coupled to a calibrated AS-1 omni-directional hydrophone (Aquarian Audio; −209.7 dB re 1 V μPa nominal sensitivity), Fig. 1. The hydrophone frequency response (10 Hz-190 kHz) comfortably exceeds the "low-frequency" (<1 kHz) and "mid-frequency" (1–10 kHz) bands.
- Acquisition schedule: An identical XML script was flashed to every unit:
  - 48 kHz / 24-bit PCM (little-endian WAV).
  - Continuous duty-cycle, segmented into 10-min or 15-min files.
  - Fixed pre-amplifier gain.

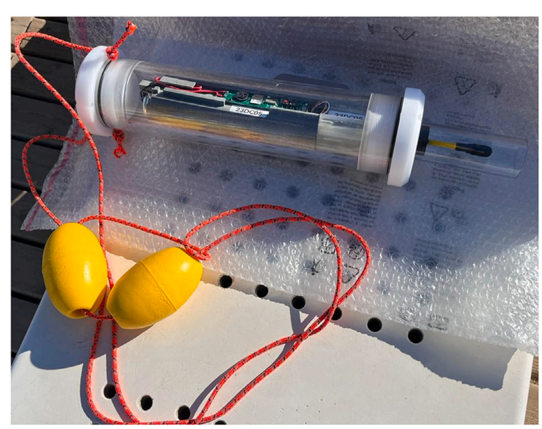


Fig. 1. Autonomous underwater acoustic recorder (uRec-384k, Nauta-RCS) coupled with an omni-directional hydrophone AS-1, Aquarian Audio; −209.7 dB re 1 V/μPa. Supported by two float buoys for vertical stability during deployment.

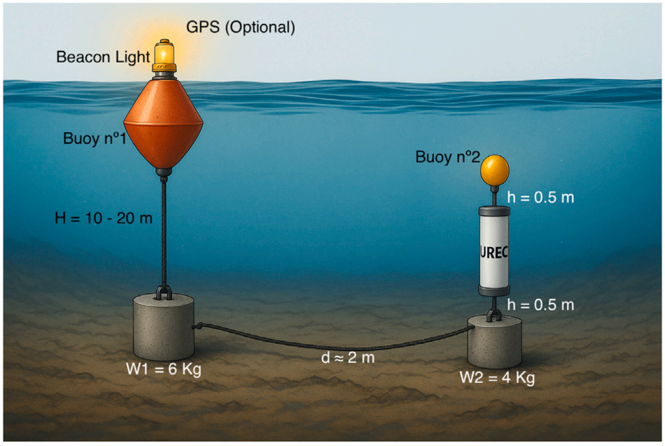


Fig. 2. Mooring system used for hydrophone deployment. The hydrophone is suspended 1.0–1.5 m above the secondary anchor, oriented capsule-downward to minimize flow noise. A surface recovery buoy with optional GPS and beacon light was included at sites where required by local port authorities. Mooring line lengths were adapted to local bathymetry.

- Power & storage: Each deployment used fresh alkaline “D” cells (+70 h autonomy) and a 128 GB SDXC card,

This fixed acquisition protocol, including sampling rate, gain, and segmentation scheme, was applied identically across all deployments. By maintaining a constant hardware configuration and factory-calibrated sensitivity, the recordings provide comparable sound pressure levels, suitable for cross-basin analyses and model validation.

3.3. Mooring configuration

The hydrophone was suspended on a vertical mooring: a bottom anchor weight, a smaller mid-line weight for tension and noise attenuation, a buoy to keep the line upright, and a surface recovery buoy, see Fig. 2.

- Anchor 1, around 6 kg (up to 10 kg on seabed; anchor 2, around 4 kg (up to 10 kg), 2 m away from the first one; in order to dampen surface-induced motion or noise.
- Hydrophone 0.5–1.5 m above anchor 2, capsule facing downwards to minimize flow noise.
- Two 25 cm float buoys maintain verticality; a surface marker buoy was added only where required by local port authorities.
- Total line length adapted to local bathymetry, all sites ∼20 m depth; see Table 1.

3.4. AIS unit

An Automatic Identification System (AIS) retrieved, in real time, the information from the vessels near the hydrophone during the deployment. A Very High Frequency (VHF) receiver was installed in LaSalle Campus Barcelona, Universitat Ramon Llull. It is driven by a Raspberry Pi and a dAISy HAT.[1] The minicomputer runs AISdispatcher, a lightweight utility that ingests AIS data following the NMEA (National Marine Electronics Association) standard validates them, removes duplicates, and compiles both global and per-message statistics. The system ran successfully for the last 3 years, with brief interruptions that did not affect the retrieval of the data, real-time status of this station can be checked.[2] The retrieved data is forwarded to AISHub, which in exchange granted free API access. A Python script then requests AISHub for all positional reports that fell inside a specific bounding box centred on the hydrophone position, once per minute — the shortest interval allowed under the free tier. On two occasions the free AISHub stream was interrupted — likely affected by the AIS restrictions imposed during the near Ukrainian conflict in Romania, the missing messages were filled with commercially sourced data. Aside from these short lapses, the AISHub feed, coupled with our antenna, proved an economical solution for continuous traffic monitoring. The Barcelona receiver is currently the only one available in the city; providing full coast coverage, detecting an average of 140 transiting vessels, daily maxima of 266 messages in transit and 100 unique vessels.

AIS messages were time-aligned with the acoustic files using coordinated timestamps, enabling a reproducible AIS-acoustic synchronisation procedure. This method allows vessel passages, distances, and speeds to be derived consistently across all campaigns, forming the basis of the event-detection workflow described in our previous work [25].

3.5. Field campaigns and study sites

The dataset was collected across three distinct European maritime environments, each one, hypothetically, representing different acoustic conditions and vessel traffic patterns. Recording campaigns were conducted in the Western Mediterranean (Catalan coast, Spain), the Black Sea (Romanian coast), and the Northern Adriatic Sea (Lagoon of Venice, Italy) between December 2023 and October 2024, Fig. 3.

3.5.1. Study sites in barcelona, Spain

The measurements along the Catalan coast include underwater acoustic recordings collected in Barcelona and Badalona, conducted by La Salle-Universitat Ramon Llull (La Salle-URL).

- **Port of Barcelona**. As one of the busiest ports in the western Mediterranean, this port hosts cargo operations, cruise ships, and private vessels, contributing to high vessel density and noise pollution. To capture the acoustic footprint of this dynamic maritime environment, hydrophones were deployed at Espai Vela (Universitat Politècnica de Catalunya, UPC), a site directly in front of the port's coast. This location was chosen to maximize the capture of vessel noise.
- **Pont del Petroli - Badalona**. Located approximately 20 km northeast of the Port of Barcelona, off the coast of Badalona. This site features a 250-meter-long pier extending into the sea, providing a stable platform for acoustic monitoring. Unlike the port area, Pont del Petroli is positioned away from direct port activities, offering a less polluted marine environment for comparative studies. The minimal anthropogenic noise in this area allows the recordings to capture natural acoustic patterns, enabling researchers to analyze underwater soundscapes with reduced anthropogenic influence and offering a more comprehensive perspective on the acoustic conditions of the Catalan coast.

3.5.2. Study sites in constanţa, Romania

Underwater acoustic measurements in Constanţa, Romania, were also conducted by La Salle-URL, focusing on maritime environments along the Romanian Black Sea coast.

- **Port of Constanţa**. Constanţa hosts three major commercial ports — Constanţa, Midia-Navodari, and Mangalia — which collectively handle over 80% of the naval traffic in the region. Additionally, three marina port facilities (Constanţa, Eforie Nord, and Mangalia) contribute seasonally to background noise levels through recreational boating and tourism activities. These areas are key contributors to underwater noise pollution, making them valuable for studying the impact of maritime traffic on the Black Sea's acoustic environment. Recordings were collected near one of the entrances to Constanţa Port, an area regularly transited by tugboats, cargo ships, and dredgers.

[1] https://www.aishub.net/ais-dispatcher (accessed on September 30, 2025).
[2] https://www.aishub.net/stations/3673 (accessed on September 30, 2025).

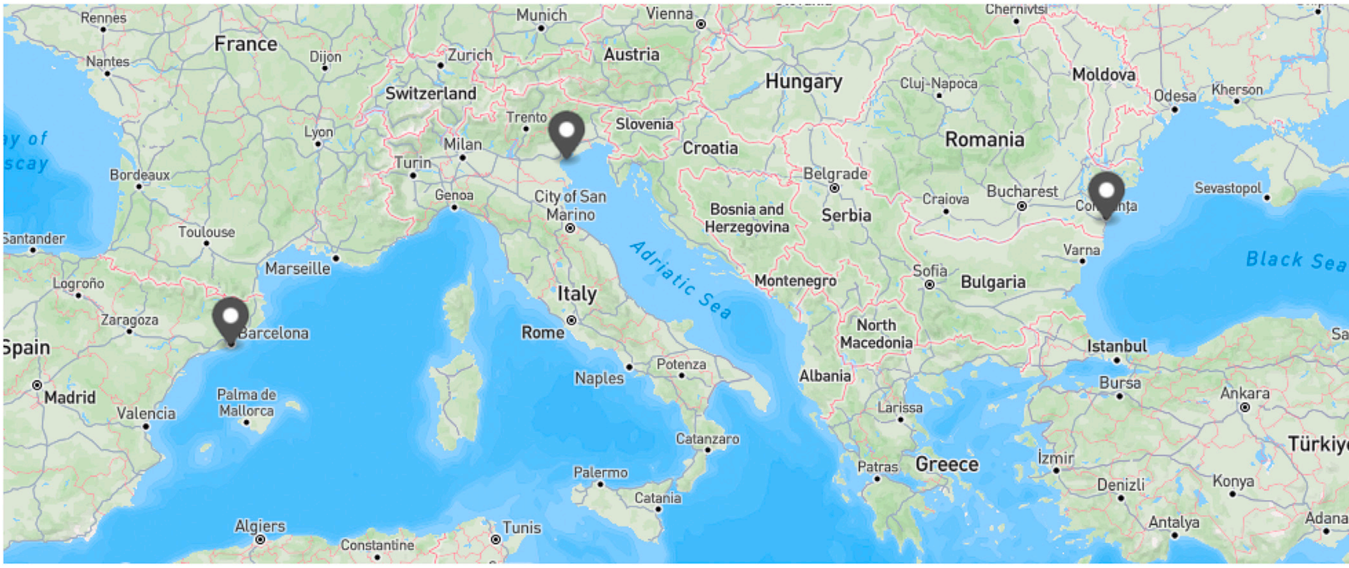


Fig. 3. Map showing the geographic locations of the three coastal sites studied: Barcelona (Spain), Venice Lagoon (Italy), and Constanţa (Romania).

Table 1
Summary of hydrophone deployment sites, including date, basin and country, station name and site type (Q = quiet, N = noisy), total raw hours recorded, deployment depth, and geographic coordinates.

Start date	Basin / Country	Station & site type	Raw hours	Depth (m)	Coordinates (lat N, lon E)
4-Dec-23	Catalan Coast Spain	Pont del Petroli, Badalona (Q)	**25.5**	**20**	41.44036 2.24865
12-Mar-24	Catalan Coast Spain	Espai Vela, Port of Barcelona (N)	**26.00**	**15**	41.36358 2.18729
12-Mar-24	Venice Lagoon Italy	AqualAlta (N)	**120,75**	**20**	45.40000 12.45000
12-Mar-24	Venice Lagoon Italy	Chioggia (Q)	**120,75**	**10**	45.22648 12.25943
14-Mar-24	Catalan Coast Spain	Pont del Petroli, Badalona (Q)	**24.00**	**20**	41.44036 2.24865
17-Jul-24	Catalan Coast Spain	Espai Vela, Port of Barcelona (N)	**46.80**	**15**	41.36358 2.18729
22-Jul-24	Catalan Coast Spain	Pont del Petroli, Badalona (Q)	**23.20**	**20**	41.44036 2.24865
3-Sep-24	Black Sea Romania	Mamaia pier (Q)	**29.20**	**20**	44.23165 28.63423
6-Sep-24	Black Sea Romania	Constanţa Port (N)	**29.70**	**20**	44.14648 28.66854
11-Oct-24	Catalan Coast Spain	Espai Vela, Port of Barcelona (N)	**50.75**	**15**	41.36358 2.18729
14-Oct-24	Catalan Coast Spain	Espai Vela, Port of Barcelona (N)	**93.25**	**15**	41.36358 2.18729
18-Oct-24	Catalan Coast Spain	Espai Vela, Port of Barcelona (N)	**100.25**	**15**	41.36358 2.18729

- **Mamaia Area**. Located north of Constanţa, this area offers a relatively quiet marine setting in contrast to the high vessel activity near Constanţa Port. During the winter months, the area experiences minimal anthropogenic noise, making it an ideal location for studying natural underwater soundscapes. However, in the summer season, the presence of motorboats and recreational vessels near the beach increases noise levels, introducing seasonal variations in the acoustic environment. The contribution of this site to the dataset provides another perspective on the acoustic conditions along the Constanţa coast.

3.5.3. *Study sites in lagoon of venice, Italy*

Underwater acoustic recordings in Italy were collected by Università degli Studi di Milano - Bicocca, focusing on the Lagoon of Venice.

- **The Lagoon of Venice**, the largest lagoon in the Mediterranean, spans approximately 550 km^2, with 85% covered by water, 6% by islands, and 7% by salt marshes. This area presents seasonal variations in underwater noise, with increased levels during summer due to heightened tourist activity and recreational boating. Additionally, the lagoon contains a port for cargo and cruise ships and a petrochemical industry, both of which contribute to the anthropogenic noise footprint.

4. The DeuteroNoise dataset

Over the course of 12 short-term deployments carried out between December 2023 and October 2024, we measured three contrasting European basins. In every basin, one recorder was moored inside a high-traffic site (noise-dominant, “-N”) while another unit was placed in a comparatively coastal site (quiet, “-Q”). By repeating this paired design across winter and summer in some basins and exceptional-traffic periods (e.g., the 37th America’s Cup logistics in Barcelona), the recordings yielded 689.70 h of calibrated 48-kHz audio that captures the full gradient from dense traffic sites to natural soundscapes. Table 1 summarises the deployment windows and the quantity of data recorded after quality control for each campaign; these recordings constitute the acoustic data of the DeuteroNoise Dataset. All 689.70 h (openly available) correspond to continuous raw underwater acoustic recordings collected at 48 kHz, segmented only into 10 or 15 min WAV files by the recorder. On the other hand, ∼12 h of labelled data is also included.

4.1. *Dataset structure*

At present the raw data from the dataset exceeds 300 GB of 48 kHz/24-bit audio. Because of the volume, all WAV files are held on dedicated RAID (Redundant Array of Independent Disks) storage at La Salle-URL; the PostgreSQL back-end keeps only file paths and summary

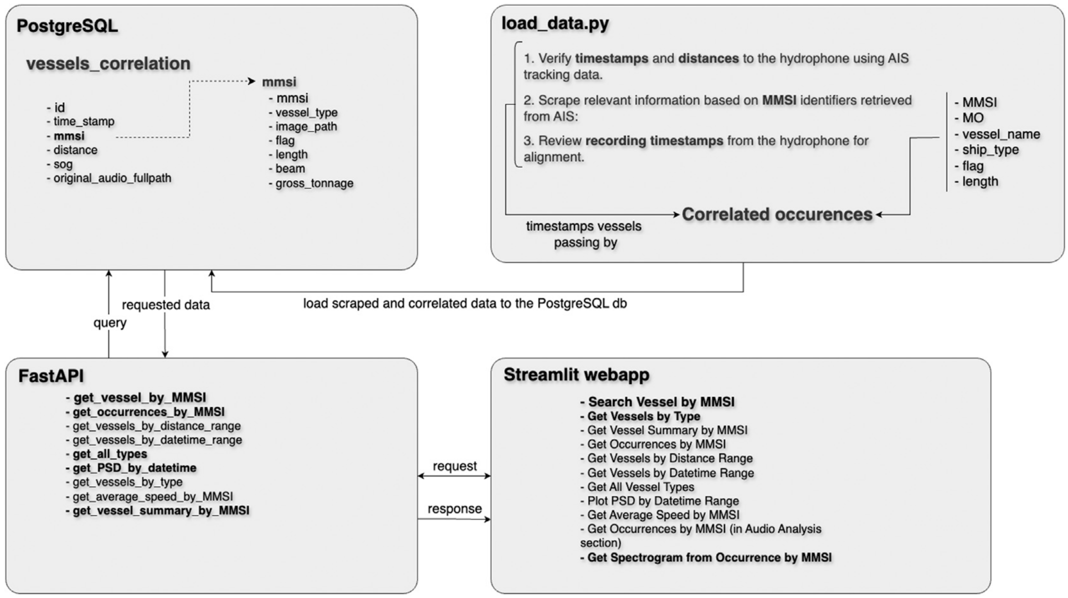


Fig. 4. Overview of the data pipeline and web architecture for the DeuteroNoise Dataset, combining PostgreSQL, FastAPI, and a Streamlit dashboard.

statistics. Each recording is time-matched to the 60 s AIS feed described above, this combines:

- AIS fields - Maritime Mobile Service Identity (MMSI), timestamp, latitude, longitude, speed-over-ground
- Scraper-derived from VesselFinder[3] - vessel class, length, beam, gross tonnage, flag and vessel image.

This linkage lets users query the dataset not just by acoustic features but also by vessel identity, dimensions and operating speed, enabling a source-level analyses and cross-basin comparisons. To accurately associate vessel passages with their acoustic signatures, as mentioned, vessel movements were correlated with the AIS dataset. Automated algorithms developed in Nou-Plana et al. [25] were then employed to:

- Detect vessel passages, determining the start and end of each event near the hydrophone.
- Identify peak noise levels during the passage. Usually at the closest point to the hydrophone (this was double checked by using the timestamps from the AIS dataset and the recordings).

Following this, a manual listening procedure was conducted to validate and adjust the data, ensuring accurate alignment between vessels and the recordings. Once vessel passages were detected, labelled, and cleaned, an initial spectral analysis was performed by computing spectrograms and frequency spectra at peak times.

4.2. *Storage management*

The dataset is built using PostgreSQL, structured to store and manage detailed acoustic data as well as metadata from the vessels. It comprises two primary relational tables: MMSI and vessels correlation, detailed information can be found in Nou-Plana et al. [26]. The MMSI table contains static vessel data, including vessel type, dimensions, tonnage, flag, and vessel images, whereas the vessels correlation table contains dynamic data such as timestamps, distance to hydrophones, Speed Over Ground (SOG), and references to the audio excerpts. Details regarding this structure and the pipeline followed are illustrated in Fig. 4.

[3] https://www.vesselfinder.com/ (accessed on September 30, 2025).

4.3. *API development and functionalities*

A RESTful API developed using FastAPI also provides access to the dataset. It enables users to query and retrieve data for their specific research objectives, such as:

- Retrieval of detailed vessel metadata using MMSI identifiers.
- Extraction of vessel occurrence data filtered by parameters such as MMSI, vessel type, distance ranges, and datetime intervals.
- Computation of statistical summaries, including average vessel speed and average distance to the hydrophone.
- Access to already processed acoustic metrics such as Power Spectral Density (PSD) values.

4.4. *Interactive web interface and visualization*

An intuitive and interactive web interface, built using Streamlit [27], complements the API by allowing users — from researchers to policymakers — to visually explore and analyze complex maritime acoustic data. Features include:

- Vessel Dashboard: Quick retrieval and visualization of vessel details, including vessel type, dimensions, tonnage, flags, and images, based on MMSI identifiers.
- Dynamic Data Exploration: Filtering of vessel occurrence events by customizable parameters such as distance to hydrophones, vessel types, and specific time ranges.
- Spectrogram Visualization: Interactive generation and analysis of spectrogram plots from selected audio excerpts, facilitating detailed examination of vessel acoustic signatures.
- Statistical Summaries: Access to summarized data, such as vessel occurrence counts, average vessel speeds, and average distances to hydrophones, supporting comparative analyses.

Furthermore, users can navigate and explore recorded campaigns across various basins (e.g., Spain, Italy, Romania), filtering data by geographical area, specific campaigns, and vessel categories.

5. Dataset overview and characteristics

The following section will focus on the composition of the DeuteroNoise Dataset, quantifying what we have rather than how

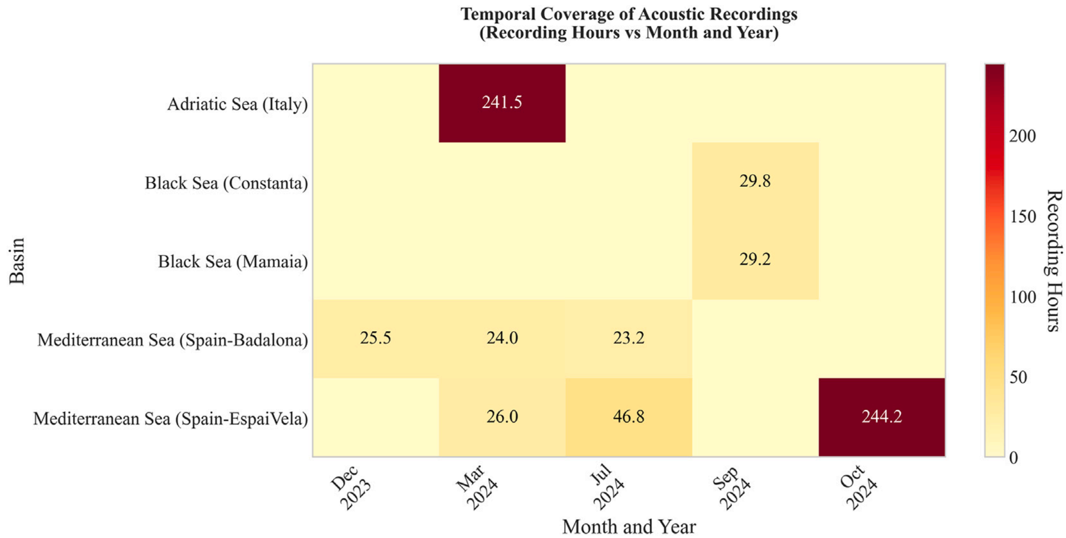


Fig. 5. Temporal coverage of acoustic recordings across the three monitored basins - the Catalan Coast (Western Mediterranean, Spain), the Venice Lagoon (Northern Adriatic Sea, Italy), and the Black Sea (Romanian coast) - between December 2023 and October 2024.

we processed it. We begin with campaign-level total acoustic volume (hours of raw WAV), the amount of annotated data, and the correlated AIS. MMSI unique data has also been studied giving an overview of ship diversity, followed by other statistics: median SPL, occurrence-length percentiles and distance. Across this section we report several quantitative observations that characterise the dataset and reveal cross-basin differences in traffic patterns, vessel diversity, and acoustic profiles.

5.1. Overall volume and temporal span:

Between December 2023 and October 2024 the DeuteroNoise campaigns retrieved 689.70 h of underwater recordings. Of this total, 389.70 h were gathered along the Catalan coast of Spain (Port of Barcelona and Badalona), 241.50 h came from the Venice Lagoon sites in Italy, and the remaining 59 h were collected in the western Black Sea in Romania (Port of Constanţa and Mamaia). Breaking down by individual campaigns, the raw data is distributed as follows:

- Lagoon of Venice (Italy) - **241.5 h**
- Port of Barcelona - Espai Vela (Spain) - **317 h**
- Badalona (Spain) - **72.7 h**
- Port of Constanţa (Romania) - **29.8 h**
- Mamaia (Romania) - **29.2 h**

The temporal distribution covers a seasonal representation during the Catalan coast campaigns. We sampled four distinct periods – early winter (4–5 Dec 2023), the early-spring (14–15 Mar 2024 plus 12–13 Mar 2024 at the Espai Vela site), mid-summer (17–19 Jul 2024 and 22–23 Jul 2024), and mid-autumn (11–22 Oct 2024). This sequence captures the full annual cycle of tourist pressure and commercial traffic. In the northern Adriatic (Italy) recorders operated from 12–19 March 2024. Finally, the western Black Sea (Romania) was monitored at Mamaia and Constanţa between 3–7 September 2024. This seasonal coverage provides the opportunity to examine temporal variability in soundscapes driven by tourist pressure and commercial traffic, particularly along the Badalona and Mamaia coasts.

In total, the analyzed dataset comprised 548 vessel passages involving 221 unique vessels over the 689.70 recording hours, equating to an average of 0.32 unique vessels per hour and 0.80 passages per hour. Taking into account that some campaign sites were considered quiet reference locations where few or no vessels were recorded, the averages drop significantly compared to the high-traffic sites. This contrast between quiet and noisy sites reflects the paired-deployment design and provides the basis for comparative acoustic analyses. On average, each vessel was recorded 2.5 times during the campaigns. The temporal and spatial distribution of these recordings across the different sites is illustrated in Fig. 5, which highlights periods of higher and lower monitoring throughout the study.

5.2. Annotated data

Every event in the DeuteroNoise Dataset is time-stamped and paired with a synchronous slice of the AIS data. This overlay turns the 250 GB of raw hydrophone output into a localized traffic recordings that can be queried for ship counts, passage density, speed profiles, and other vessel-focused statistics.

5.2.1. Vessel-type distribution and unique-ship coverage

Across the eleven recording campaigns, a total of 221 unique MMSI identifiers were detected. The two long-duration deployments in the Venice Lagoon (Adriatic Sea) contributed the largest share, with 100 unique vessels (364 transits). This basin displayed the greatest vessel type diversity (14 categories), but was heavily skewed toward fishing vessels, which accounted for 61% of all unique MMSI. Sailing vessels (13.0%) and tugs (7.0%) were the next most common, with the top three categories alone representing 81.0% of the fleet. Other types present included general cargo ships, tankers, passenger vessels, and a range of specialized workboats, reflecting the mixed-use nature of the lagoon and adjacent coastal waters. These observations indicate that the Adriatic soundscape tends to be shaped primarily by fisheries and mixed commercial activities.

The Western Mediterranean sites in Barcelona (Espai Vela) recorded 65 unique vessels (123 transits), with 8 different vessel categories. This dataset was strongly dominated by recreational traffic: sailing vessels (46.2%) and pleasure craft (32.3%) together comprised more than three-quarters of all unique MMSI, with the addition of a small number of yachts, pilot launches, and high-speed craft (HSC). The high dominance of leisure vessels in the Catalan sites contrasts with the more industrial and mixed-use profile of the Adriatic dataset. This pattern reveals a coastal soundscape predominantly influenced by small recreational craft operating near shore.

The Black Sea dataset from Constanţa presented a different picture, with 38 unique vessels recorded over just 29.8 h (38 transits). Although vessel type diversity was again high (13 categories), the distribution

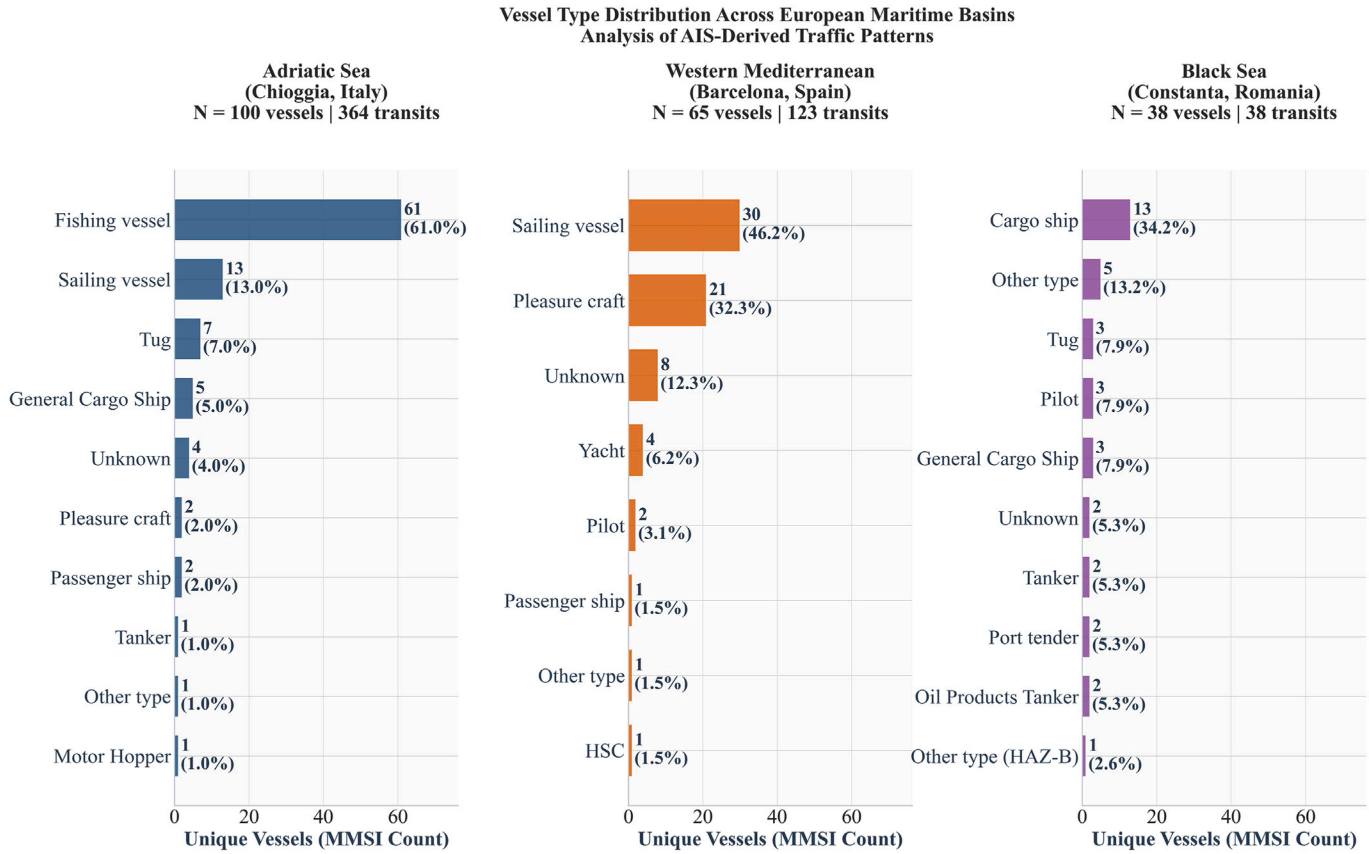


Fig. 6. Vessel type distribution across the three main basins monitored during the DeuteroNoise campaigns: Adriatic Sea (Chioggia, Italy), Western Mediterranean (Barcelona, Spain), and Black Sea (Constanţa, Romania). Bars represent the number of unique vessels (MMSI count) per type.

was less dominated by a single category compared to the Adriatic or Catalan datasets. Cargo ships represented 34.2% of the fleet, followed by "other type" (13.2%), "general" cargo ships (7.9%), pilots (7.9%), and tugs (7.9%). This more balanced and commercially oriented traffic composition reflects the functional role of Constanţa as a major port complex.

Fig. 7 demonstrates that fishing vessels, sailing vessels, and pleasure craft together account for over 80% of all unique MMSI identifiers, leaving a long tail of less common categories. This dominance is visually reinforced in the basin-specific distributions (Fig. 6), where the Adriatic fleet is heavily fishing-focused, the Catalan fleet is primarily leisure-oriented, and the Black Sea fleet is commercially diverse but led by cargo vessel traffic. Taken together, these results show pronounced cross-basin contrasts in fleet composition, confirming that each monitored region exhibits a distinct acoustic regime.

5.2.2. Labelled events

The vessel passages in the DeuteroNoise Dataset were labelled primarily by combining the audio recordings with their corresponding AIS metadata. The process is based on a semi active listening workflow. First, the audio files were correlated with the retrieved AIS stream: timestamps were aligned, and vessel positions were interpolated relative to the hydrophone to determine the closest point of approach and its associated time. During the AIS correlation stage, any detected time windows that contained overlapping passages of more than one vessel were discarded. This reduced the total amount of annotated material but ensured that the resulting excerpts corresponded to single, clearly identifiable sources. After this correlation step, a wav excerpt of approximately one minute was extracted around the moment when the vessel reached the minimum distance. These excerpts were then passed through the trained model described in Ref. [9], which produced a confidence value indicating whether the excerpt contained vessel noise or a background noise. Excerpts identified as containing vessel sound were further processed using a Python script designed to provide an initial estimate of the event boundaries. This script produced approximate start and end points based on background conditions, local signal to noise ratio, and a predefined threshold. These automatically generated boundaries were then used as reference points for a final manual validation step. Each excerpt was actively listened to, and the event limits were refined by examining the spectrogram to ensure that the annotated start and end times matched the onset and offset of the acoustic signature. During the active listening stage, excerpts in which more than one vessel could be heard simultaneously were also excluded. Only events with a unique and unambiguous acoustic signature were retained, further improving the clarity and reliability of the labelled dataset. This combination of automatic estimation and manual verification results in a high-confidence set of single-source events suitable for acoustic characterisation. The resulting annotations were compiled into CSV files that include the MMSI of the vessel, the vessel type, the distance to the hydrophone at closest approach, and the corrected start and end timestamps referenced to the corresponding audio file. All annotated excerpts in this release correspond to single-vessel passages, and any overlapping, ambiguous, or multi-source events were systematically removed during the AIS correlation and manual validation steps; therefore, users may still encounter unlabelled portions of the raw recordings that contain overlapping vessel sounds or mixed-source acoustic scenes. Across all three monitored basins, the labelled dataset comprises 11.73 h of verified vessel-related audio. To provide a clearer overview, for this analysis, vessel types were grouped into six operational categories: recreational and leisure, fisheries, commercial cargo and transport, service and workboats, passenger vessels,

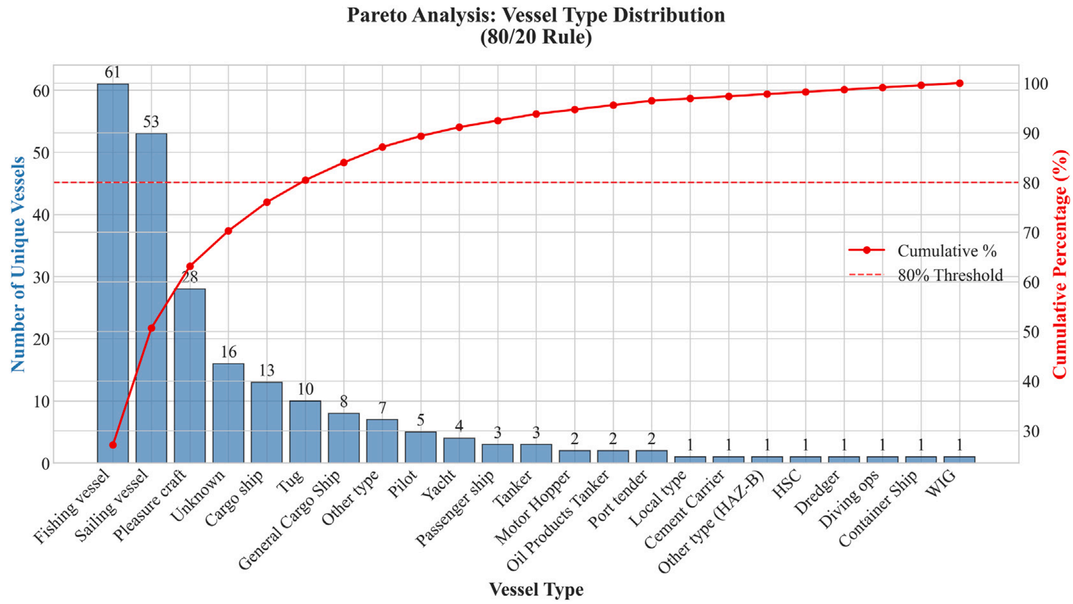


Fig. 7. Pareto analysis of vessel type distribution based on unique MMSI counts from all monitored basins. Blue bars represent the number of unique vessels per type, while the red line shows the cumulative percentage. The dashed horizontal line marks the 80% threshold, illustrating that fishing vessels, sailing vessels, and pleasure craft together account for over 80% of the recorded fleet.

and unknown. This grouping helps summarise activity patterns across basins without relying on fine-grained AIS classes, i.e., 22 fine-grained different types. In the Western Mediterranean (Espai Vela, Barcelona), the annotated material totals 4.41 h and is dominated by recreational and leisure vessels, reflecting the high density of small craft operating close to the urban coastline. Only a minor proportion originates from service and workboats or passenger activity, and a small share corresponds to vessels without reliable classification. Since the entrance of that port is narrow, the vessel passages that appear in the recordings are brief, with event durations generally between 10 s and 1 min. In the Western Black Sea (Constanţa), the annotated subset amounts to 1.54 h. Here, commercial cargo and transport vessels represent most of the labelled time, consistent with the port-oriented nature of the site. Service and workboats also contribute significantly, while other categories remain marginal. Because the port entrance in Constanţa is wider, the recorded events tend to last longer, with durations ranging from 39 to 270 s and a mean of approximately 146 s. In the Northern Adriatic (Chioggia, Venice Lagoon), 5.78 h of annotated data were obtained. Fisheries dominate the annotations, matching the well-known role of Chioggia as an active fishing area. Additional contributions come from service vessels and a smaller share of leisure traffic. Chioggia events usually occur at larger distances from the hydrophone, and the individual vessel recordings are relatively short, with durations ranging from 27.77 to 1 min and a mean of 57.12 s. Aggregated across basins, the annotated dataset highlights three distinct regional profiles: leisure-driven activity in Barcelona, commercial-port traffic in Constanţa, and fishing-oriented operations in Chioggia; illustrating how local maritime functions shape the structure of underwater soundscapes across regions.

5.3. Distance range

The spatial distribution of vessel detections relative to hydrophone positions varied significantly across the three monitoring basins, reflecting distinct maritime traffic patterns, bathymetric conditions, and deployment characteristics of each site, Fig. 8. Distance measurements ranged from 12 to 1.152 m across all monitoring stations, with each basin exhibiting characteristic detection patterns that provide insights into local vessel behavior and acoustic monitoring effectiveness.

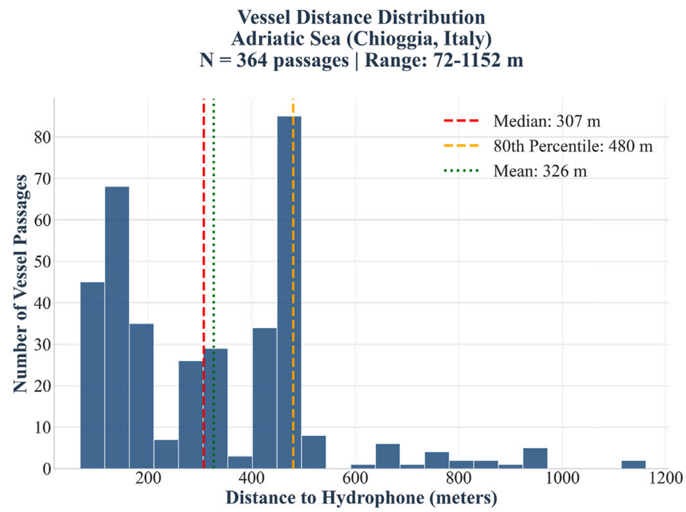
The Adriatic Sea stations in the Venice Lagoon recorded vessel passages across the broadest distance range (72–1.152 m), with a median detection distance of 307 m, Fig. 8(a). This wide distribution, is characterised by peaks at 150–200 m and 400–500 m. In contrast, the Western Mediterranean stations near Barcelona contain near-field detections, with 89% of the 123 recorded passages occurring within 50 m of the hydrophone (range: 12–148 m, median: 31 m), Fig. 8(b). This reflects the recreational nature of the monitored traffic, where sailing vessels and pleasure craft navigate in close proximity to the coastal deployment sites at Espai Vela. Finally, the Black Sea monitoring at Constanţa Port demonstrated intermediate detection characteristics, with vessel passages distributed across a 57–195 m range (median: 98 m), Fig. 8(c). Despite the smaller sample size (N = 38), this distribution pattern reflects the structured approach routes typical of a commercial port entrance, where cargo vessels and other commercial traffic maintain relatively consistent distances while following established navigation channels.

These distance distribution patterns have implications for the acoustic analysis and interpretation of vessel noise impacts. The extended ranges achieved in the Adriatic Sea provide comprehensive spatial coverage suitable for assessing cumulative noise exposure across different vessel types and operational scenarios. The near-field detections in the Western Mediterranean offer exceptional detail for individual vessel characterisation but may underrepresent more distant traffic. The intermediate ranges at Constanţa represent an optimal balance between detection coverage and signal quality, particularly valuable for analyzing the acoustic signatures of commercial vessels in port approach scenarios.

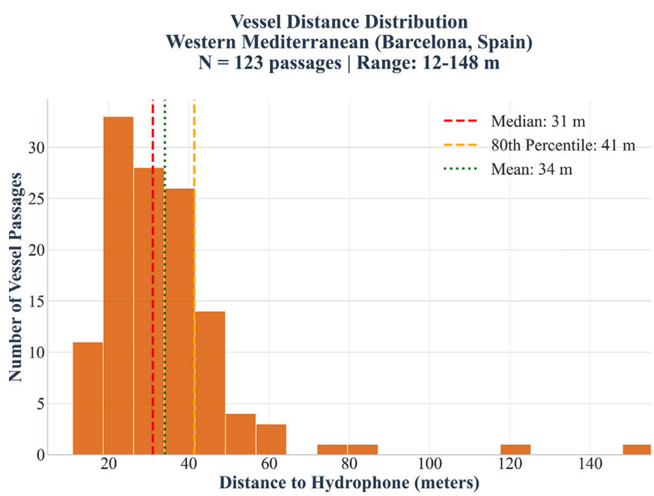
These relationships between spatial detection patterns and fleet characteristics provide context for interpreting the underwater noise measurements and their potential environmental impacts across different European maritime environments. As a result, distance distributions serve as another essential component for understanding acoustic profile variability within the DeuteroNoise Dataset.

5.4. Constraints of AIS correlation and site-specific propagation conditions

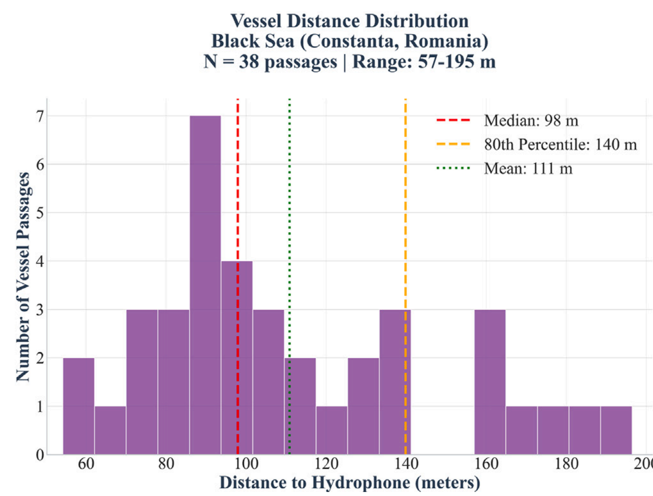
Although AIS linked annotation provides a practical way to associate acoustic events with specific vessels, it introduces several sources



(a) Adriatic Sea (Italy)



(b) Western Mediterranean (Spain)



(c) Black Sea (Romania)

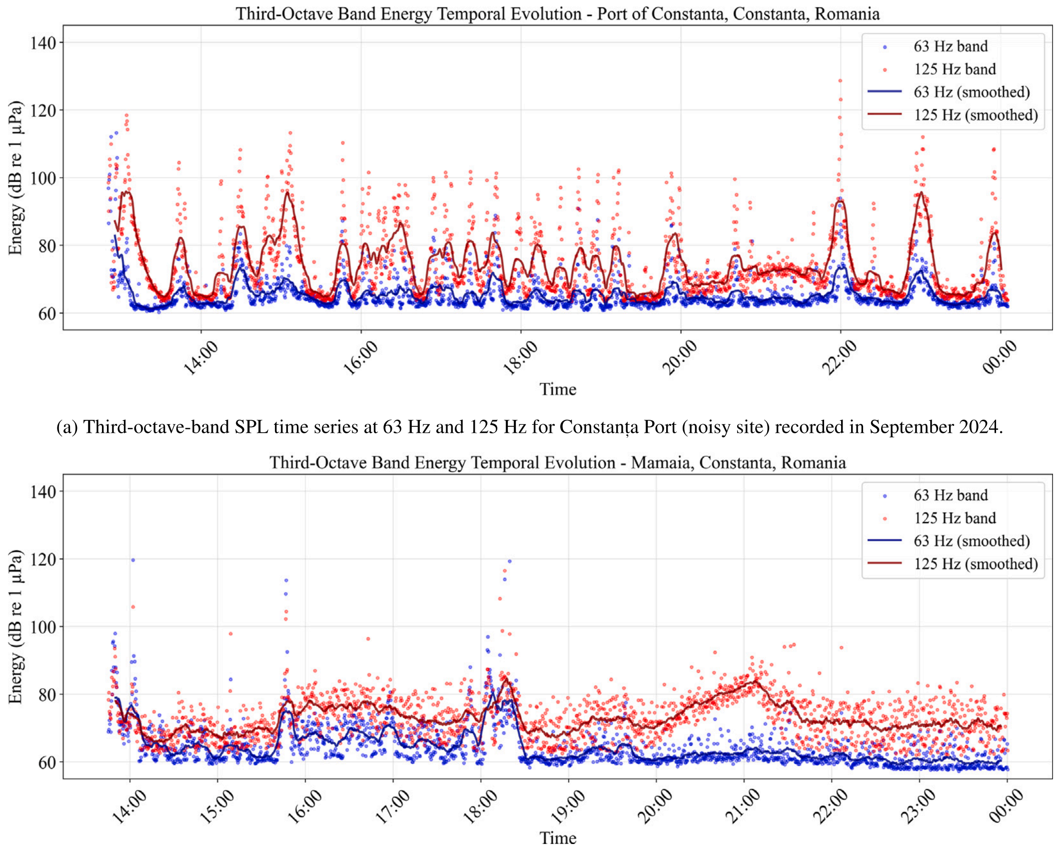
Fig. 8. Distance range in three European maritime basins during the DeuteroNoise monitoring campaign. Distance histograms reveal distinct vessel traffic patterns: (a) Adriatic Sea shows wide-range commercial traffic with median detection at 307 m, (b) Western Mediterranean exhibits predominantly near-shore recreational activity with 31 m median distance, and (c) Black Sea displays moderate port-related traffic with 98 m median range. These patterns reflect regional differences in vessel types, navigation channels, and coastal bathymetry.

of uncertainty that must be considered. AIS data quality can vary considerably, since positional information depends on the transmitting equipment, the update rate of the vessel, and the reception and aggregation performed by external services. Down sampling, and inconsistent reporting intervals can introduce temporal inaccuracies in the vessel track, and user entered static metadata such as vessel dimensions or type are often incomplete or erroneous. In addition, small recreational craft, fishing boats, and other local traffic may not transmit AIS at all, may do so intermittently, or may deactivate the system. As a result, AIS cannot be considered an exhaustive or fully reliable representation of vessel activity.

To minimize these limitations, all AIS tracks associated with acoustic events were manually inspected. Any cases where the AIS trajectory or timing did ambiguously correspond to a single vessel passing sound recorded near the hydrophone, were discarded. Similarly, if the acoustic record indicated the presence of more than one vessel near the AIS timestamp, the event was removed from the labelled set. This filtering process reduced the final number of annotated events but ensured that the retained examples corresponded to unique and clearly identifiable passages, as mentioned in Section 5.2.2. This conservative approach prioritises annotation reliability over dataset volume, resulting in a high-confidence subset of vessel events.

Another factor influencing the precision of source related interpretation is the local bathymetry. Water depth, bottom type, and coastal morphology affect sound propagation and can introduce uncertainties when comparing signal levels across basins. Bathymetric data were not integrated into the present release of the dataset, and no propagation model was applied. While this does not affect the integrity of the labelled events, it does limit the ability to derive absolute source levels or to directly compare propagation conditions between sites. Future extensions of the dataset will incorporate detailed bathymetry to support more accurate source level estimation and cross basin analyses.

Meteorological data such as wind speed, wave height, and precipitation are not included in the present release. The focus of this initial dataset is on providing continuous calibrated underwater recordings paired with AIS metadata, while the integration of environmental parameters requires access to region-specific meteorological archives from different national agencies, which were not consistently available during this study. These environmental variables can, however, be retrieved independently by users if needed, and future extensions of the dataset



(a) Third-octave-band SPL time series at 63 Hz and 125 Hz for Constanța Port (noisy site) recorded in September 2024.

(b) Third-octave-band SPL time series at 63 Hz and 125 Hz for Mamaia (quiet site) recorded in September 2024.

Fig. 9. Example of low-frequency SPL data from two contrasting sites in the western Black Sea basin.

will incorporate harmonised environmental and bathymetric metadata to support more detailed soundscape analyses.

Taken together, these limitations reflect the challenges inherent to AIS based acoustic correlation, particularly in coastal environments with mixed traffic and heterogeneous reporting practices. The combination of automatic detection, manual inspection, and the exclusion of ambiguous events helps mitigate these issues, but users of the dataset should remain aware of the constraints associated with the site-specific propagation conditions.

6. Comparative analysis of SPL time series at contrasting sites

The DeuteroNoise Dataset offers calibrated underwater acoustic recordings from multiple European basins, enabling a consistent basis for quantifying and comparing underwater SPL across a wide range of environments. The SPL comparisons presented below highlight how vessel activity, site characteristics, and local navigation patterns shape the low-frequency acoustic environment across regions. As previously mentioned in section, each basin was monitored with paired deployments: a high-traffic “noisy” site positioned in close proximity to major shipping or port operations, and a comparatively “quiet” site located outside the main traffic corridors but within the same hydrogeographic context. This experimental layout allows researchers to capture both ends of the anthropogenic noise gradient, and to examine how spatial differences interact with diurnal cycles, seasonal patterns, and specific vessel activities.

Such data can be used to produce broadband or frequency-specific SPL metrics, diurnal versus nocturnal noise profiles, or traffic-correlated sound signatures, which are valuable for environmental assessment and regulatory compliance under frameworks such as the EU MSFD (Descriptor 11) [9]. The low-frequency bands around 63 and 125 Hz, used here as an example, are particularly relevant as they are strongly influenced by shipping activity and overlap with the hearing ranges of many marine animals, including fish and marine mammals.

Fig. 9 provides a representative case study from the western Black Sea basin, based on two short-term deployments in September 2024. At Constanța Port, the SPL time series shows a persistently elevated baseline in both bands, punctuated by frequent peaks exceeding 100 dB re 1 μPa. These elevated levels confirm the strong contribution of port-related traffic to the local low-frequency acoustic environment and these events coincide with vessel passages logged in the synchronised AIS record, associated with high-tonnage cargo vessels, tugs, and dredging operations concentrated at the port entrance. The temporal distribution of these peaks follows a diurnal pattern, with the highest intensity and frequency of events during working daylight hours, and reduced, but still elevated, activity at night. This recurring diurnal pattern reflects the operational schedule of port manoeuvres and commercial traffic flows.

In contrast, the Mamaia site, located approximately 10 km north of Constanța Port, presents a different acoustic profile. Median SPL levels derived directly from the dB time series are 10–15 dB lower in both bands, with minimal fluctuations during night hours and only occasional daytime increases. These sporadic peaks are likely attributable to small recreational vessels, but remain well below the intensity of port-related traffic noise.

This comparison illustrates how the DeuteroNoise Dataset can be used to extract detailed noise profiles and to isolate the acoustic signatures of distinct vessel activities. Beyond shipping studies, such analyses can inform spatial planning, identify areas where noise mitigation may be most beneficial, and provide baseline conditions for assessing compliance with underwater noise thresholds. These findings underscore the dataset’s ability to support regulatory evaluation and ecological impact assessments across multiple European environments.

Table 2
Comparison of major publicly available underwater acoustic datasets. DeuteroNoise is the only one with full open access (audio, AIS tracks, range data, and metadata) under permissive CC-BY-4.0 / MIT licensing.

Feature	ShipsEar	DeepShip	VTUAD	DeuteroNoise
Geographic coverage	Single ria (Ría de Vigo, NW Spain, Atlantic)	Single cabled site Strait of Georgia (NE Pacific)	Strait of Georgia (NE Pacific), multi-scenario distance benchmark	**Three distinct basins: Western Mediterranean (ES), Northern Adriatic (IT), Western Black Sea (RO)**
Number of campaigns / years	2 campaigns (2012-2013)	Continuous 2016–2018	Continuous ONC dataset (no explicit cruise info)	**6 multi-day campaigns (Dec 2023 - Oct 2024)**
Deployment environment	Port quays & near-shore reference	Fixed ONC cabled node (141-147 m depth)	ONC hydrophones with three distance scenarios	**Shallow-water moorings (15-25 m)**
Recording chain	MarSensing SR-1, 52.734 kHz, 24-bit	icListen AF smart hydrophone, 32 kHz	ONC hydrophones (sampling rate not specified in papers)	**URec-384k + Aquarian AS-1, 48 kHz/24-bit, identical calibration at all sites**
Accessible audio volume	~2 h (90 wav files)	47 h (613 wav files)	~49 h total (1-s clips, ~59k in curated set)	**> 100 h (continuous 10-/15-min blocks)**
Vessel classes	11 fine-grained types (merged into 4 for experiments)	4 (Cargo, Tanker, Tug, Passenger)	5 (Cargo, Tanker, Tug, Passenger, Background)	**11 + raw AIS labels, collapsible into coarse or fine taxonomies**
Range annotation	Approx. distance bins (< 50, 50–100, 100–150, > 150 m)	AIS-based single-vessel events (≤ 2 km, 3 distance points per pass)	Three distance scenarios: S1: 2 km incl / 4 km excl; S2: 3 km / 5 km; S3: 4 km / 6 km	**Continuous AIS tracking; exact range to hydrophone stored for every passage**
Environmental metadata	Wind speed (manual)	None beyond water depth	Conductivity, Temperature, Depth, Salinity, Sound Velocity	**None beyond water depth**
Natural soundscape baselines	Wind, rain, wave clips	Absent	Ambient noise when no vessel present	**Dedicated quiet-site recordings**
Primary research use so far	GMM vessel classifier	Deep learning benchmark	Range-aware vessel classification	**Cross-basin analysis, SPL analysis, invertebrate-exposure assays, ML automatic classification [25]**
Licence / access	CC-BY-NC via project website	CC-BY-4.0 via Zenodo	CC-BY-4.0 via IEEE DataPort	**CC-BY-4.0 / MIT; full API & dashboard (all data open)**

7. Conclusions

The DeuteroNoise Dataset represents a significant advancement in underwater acoustic monitoring, addressing critical limitations in existing datasets through its comprehensive, multi-basin approach to documenting vessel noise and natural soundscapes across European coastal waters. It provides broader geographic coverage, standardized methodology, and a substantially larger volume of continuous raw audio (approximately 700 h) compared to existing open datasets that usually provide only 2–50 h of vessel-only recordings. In addition, the DeuteroNoise Dataset includes 12 h of labelled single-vessel passages. The analyses presented in this work demonstrate clear cross-basin differences in vessel-type composition, event durations, and distance-to-hydrophone distributions, helping to highlight the diversity of maritime activity across regions. This open-access resource bridges gaps in underwater acoustic research by providing calibrated, standardized recordings from three hydrogeographically distinct European basins — the Western Mediterranean (Catalan coast), Northern Adriatic Sea (Venice Lagoon), and western Black Sea (Romanian coast). Our results show how these contrasting environments generate distinct acoustic profiles shaped by local navigation patterns, port operations, and recreational activity.

The dataset compiled 689.70 h of calibrated 48 kHz hydrophone recordings, documenting 548 vessel passages from 221 unique vessels. This substantial dataset encompasses diverse vessel types, from recreational sailing vessels and pleasure craft to commercial cargo ships, tankers, and specialized workboats, providing comprehensive coverage of European maritime traffic patterns. These differences translate into measurable contrasts in low-frequency SPL levels, near-field versus far-field detection patterns, and the relative influence of anthropogenic noise across basins. The paired deployment strategy, combining high-traffic “noisy” sites with comparatively “quiet” reference locations, captures the full gradient of anthropogenic noise impacts across different coastal environments; enabling a robust comparison of background conditions and vessel-driven acoustic variability within and between basins.

The standardized methodology employed across all monitoring sites, utilizing identical uRec-384k recorders coupled with calibrated AS-1 hydrophones, ensures data comparability and enables robust cross-basin analyses. This uniformity addresses a fundamental limitation of previous datasets, where differing hardware and calibration procedures prevented direct comparison of absolute sound pressure levels between sites. The reproducibility of the acquisition protocol strengthens the dataset’s value as a benchmark resource for future underwater acoustic studies.

The integration of real-time AIS data with acoustic recordings provides vessel identification and tracking capabilities, enabling researchers to correlate specific acoustic signatures with vessel identity, position, speed, and physical characteristics. The PostgreSQL dataset architecture, coupled with a FastAPI backend and an interactive Streamlit web interface, delivers a scalable platform for large-scale data retrieval and exploration that significantly enhances accessibility for researchers worldwide. The dataset’s open-access design, under CC-BY-4.0, promotes collaborative research and supports the development of advanced analytical methods, including machine learning models for vessel classification and acoustic propagation studies.[4]

The DeuteroNoise Dataset provides essential baseline data for environmental impact assessments and regulatory compliance under frameworks such as the EU Marine Strategy Framework Directive Descriptor 11. The documented sound pressure level profiles across different European maritime environments offer critical insights into spatial and temporal variations in underwater noise pollution, supporting evidence-based policy decisions for marine conservation. The dataset’s capacity to isolate acoustic signatures of specific vessel activities enables detailed analyses of noise sources, facilitating the development of targeted mitigation strategies and the observed differences in vessel

[4] Details on dataset access and usage are available at https://github.com/Ignasinou/deuteronoise-dataset-preview.

type distribution and acoustic characteristics between basins can highlight the importance of region-specific approaches to underwater noise management.

While the current dataset provides substantial coverage across three European basins, expansion to additional geographic regions would further enhance the dataset's global relevance. The DeuteroNoise project represents a model for international collaboration in marine acoustic research, demonstrating how standardized methodologies and open data sharing can improve scientific progress while raising awareness among the general public about the impacts of anthropogenic noise on marine ecosystems. While the present release provides calibrated, multi-basin recordings and vessel-synchronised annotations, several extensions are planned to enhance its scientific scope. Future versions of the DeuteroNoise Dataset will incorporate harmonised environmental metadata (for example wind speed, wave height, and rainfall), detailed bathymetric information, and additional recording sites to expand geographic coverage. These additions will support more advanced applications, including propagation modelling, long-term SPL trend analysis, ecological impact studies, and the development of machine-learning models for vessel classification. By progressively integrating these elements, the dataset will offer an increasingly comprehensive resource for underwater acoustic research across disciplines (Table 2).

CRediT authorship contribution statement

Ignasi Nou-Plana: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Adrian Teaca:** Writing – review & editing, Supervision, Resources, Investigation, Data curation. **Giovanni Zambon:** Writing – review & editing, Supervision, Resources, Investigation, Data curation. **Rosa Ma Alsina-Pagès:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Dr. Xavier Martínez de Osés and Norbert Gelardo from UPC for providing access to Espai Vela, and facilitating the necessary measurements. The authors also gratefully acknowledge Teodor Muşat from GeoEcoMar-Constanţa for his crucial support and hands-on assistance during the deployments. This research has been funded by JPI Oceans, Project NoiseInTheSea-2022–0011 "Characterization of maritime noise in different european basins and its impact on ecological relevant deuterostome invertebrates", via the Spanish Ministerio de Ciencia, Innovación y Universidades within the call Proyectos de Colaboración Internacional and the European Union NextGenerationEU/PRTR PCI2022-135063–2 and PCI2022-135017–2. This study was also co-funded by the Departament de Recerca i Universitats (Generalitat de Catalunya) under Grant Ref. 2021-SGR-01396 (HER).

Data availability

Data will be made available on request.

References

[1] Shannon G, McKenna MF, Angeloni LM, Crooks KR, Fristrup KM, Brown E, Warner KA, Nelson MD, White C, Briggs J, et al. A synthesis of two decades of research documenting the effects of noise on wildlife. Biol Rev 2016;91(4):982–1005.
[2] Farcas A, Thompson PM, Merchant ND. Underwater noise modelling for environmental impact assessment. Environ Impact Assess Rev 2016;57:114–22.
[3] Zhu P, Zhang Y, Huang Y, Zhao C, Zhao K, Zhou F. Underwater acoustic target recognition based on spectrum component analysis of ship radiated noise. Appl Acoust 2023;211:109552.
[4] Bittencourt L, Carvalho RR, Lailson-Brito J, Azevedo AF. Underwater noise pollution in a coastal tropical environment. Mar Pollut Bull 2014;83(1):331–6.
[5] Finneran JJ, Schlundt CE, Carder DA, Clark JA, Young JA, Gaspin JB, Ridgway SH. Auditory and behavioral responses of bottlenose dolphins (*tursiops truncatus*) and a beluga whale (delphinapterus leucas) to impulsive sounds resembling distant signatures of underwater explosions. J Acoust Soc Am 2000;108(1):417–31.
[6] Wenz GM. Review of underwater Acoustics research: noise. J Acoust Soc Am 1972;51(3B):1010–24.
[7] Erbe C. Underwater acoustics: noise and the effects on marine mammals. A pocket handbook 2011;164(9):10–35.
[8] Zambon G, Benocci R, Sabbadin G, Manni L, Candiani S, Bozzo M, Pennati R, Chatzigeorgiou M, Norland S, Cavazos Contreras IM, et al. The european deuteronoise project: impact of anthropogenic noise on marine invertebrate deuterostomes. In: Inter-Noise and Noise-Con congress and conference proceedings, vol. 270. Institute of Noise Control Engineering; 2024, p. 9957–61.
[9] European Parliament, Council of the European Union, Directive 2008/56/ec of the european parliament and of the council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (marine strategy framework directive), Official Journal of the European Union, L164, 19–40, 2008, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32008L0056.
[10] Nou Plana I, Vaquerizo Serrano J, Freixes M, Arnela M, Rodriguez Quintana E, Cañestro C, Alsina Pagès RM. Analysis of the underwater acoustic environment in the Catalan shore: badalona and port of barcelona cases in the framework of the deuteronoise jpi-oceans project. In: Inter-Noise and Noise-Con congress and conference proceedings, vol. 270. Institute of Noise Control Engineering; 2024, p. 216–27.
[11] Hildebrand JA. Anthropogenic and natural sources of ambient noise in the ocean. Mar Ecol Prog Ser 2009;395:5–20.
[12] Merz S, Kinns R, Kessissoglou N. Structural and acoustic responses of a submarine hull due to propeller forces. J Sound Vib 2009;325(1–2):266–86.
[13] Guerra F, Grilo C, Pedroso NM, Cabral H. Environmental impact assessment in the marine environment: a comparison of legal frameworks. Environ Impact Assess Rev 2015;55:182–94.
[14] Trenkel VM, Ressler PH, Jech M, Giannoulaki M, Taylor C. Underwater Acoustics for ecosystem-based management: state of the science and proposals for ecosystem indicators. Mar Ecol Prog Ser 2011;442:285–301.
[15] Guan S, Brookens T, Vignola J. Use of underwater Acoustics in marine conservation and policy: previous advances, current status, and future needs. J Mar Sci Eng 2021;9(2):173.
[16] Liu F, Shen T, Luo Z, Zhao D, Guo S. Underwater target recognition using convolutional recurrent neural networks with 3-d mel-spectrogram and data augmentation. Appl Acoust 2021;178:107989.
[17] Yao Q, Wang Y, Yang Y. Underwater acoustic target recognition based on data augmentation and residual CNN. Electronics 2023;12(5), https://doi.org/10.3390/electronics12051206, https://www.mdpi.com/2079-9292/12/5/1206.
[18] ZoBell VM, Frasier KE, Morten JA, Hastings SP, Peavey Reeves LE, Wiggins SM, Hildebrand JA. Underwater noise mitigation in the santa barbara channel through incentive-based vessel speed reduction. Sci Rep 2021;11(1):18391.
[19] Mobley JR. Results of marine mammal surveys on US Navy underwater ranges in Hawaii and Bahamas. Final report submitted to Office of Naval Research, Marine Mammal Program; 2004.
[20] Erbe C. Underwater noise of small personal watercraft (jet skis). J Acoust Soc Am 2013;133(4) EL326–EL330.
[21] Roth EH, Schmidt V, Hildebrand JA, Wiggins SM. Underwater radiated noise levels of a research icebreaker in the central Arctic ocean. J Acoust Soc Am 2013;133(4):1971–80.
[22] Santos-Domínguez D, Torres-Guijarro S, Cardenal-López A, Pena-Gimenez A. Shipsear: an underwater vessel noise database. Appl Acoust 2016;113: 64–9.
[23] Irfan M, Jiangbin ZHENG, Ali S, Iqbal M, Masood Z, Hamid U. Deepship: an underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification. Expert Syst Appl 2021;183:115270.
[24] Nathala SS, Yakkati RR, Dayal A, Manikandan MS, Zhou J, Cenkeramaddi LR. Vessel type classification utilizing underwater acoustic data and deep learning. In: 2024 IEEE 19th conference on industrial electronics and applications (ICIEA). IEEE; 2024, p. 1–6.
[25] Nou-Plana I, Freixes M, Vaquerizo-Serrano J, Arnela M, Cañestro C, Quintana ER, Teaca A, Chatzigeorgiou M, Ristoratore F, Zambon G, et al., The deuteronoise dataset: an initial exploration of an underwater vessel sound dataset within the framework of a jpi-oceans project. In: Artificial intelligence research and development. IOS Press; 2024, pp. 260–9.
[26] Nou-Plana I, Teaca A, Gazzada F, Freixes M, Zambon G, Alsina-Pagès RM. An open dataset for exploring underwater soundscapes and vessel noise in barcelona, constanţa and lagoon of venice basins. In: Proceedings of the 11th convention of the european Acoustics association (forum acusticum 2025). Málaga, Spain: European Acoustics Association; 2025, p. 260–9.
[27] Inc. S. Streamlit, https://streamlit.io, [Accessed: 30 Sep 2025].