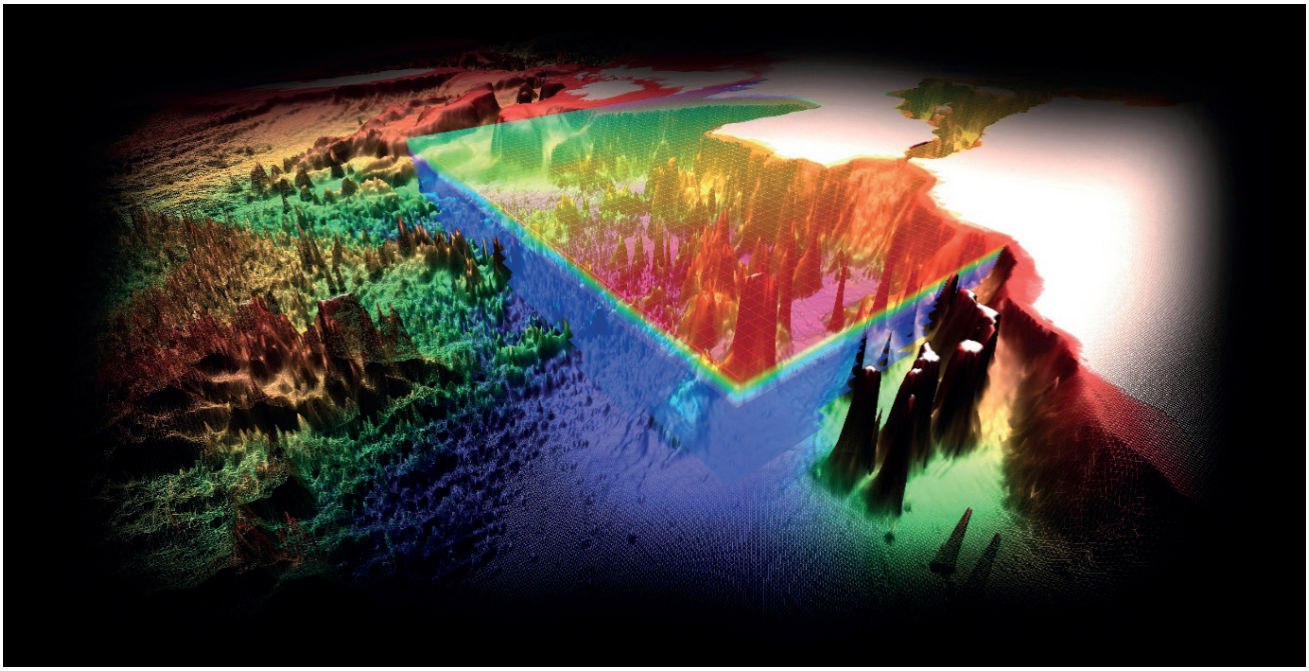
ISSN 1726-5886

CIESM
WORKSHOP
MONOGRAPHS

53

An ocean of gradients - towards a 3D mapping and visualization of high-resolution marine data



The Mediterranean Science Commission

Cover: 3D interactive application prototype displaying a combined bathymetric grid (GEBCO and EMODNET 2020) with gridded 3D aggregations of temperature data from March 2013 (NOAA National Centers for Environmental Information, 2013 - Global NCOM Region 2 dataset).
Credit: Kaveh Rassoulzadegan, CIESM.

CIESM Workshop Monographs

An ocean of gradients – towards a 3D mapping and visualization of high-resolution marine data

CIESM Workshop Monographs ◊ 53.

This collection offers a broad range of titles in the marine sciences, with a particular focus on emerging issues. CIESM Monographs do not aim to present state-of-the-art reviews; they reflect the latest thinking of researchers gathered at CIESM invitation to assess existing knowledge, confront their hypotheses and perspectives, and to identify the most interesting paths for future action.

A collection founded by Frédéric Briand.

CONTENTS

AN OCEAN OF GRADIENTS — TOWARDS A 3D MAPPING AND VISUALIZATION OF HIGH-RESOLUTION MARINE DATA

AN OVERVIEW [1]

[1] To be cited as: D'Auria G., El Bour M., Furla P., Giovannelli D., Giuliano L., Jebber M., Moschella P., Œn Z.B., Pinheiro L., Rassoulzadegan K., Reyes Prieto M., Richardson L., Schmitt T., Tamames J., Tassin Campanella V., Vitale Brovarone A., Yakimov M., Yücel M., Zaaboub N. 2025. An ocean of gradients - towards a 3D mapping and visualization of high-resolution marine data – An overview. pp 5 – 28 In CIESM Monograph 53 [L Giuliano and A. Rodriguez y Baena, eds.] CIESM Publisher, Paris, Monaco, 162 p.

This synthetic overview was first outlined during an immersive CIESM Workshop (Monaco, International Hydrographic Organization, March 19-21, 2024). In the months that followed, it was developed based on written contributions from participants, coordinated by Monia El Bour, chair of the CIESM Committee on Marine Microbiology & Biotechnology, and Luis Pinheiro, chair of the Marine Geosciences Committee. Laura Giuliano, editor of the CIESM Monographs Series, and Alessia Rodriguez y Baena, who co-edited this volume, reviewed the entire manuscript, with particular attention to the introductory chapter. A special thanks is also extended to Mustafa Yücel for his critical reading of the Overview chapter of the monograph. The authors are grateful to Annelyse Gastaldi for the timely physical production of the volume at CIESM headquarters. We also extend our sincere gratitude to R.A. Luigi Sinapi and Sarah Jones (International Hydrographic Organization) and Dorota Czerucka (Centre Scientifique de Monaco). Their invaluable logistical support, dedicated efforts, and time were instrumental in the successful co-hosting of CIESM Workshop 53. A special thanks for their generous arrangement for participants to visit the Principality Aquarium during Monaco Ocean Week 2024, which significantly enhanced the workshop experience.

1. BACKGROUND

Understanding the spatial heterogeneity of marine environments is fundamental to comprehending the complex dynamics of ocean ecosystems. The ocean is characterized by pervasive variations in key properties such as temperature, salinity, and nutrient concentrations, forming what can be conceptualized as an "ocean of gradients". These gradients play a crucial role in structuring habitats and influencing the distribution of marine life (Nezlin 2010). Effectively analysing and interpreting the increasing volume of high-resolution marine data necessitates the use of sophisticated tools capable of visualizing and quantifying these multi-dimensional gradients.

Geographic Information Systems (GIS) and related software have become indispensable for exploring the spatial complexity of ocean gradients. These tools allow researchers to integrate diverse datasets, including bathymetry, surficial geology, and hydrographic information, to create comprehensive representations of the marine environment (Kostylev *et al.* 2001). By visualizing data in a spatial context, scientists can identify critical habitat features and understand how environmental gradients influence their characteristics (Tufekci *et al.* 2010). For instance, mapping benthic habitats often relies on overlaying bathymetric data with information on sediment type and water column properties, revealing how gradients in these factors contribute to habitat complexity and suitability for different species.

Furthermore, the analysis of oceanographic gradients extends beyond static spatial representations. Understanding the interplay between different environmental variables and their combined influence on marine ecosystems requires tools that can handle multi-layered data. By integrating data on temperature, salinity, nutrients, and other relevant parameters, researchers can gain insights into the complex relationships that drive ecological patterns (Nezlin 2010). The ability to visualize these gradients in a spatially explicit manner allows for the identification of key ecological boundaries, such as oceanic fronts and upwelling zones, where sharp changes in environmental conditions can lead to distinct biological communities.

In conclusion, the study of ocean gradients is crucial for advancing our understanding of marine ecosystems. The development and application of sophisticated tools, including GIS and specialized oceanographic software, enable researchers to effectively analyse and visualize the multi-dimensional nature of these gradients. By integrating diverse datasets and employing spatial analysis techniques, scientists can unravel the complex relationships between environmental variability and the distribution of marine life, ultimately contributing to more informed conservation and management efforts (Kostylev *et al.* 2001; Tufekci *et al.* 2010; Nezlin 2010).

2. DEFINING OCEAN GRADIENTS

Having discussed the widespread presence of oceanic gradients and the utility of tools such as GIS in their analysis, it's vital to precisely define what a "gradient" entails. This section, therefore, broadens the concept beyond basic interpretations, exploring its multifaceted spatial, temporal, and variable characteristics.

The Merriam–Webster dictionary defines "gradient" as "the change in the value of a quantity (such as temperature, pressure, or concentration) with change in a given variable and especially per unit distance in a specified direction" (Merriam–Webster n.d.). While this definition offers a foundational understanding, it necessitates further elaboration to enhance its clarity and applicability across diverse contexts.

The term "distance" in the definition is particularly crucial and warrants specific attention, as it can encompass various forms of measurement. It may refer to Euclidean distance or another metric. Importantly, "distance" can also denote temporal separation, such as observations taken at different times, or represent a spatiotemporal dimension. This is evident in phenomena like the El Niño-Southern Oscillation (ENSO), a large-scale recurring ocean-atmospheric phenomenon where observations are recorded across various locations and times. The variability in distance scales, from micro to macro, ultimately depends on the specific research question.

Moreover, the phrase "change in a given variable" also requires expansion. This variable can be any directly observable quantity (Figures 1a, 1b), such as temperature readings from a conductivity-temperature-depth (CTD) profile–a tool that records vertical structure of marine systems. Alternatively, the variable might be an inferred measure derived from primary observations, like the variance of a measured quantity (Figure 1c, 1d) or fluctuations in periodic components of a system (Figure 1e), as observed in Quaternary sea-level curves around 800 kyr ago (Paillard 2001).

Beyond individual measurements, entire systems or subsystems within a larger framework can undergo either abrupt shifts or gradual transitions from one state to another over time. These

changes are often termed "tipping points," representing critical moments where a system's behavior abruptly shifts, leading to significant alterations in its dynamics. Tipping points can be viewed as manifestations of gradients, as they reflect pronounced changes in the system's state variables across spatial or temporal scales. These shifts may arise from nonlinear feedback mechanisms that can amplify small perturbations, pushing the system towards a new stable state or an unstable condition. Changes in feedback amplitude, periodicity or directionality constitute a type of gradient that is beyond directly measured variables and it is well suited to describe the complex relationship present at the ecosystem and planetary scale. Understanding tipping points and their relationship to gradients is essential for predicting and managing the complex behavior of natural systems, particularly under anthropogenic forcing.

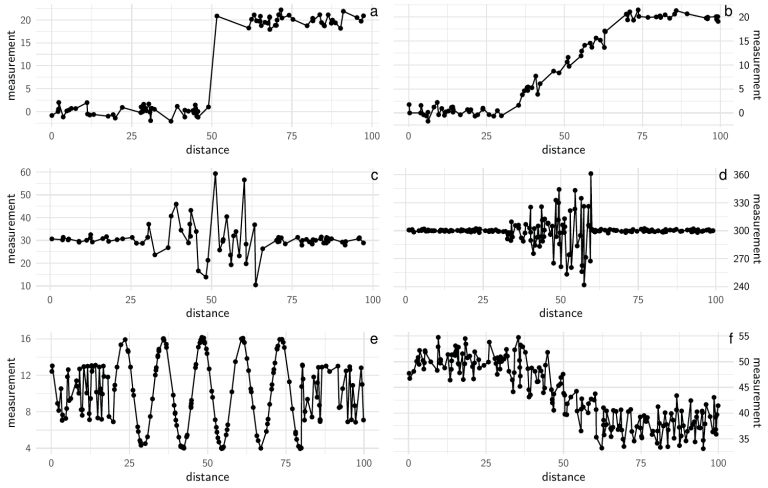


Figure 1. These six scenarios illustrate various simulated gradients: (a) a distinct shift in the overall level; (b) a period of stability, followed by a linear increase, and then a return to a new stable level; (c)and (d) a noticeable change in data variance occurring after a critical point, with the variance subsequently returning to its original state. This example highlights the concept of "temporal distance." (e) changes in the periodic component of a variable; (f) variability is stable over time but the system is slowly drifting toward a new state. The figures are inspired by Lindeløv (2020), and the corresponding code for reproduction can be found at https://github.com/zboraon/gradients_monograph_simulation/tree/main .

In refining these concepts, it becomes apparent that the definition of a gradient extends beyond merely measuring changes across distances in physical space. It also involves changes across temporal and spatiotemporal scales and can encompass both explicit and implicit changes in observed variables. This broadened perspective is crucial for accurately describing gradients in a scientific context.

The concept of a gradient applies broadly to geology and geological features, encompassing both their manifestation and their causes. A prime example is the variation in temperature with Earth's depth, known as a geothermal gradient or geotherm. Geothermal gradients can change spatially due to the intrinsic properties of the rock, such as the concentration of heat-producing radiogenic elements. They can also shift in response to localized events like tectonic activity. Beyond these directly measured variables, dynamic changes in a system's feedback loops—such as their amplitude, frequency, or direction—also constitute a type of gradient. This concept is well-suited for describing complex relationships at both the ecosystem and planetary scales,

particularly in relation to the Earth's internal systems (*e.g.,* magma and geothermal processes).

Nevertheless, the concept of gradient extends to a much wider variety of physicochemical and spatiotemporal features. This includes, but is not limited to, pressure variations, rock and fluid compositions, mechanical properties, tectonic stresses and strains, porosity and permeability, the state of and distance from thermodynamic equilibrium (*e.g.,* metastable, unstable, stable), and geomorphological features. Additionally, the geological concept of a gradient is intrinsically scale-dependent, with detectable gradients ranging from the nanoscale to the scale of the Earth itself. Geological gradients are also inherently transient or evolving, with rates and lifetimes spanning from geologically instantaneous phenomena, such as strain gradients during a fault-localized seismic event, to millions or billions of years, like lithospheric-scale, plate tectonic-driven modifications.

Geological manifestations, such as rocks, magmas, sediments, faults, and gravitational processes, can be either the cause or the consequence of gradients. For instance, chemical gradients in rocks and/or fluids can arise from initial compositional disequilibria between two neighboring materials or result from the re-equilibration of those materials over space and time. The interconnections among this vast range of parameters, scales, and times can determine or be the consequence of geological gradients at all spatial and temporal scales. Local phenomena, whether in space or time, can profoundly alter larger and longer-lived processes and their associated gradients, imposing feedbacks upon them.

Gravity is a primary driver of Earth gradients and controls several processes contributing to the differentiation and stratification of the crust, ocean and atmosphere. In complex systems, the effect of gravity allows for the creation of complex dynamic (both spatially and temporally) gradients controlling complex geo-biological feedback. Winogradsky columns are a great example of the power of gravity in structuring layers. Winogradsky columns are microbial microcosms conceived by Sergei Winogradsky in the 1880s to reproduce complex microbial ecosystems in the laboratory. They are a great example of the power of gravity in structuring layers. The column's construction starts with preparing a homogenous mixture of natural samples (typically soil or sediments) and water, which is either collected from the natural site or designed in the lab to enrich specific microbial metabolisms. The mixture is then poured into a high aspect-ratio flask, closed at the top and exposed to light. Gravity will immediately start to act on the system favoring size sorting within the sediment column, compacting the sediment grains and favoring the ascent of the gases produced in the deeper layers by anaerobic microorganisms. Gravity and light will favor the presence of specific microniches controlled by geochemical gradients. Within a few days to weeks the column will develop strong geochemical and biological gradients visible as distinct colored layers within the column. The geo-biological spatial gradient developed will change in time providing temporal gradients (Figure 2).

Figure 2. Picture of two months old Winogradsky columns obtained from sediment of the coastal brackish Lucrino Lake near Naples (Italy). The distinct color represents diverse microbial communities occupying distinct ecological niches created by the interaction of gravity, light and microbial metabolisms. The spatial geo-biological gradients are everchanging and provide temporal gradients as well. The different columns started with the same sediments and water, but were amended with different nutrients. Columns realized in the Giovannelli Lab at the University of Naples Federico II.

Biology can directly influence both physical and chemical environmental parameters. Life, especially microbial life, is capable of interacting with a large swath of the periodic table and can oxide and reduce a very large number of molecules. The accumulation, diffusion and transport of these molecules can create strong local gradients that in turn influence other organisms. This happens for example in the sediments hosting Posidonia prairies where sulfate reducing bacteria use the high organic carbon accumulating in the rhizome and produce hydrogen sulfide creating strong sedimentary gradients influencing the distribution of all other burrowing organisms. Similarly, the ability of microbial life to precipitate and dissolve a large number of minerals can alter the physical properties and the flow patterns of fluids in sediments and rocks in the subsurface hosting complex microbial communities. A last example is constituted by the ability of microbial life to create and control the distribution of hyperacidic environments. Environments with pH below 3 on our planet are widespread and generally linked to the presence of sulfide minerals or degassing magmatic chambers. In the presence of sulfide minerals exposed to the surface, like in areas of intense mining activities, specialized microbial consortia oxidize sulfide minerals to sulfuric acid therefore acidifying the surrounding pH to levels toxic to most other life. There are just a few examples of the complex interactions through which biology and geology create and control environmental gradients.

In the Mediterranean Sea, several decades of both sparse and collected geophysical data—including bathymetric, gravimetric, seismic, and magnetic information—reveal a profound diversity of subseafloor geological features, horizontally and vertically. Some data, like heat flow data, are available across the entire Mediterranean surface at rather constant resolutions.

Others, such as lithological features, are sparser and tied to localized direct investigations, such as a few boreholes. Combining and visualizing these existing data, whether sparse or continuous, and both onshore and offshore, provides an invaluable tool to predict and guide investigations into the geology of underwater Mediterranean features, their associated gradients, and their potential control over biological, chemical, and physical manifestations. Furthermore, biological processes and evolutionary adaptations significantly shape and respond to these oceanic gradients, from microbial community distributions influenced by nutrient availability to the emergence of distinct faunal zones driven by pressure and light variations.

3. PRACTICAL WAYS TO USE AVAILABLE RESOURCES

Understanding the ocean's intricate, ever-changing nature requires advanced computer tools. This section explores innovative ways to use technology for better 3D mapping and visualizing marine data. This includes grouping similar environmental areas (what we call "environmental niches"), simplifying vast ocean genetic data (metabolomics pathways), and developing a Microbiome MetaCube (MMC) prototype. These new tools help us bring together various types of data into one easy-to-access system.

3.1. Accessing data

Getting access to diverse marine data is key to understanding and visualizing ocean gradients – those changes in conditions over space or time. We do this by using well-known global data portals like GEBCO, EMODNET, MGnify, EarthChem, and ODATIS, as well as project specific data repositories, such as TARA and MALASPINA, which offer a wide range of large, publicly available datasets. We also use data from more specialized, research-focused initiatives, which often provide specific datasets or have their own dedicated portals. Good examples include data from Deep Hypersaline Anoxic Basins (DHABs), Hydrothermal Vents, the CoEvolve project and the eGEOTRACES portal (which is part of the larger GEOTRACES program and helps visualize its datasets). All these different data sources are excellent for describing and showing these ocean gradients.

Beyond these broad portals, effective data access also means making specific searches within large biological and geochemical databases, such as those holding metagenome data (genetic material from entire communities of microbes). This lets us pull out targeted, interdisciplinary information on various elements vital to marine systems. Such information components may include concentration of redox-active elements, metals, carbon, different pollutants (like plastics), and even pathogens. These combined data access strategies are essential for creating advanced computer tools for managing marine environments, which we'll explain in detail in the next subsections.

3.2. IT tools

3.2.1. Clustering environmental niches

Each geographic location where data is collected acts as a unique example, helping to define the specific characteristics of the living organisms found there—this is their biological niche. By gathering and then grouping these data points, the objective is to build a comprehensive database of Niche Profiles (NP). The idea is to put identical ocean niches into the same NP

group. This fundamental concept isn't just for the ocean; it applies to any ecological setting.

A tool under development is designed to process physical and chemical data obtained from public repositories, with every data point carefully described. Once this complete dataset is assembled, rigorous mathematical modeling is then applied to create an NP classifier—essentially, a system that categorizes these niches.

A crucial part of such a tool's development is its ability to work seamlessly with other oceanographic data visualization software like Ocean Data View (ODV) and the pilot CIESM 3D visualization tool (as described by Rassoulzadegan in this volume). The physical and chemical data drawn from public repositories—used for defining biological niches and building the NP classifier—can be easily connected with ODV and the pilot CIESM 3D visualization tool, among others. This interoperability (the ability for different systems to work together) lets users directly import the generated niche profiles and related data for advanced visual exploration, analysis, and mapping.

This type of tool and its database are expected to produce a significant side benefit: a detailed inventory of marine biological niches. This inventory can potentially be linked with the information presented in Section 3.2.3 (Microbiome MetaCube).

3.2.2. Simplifying ocean microbial metabolic pathways

Metagenomic data (genetic material from entire microbial communities) from the Mediterranean Sea can be analyzed using MetaDAG. This is a new bioinformatics tool specifically designed to model and simplify the complex metabolic networks—essentially, all the chemical reactions—that occur within organisms and whole microbial communities (biomes). Other specialized tools can be used to annotate large metagenomic collections to obtain information on the taxonomic and functional diversity of each sample, such as DRAM, Metabolic or Geomosaic.

In order to uncover which types of microbes are present and what they're doing, specialized tools like Kaiju, SqueezeMeta, PiCrust, GTDBtk, or q2-metnet can be employed. Functional annotation of the metagenomes can be accomplished using a large number of tools, including funprofiler, DRAM, KOfamScan, KEGG, or database specific searches using Diamond or HMM. Several of these tools are packaged into pipelines or available through the Galaxy or KBASE platform. The choice of tool depends on how the genetic information was originally sequenced and the specific objectives of the research. Several of these tools help identify specific KEGG Orthology (KO) gene IDs within the samples, which are like unique barcodes for certain genes involved in metabolic processes. Based on these KO IDs, the metabolic networks can be reconstructed into detailed diagrams called reaction graphs, then into directed acyclic graphs (DAGs), and finally into simplified versions known as compressed metabolic directed acyclic graphs (metaDAGs).

Another important use of this method is building the core and pan-metabolism of these microbial communities. The "core metabolism" refers to the essential chemical reactions shared by most microbes in a sample, while "pan-metabolism" includes all the reactions present across all samples combined. Once data from all relevant samples are collected, the pan-metabolism of Mediterranean microbiomes can be assembled. This comprehensive metabolic dataset is currently being developed and will form the foundation for building a MetaCube.

3.2.3. Microbiome MetaCube (MMC)

A universal way to represent metabolic pathways, those intricate chains of chemical reactions that sustain life, is required. While current tools rely on metabolic maps and network arrangements, challenges persist in comparing and combining these connections, especially in three-dimensional environments where ocean conditions, for example, change continuously, creating diverse gradients.

Some researchers (like Reyes-Prieto *et al.*, discussed in this volume) are actively developing a tool to create standard 3D maps of metabolic pathways. These maps, when layered with metagenomic data from global genetic databases, are designed to highlight the specific metabolic abilities linked to different genetic data points. These 3D representations of metagenomic data could also aim to bring together geological, physical, chemical, and biological data into a single, globally accessible system. This helps contribute to wider efforts in digital ocean mapping and modeling. For example, core and pan-metabolic data described in Section 3.2.2 (Ocean's Metabolomics Pathways Compression Tool) can be visualized within the MMC (MetaCube) environment, as described in Section 5.2.

3.2.4. Integrating physical, geochemical and metabolic data

An urgent need exists to develop innovative visualization and analysis tools that go beyond those currently available and those described in this volume, integrating data points from the physical, chemical, geochemical, and biological domains in novel ways. These tools must address the complexity of natural ecosystems in a truly interdisciplinary fashion. Their primary purpose will be to aid the interpretation of complex geobiological feedbacks that characterize the ocean and to carry predictive power useful for understanding the ecosystem's response to future changes, thereby defining new theoretical and practical approaches to investigation.

4. SAMPLING AND DETECTING GRADIENTS OF RELEVANCE FOR MICROBIAL LIFE

A deeper understanding of microbial life requires moving beyond measures of total biomass to focus on the fundamental units of organization: microbial communities and their interactions. These communities, rather than sheer abundance, drive the net process rates underlying biogeochemical cycles that sustain life on Earth. Capturing the spatial and temporal dynamics of microbial community structure—alongside the gene expression patterns that mediate trophic interactions—is therefore essential to a comprehensive understanding of ecosystem function. The spatial organization of microorganisms, shaped by environmental gradients and evolutionary histories, remains largely unmapped. Uncovering these patterns will require intensive, systematic surveys of microbial genetic and functional diversity across a wide range of ecological and geological contexts.

Here follow two examples of standardized, integrated analyses of the tremendous complexity of natural microbial systems based on innovative technology and a shared vision: to fully understand genomes, they must be studied within the environmental frameworks in which they evolved, function, and continue to adapt.

4.1. Sampling extreme marine interfaces: a methodological imperative for microbial research

Understanding the distribution and functional roles of microbial communities in extreme marine environments demands precise, high-resolution sampling. This is especially true along the steep physicochemical gradients that define the interfaces between seawater and hypersaline brines. These interfaces, often less than two meters thick, are ecological hotspots where intense chemical stratification supports unique microbial metabolisms, like chemolithoautotrophy. However, accessing and accurately sampling these gradients at depths exceeding 3,000 meters poses significant technical challenges.

The need for accurate, gradient-resolving sampling is perfectly exemplified by studies on Deep Hypersaline Anoxic Lakes (DHALs) in the Eastern Mediterranean Sea. Formed from the dissolution of ancient Messinian evaporites, these brine lakes are isolated, high-salinity ecosystems confined within deep depressions. Their exceptional chemical variability and sharp density stratification create extreme, layered environments, ideal for studying microbial adaptation and community structure.

To target these unique environments, exploration begins with high-resolution digital terrain models (with a resolution of 500 meters or less) to pinpoint confined depressions likely to host brines. These candidate basins are then further investigated using 3.5 kHz Chirp swath-bathymetry profilers. These instruments detect the brine-seawater interface by capturing sharp acoustic reflection lines resulting from the high salinity contrast, typically sealed by the basin walls.

Once a promising site is identified, in situ verification is performed using a rosette sampler equipped with conductivity-temperature-depth (CTD) sensors. Key physical parameters—salinity (via conductivity), temperature, and oxygen—are recorded and depth-calibrated to precisely locate the interface and determine its thickness. Sampling at this stage remains qualitative; the exact chemical composition must be confirmed through subsequent laboratory analyses.

Only after this multi-step verification can the interface be effectively sampled. In addition to the traditional single-point sampling approach, a specialized method allows for 10 cm-resolution sampling across the interface layer. This technique ensures discrete collection along the steep salinity gradient.

The process involves several meticulous steps outlined in Figure 3:

1. Initial Descent and Stabilization: The rosette, carrying Niskin bottles and CTD sensors, is slowly lowered (at 0.3—0.6 m/s). The CTD sensor precisely indicates the interface's vertical position by a significant increase in conductivity. The rosette continues into the brine for another 25 meters, then pauses to straighten the cable and stabilize its vertical position.

2. Precise Withdrawal and Bottle Closure: The rosette is then slowly withdrawn from the brine at a speed of less than 0.1 meters per second. Niskin bottles are triggered to close at the identified interface position, ensuring samples are collected from specific, narrow bands within the gradient.

This method provides intact stratified samples, which are crucial for downstream microbiological, geochemical, and genomic analyses.

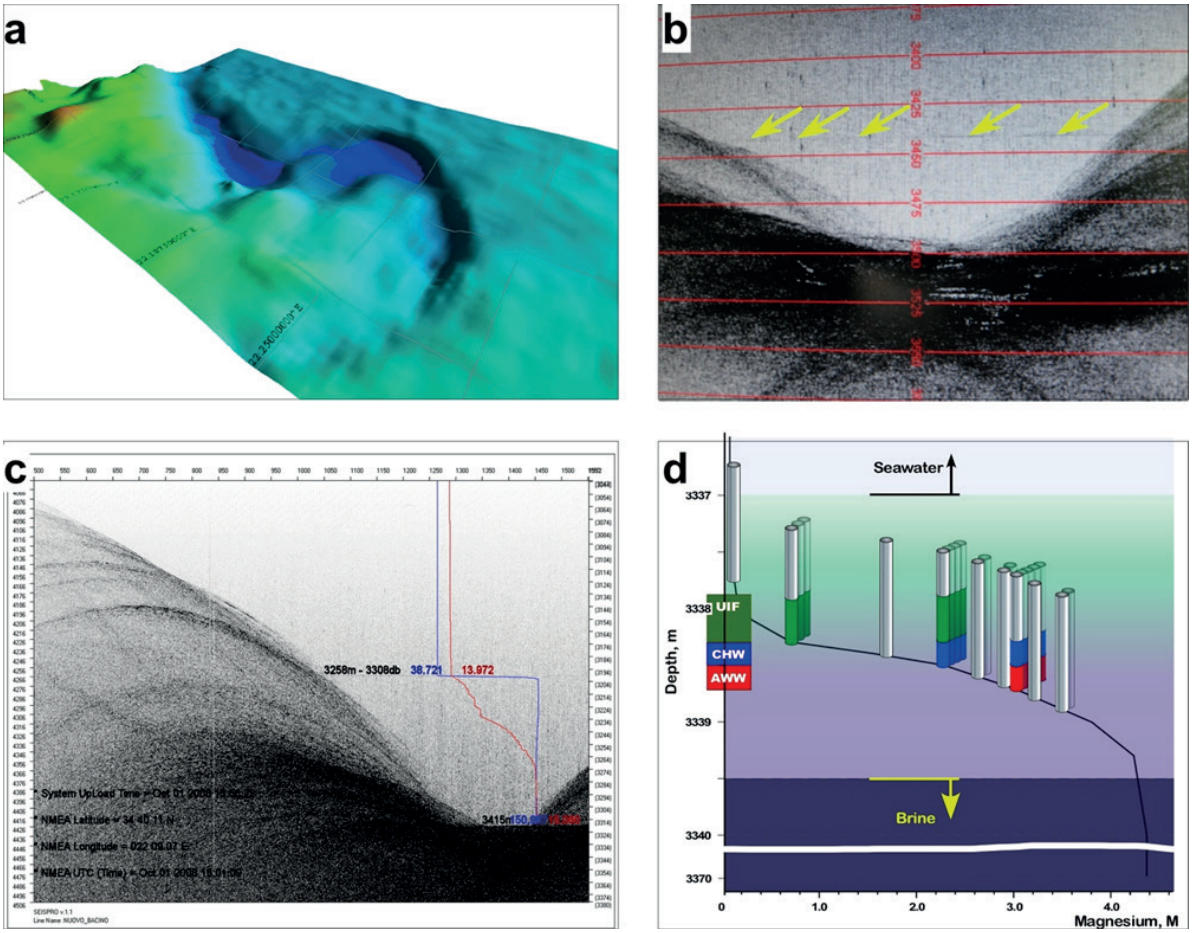
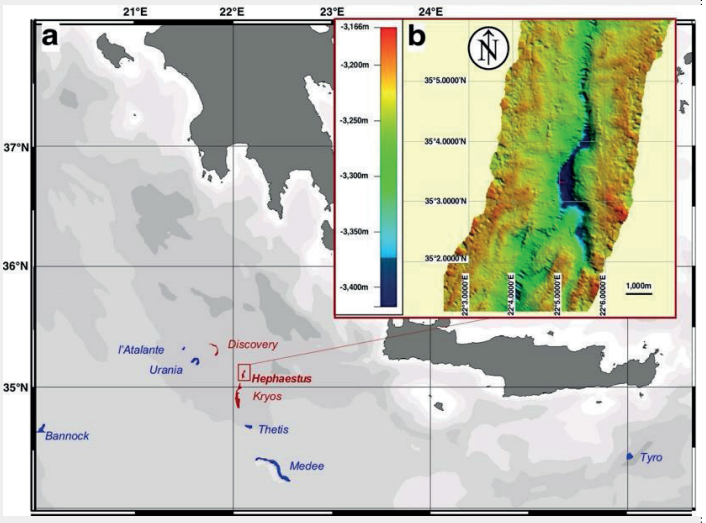


Figure 3. Four key steps in exploring the sharp physicochemical gradients at the seawater-brine interface of Mediterranean Deep Hypersaline Anoxic Lakes (DHALs): (a) Identifying deep, confined depressions using a morpho-bathymetric digital terrain model of the Mediterranean Sea to pinpoint potential DHAL locations. (b) Utilizing 3.5 kHz Chirp swath-bathymetry profiling with multibeam mapping to reveal a perfectly straight reflection line (yellow arrows), indicating the interface. (c) Performing an initial cast with a CTD-equipped rosette to measure essential parameters (conductivity/salinity and oxygen concentration) that highlight the sharp physicochemical gradients. (d) Implementing an interface subsampling strategy with Niskin bottles mounted on a CTD-rosette.

This refined protocol was successfully applied in the identification and characterization of the Hephaestus brine lake, one of the most saline natural hydrological formations known. Detected approximately 10 km north-northeast of DHAL Kryos, the lake is hosted within a steep, sinuous fracture reaching a depth of 3,423 meters. The presence of a sharp brine-seawater interface was first indicated by bathymetric and acoustic data, and later definitively confirmed through CTD profiling and targeted brine sampling. This case vividly illustrates the critical need for high-precision, adaptive sampling methodologies to explore the unique microbial ecosystems associated with marine brine lakes. Without such rigor, the steep physicochemical gradients that define these habitats—and the microbial communities they support—would remain inaccessible, under-characterized, and disconnected from broader biogeochemical understanding.

Deep-sea, hypersaline, anoxic lakes (DHALs)

in the Eastern Mediterranean, likely formed through main geological processes associated to tectonic events, which led to the submarine dissolution of exposed 0.9 to 5.3-million-year-old Messinian evaporites. The chemistry of the DHAL brines is highly variable, primarily due to the dissolution of different layers within the Messinian evaporitic suite (La cono *et al*. 2011; Yakimov *et al.* 2014).



Most Mediterranean DHALs are thalassohaline, meaning they have a seawater-like proportion of major ions, but with a total salinity 7–10 times higher than that of normal seawater.
Due to the density difference of about 20–30% between brines and seawater, DHALs are separated from the overlying seawater by a sharp interface, typically less than 2 meters thick, characterized by very steep chemical gradients and considered a hotspot for biological activity, including chemolithoautotrophy.

4.2. Metagenomics at the crossroads of medical and environmental research

Understanding the intricate interplay between microorganisms and their physicochemical environments—whether in natural ecosystems or within the human body—has become a central focus in microbiome research. Thanks to advancements in experimental methodologies and analytical tools, we can now precisely characterize microbial communities and the environmental gradients that shape them.

For example, metataxonomy (also known as metabarcoding), which uses taxonomic marker genes, and metagenomics (or shotgun sequencing), which captures a sample's entire genomic content, are rapidly transforming our understanding. These approaches place microbes within the environmental frameworks in which they evolved, function, and continue to adapt. Metagenomic approaches, in particular, capture the complexity of microbial communities in situ, offering insights into the spatial and temporal dimensions of microbial diversity, along with the gene expression patterns driving trophic interactions and system-level functionality.

Crucially, both approaches are applicable across environments, enabling a comparative framework that spans oceanographic samples and human-derived material. The basic workflows for DNA extraction are similar regardless of sample origin, though protocols

adapt for differences in biomass density, matrix composition, and preservation needs. Table 1 summarizes key considerations in DNA extraction techniques across different settings.

In human microbiome studies, stool samples are commonly used due to their ease of collection, high microbial load (~106 cells/mL), and straightforward preservation—typically by freezing at -20°C. RNA-focused studies often incorporate stabilizing agents prior to extraction. For environmental samples, where microbial biomass may be more diluted or spatially heterogeneous, field-deployable filtration systems and preservation protocols have been developed to capture and stabilize the microbial signal across diverse physicochemical gradients.

Step	Human Gut Microbiome	Free-Living Environmental Niches
1. Sample Collection	Collect stool samples, swabs, or biopsies from the gut or other matrices.	Collect soil, water, air, or surface samples.
2. Dilution/Concentration		Dilute or concentrate samples to achieve optimal microbial concentration, avoiding/reducing chemical contaminants which may impair the PCR and library preparation steps.
3. DNA Extraction	Extract DNA from the collected samples using usually specialized kits.	Extract DNA from environmental samples using ad-hoc adapted methods or specialized kits.
4. Primer Selection	Choose appropriate primers targeting specific regions of the 16S rRNA gene or other relevant genes to ensure comprehensive coverage and minimize biases.	Choose appropriate primers targeting specific regions of the 16S rRNA gene or other relevant genes to ensure comprehensive coverage and minimize biases.
5. Library Preparation	Prepare DNA libraries for sequencing, including amplification and barcoding.	Prepare DNA libraries for sequencing, including amplification and barcoding.
6. Sequencing	Perform sequencing using platforms like Illumina or PacBio.	Perform sequencing using platforms like Illumina or PacBio.
7. Data Analysis & Visualization	Analyze sequencing data to identify microbial taxa and functions. Visualize results using bioinformatics tools.	Analyze sequencing data to identify microbial taxa and functions. Visualize results using bioinformatics tools.

Tabe 1. Considerations for microbiome analysis comparing methodological steps for human gut and free-living environmental samples.

These studies demonstrate that the human microbiota, much like microbial communities in oceans, soils, or sediments, plays a critical role in regulating key metabolic processes. In both cases, microbial communities aren't passive residents but active agents in maintaining

system homeostasis—be it ecosystem function or human health. Disruptions in these microbial systems, whether due to pollutants in nature or antibiotics and dietary shifts in humans, can trigger dysbiosis, undermining the resilience and functionality of the host environment.

New techniques such as single-cell genomics, spatial transcriptomics, and in situ hybridization are extending the analytical frontier. These allow researchers to examine microbial function and identity at the level of individual cells and in direct relation to their micro-environmental conditions.

Finally, the success of molecular analyses heavily depends on the computational tools used to process and interpret sequencing data. Metataxonomy remains widely adopted due to its cost-effectiveness and the relative maturity of its bioinformatics pipelines. However, its performance is influenced by primer design, which must be carefully chosen to ensure broad and unbiased taxonomic coverage for the target sample.

Overall, the integration of advanced sampling strategies, high-throughput molecular techniques, and powerful bioinformatics platforms is opening new pathways for characterizing microbial communities and the gradients that structure them—whether in the open ocean, coastal sediments, or the human gastrointestinal tract. This convergence of technology and ecological insight is central to the next generation of microbiome science.

5. PROTOTYPE EXERCISES TO VISUALIZE GRADIENTS IN 3D

The pilot 3D visualization tool being developed under the auspices of CIESM (Rassoulzadegan, this volume) represents a significant advancement in marine data analysis. This versatile and user-friendly tool was developed to improve upon previous limitations in assessing gradients within the Mediterranean Sea. It facilitates the visualization and interpretation of complex marine data, including biological, physical, chemical, and even genetic parameters. By allowing for the overlay of these diverse datasets onto 3D bathymetric maps, among other features, the tool enables volumetric visualization of nutrient hotspots and enhances the detection of biogeochemical gradients. This capability is crucial for understanding the intricate interactions and environmental changes occurring across the Mediterranean Sea. Below, two examples will be presented to demonstrate how CIESM scientists collaborate to test this tool through prototypes, visualizing and analyzing physicochemical, nutrient, and genetic data to gain a more comprehensive understanding of the marine environment.

5.1. Physicochemical parameters and nutrients

The pilot CIESM 3D visualization tool is powerful for displaying interactions, geochemical gradients, and areas of high reactivity and elemental flux within marine systems. This approach significantly enhances our ability to assess gradients intuitively compared to other methods. Specifically, one of its prototypes allows for the volumetric visualization of nutrient hotspots and improves the detection of biogeochemical gradients by overlaying physicochemical and nutrient data onto 3D bathymetric maps. These visualizations are most useful for identifying both vertical and horizontal environmental gradients across the Mediterranean Sea. The distinct features of the Mediterranean basins, along with Atlantic water inflow, justify using both 2D and the pilot CIESM 3D visualization tool's mapping techniques to effectively display these gradients.

Biogeochemical processes, involving the transformation of organic and inorganic components, are influenced by factors like temperature, salinity, photosynthetically available radiation (PAR), and other biogeochemical properties, all contributing to environmental gradients. Temperature is a key driver in assessing these gradients. The CIESM 3D visualization presented in this section clearly shows the Mediterranean's temperature gradient: cooler waters (16–18°C) in the west and warmer waters (21–22°C) in the east, particularly in the Levantine Basin and near Egypt and southern Turkey (Figure 4). This visualization helps interpret how oceanographic factors, such as Atlantic water inflow and regional circulation, shape this gradient.

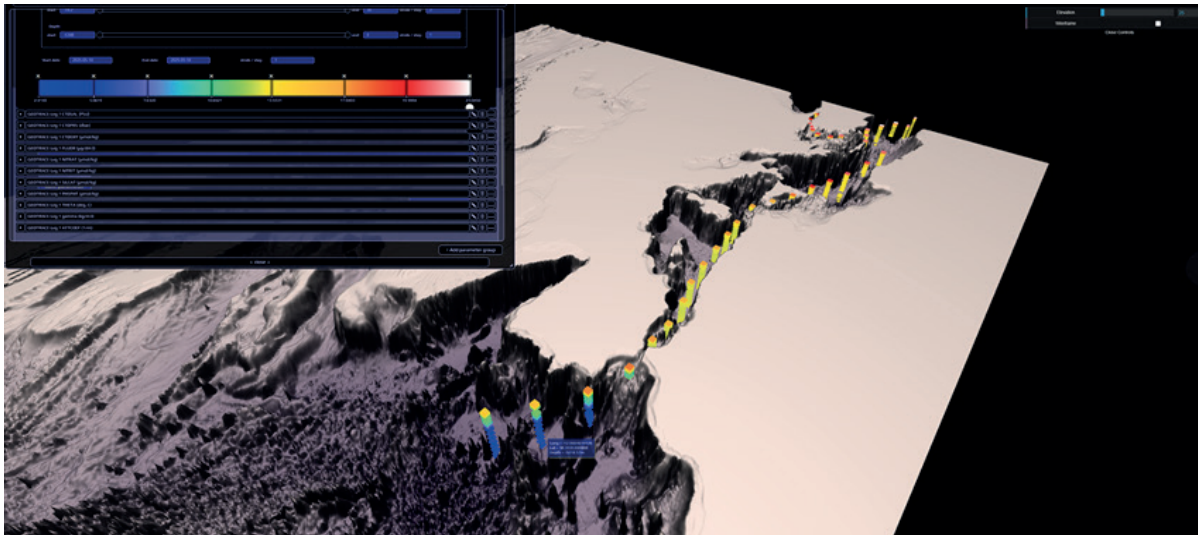


Figure 4. Temperature (C°) in the Mediterranean Sea along the Southern Leg (S) of the GEOTRACES-A04N section during R/V Pelagia cruises (64PE370/374) in 2013.

The CIESM 3D visualization significantly aids in understanding and interpreting nutrient distributions. For instance, while salinity plots show distinct water masses and nutrient distributions indicate oligotrophic conditions in the eastern basin, the CIESM 3D visualization of nitrate concentrations in the 0–50 m layer (Figure 5) aligns well with modeling results. This tool facilitates the interpretation of a clear gradient in deeper waters, with higher nitrate concentrations near the Atlantic inflow at the Strait of Gibraltar, gradually decreasing eastward.

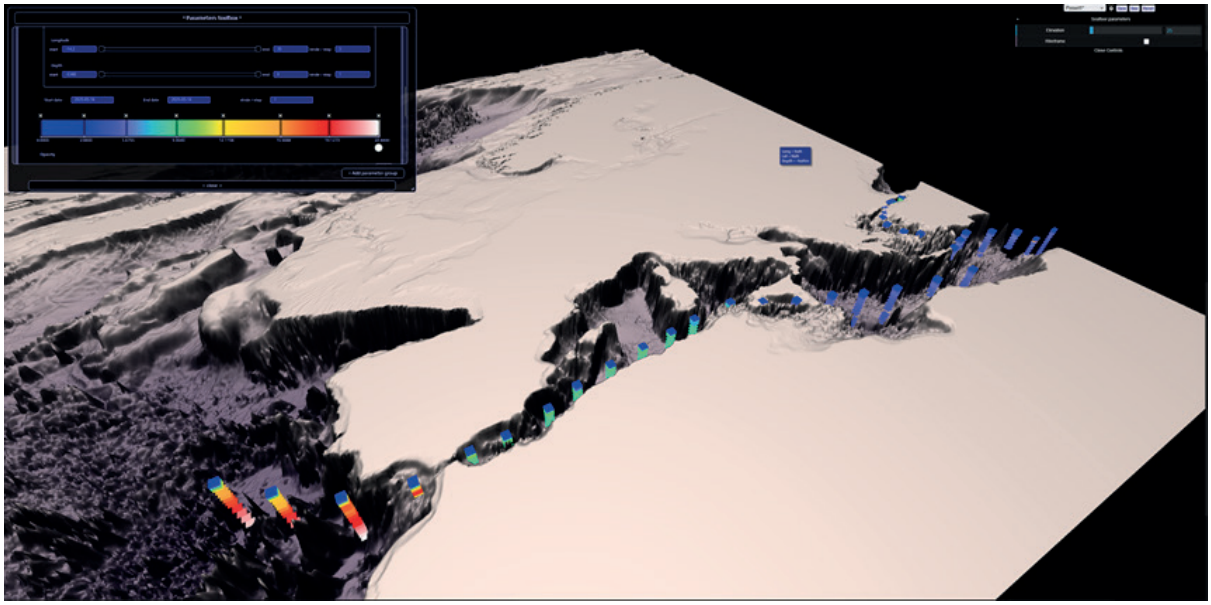


Figure 5. Nitrates (µmol/Kg) in the Mediterranean Sea along the Southern Leg (S) of the GEOTRACES-A04N section during R/V Pelagia cruises (64PE370/374) in 2013.

This highlights how the pilot CIESM 3D visualization tool doesn't just display data; it also helps us understand complex nutrient patterns, like the decrease in concentrations along the anti-estuarine circulation from the Strait of Gibraltar, where nutrient-poor Atlantic Water enters. We can also effectively visualize and interpret the main sources of dissolved silicate in the Mediterranean Sea, including continental river inputs and submarine groundwater discharge, using these 3D models.

5.2. Human gut metabolomics

The pilot CIESM 3D visualization tool proves to be an exceptionally powerful resource for exploring the intricate details of complex biological systems, including microbial metabolic networks within the gut. This advanced tool significantly enhances our ability to analyze and interpret the vast datasets generated in microbiome research, moving beyond traditional 2D representations to offer a truly immersive and insightful view of microbial functions. Building on this foundation, the MetaCube application, a prototype of the pilot CIESM 3D visualization tool, provides a dedicated 3D viewer for metabolic reconstructions, facilitating the visualization and comparison of complex metabolic maps.

Microbiome research aims to understand how microorganisms interact with their environments. Advances in metataxonomy and metagenomics now allow precise characterization of microbial communities, revealing their evolution and function. Metagenomics, in particular, captures the spatial, temporal, and functional complexity of microbes in situ, offering insights into gene expression and system-level functionality. Both approaches are applicable across diverse environments, from oceans to human samples, with adaptable DNA extraction protocols. For human studies, stool samples are common, while environmental samples use specialized filtration and preservation.

In this example, we apply the pilot CIESM 3D visualization tool, through its MetaCube prototype, to understand how different dietary patterns—specifically Western and Korean diets—affect the structure of microbial metabolic networks within gut microbiome samples. The Western diet is typically characterized by a high intake of animal proteins, saturated fats, refined sugars, and processed foods. In contrast, the Korean diet emphasizes plant-based foods, including vegetables, legumes, and fermented products, alongside rice, with a lower intake of animal fats and processed sugars.

These dietary differences were reflected in the metabolic networks reconstructed using MetaDAG, a web-based tool that compares metabolic networks from microbiome data (Palmer *et al.* 2025; available at https://bioinfo.uib.es/metadag/). As a demonstration, MetaDAG was applied to analyze 24 gut microbiome samples from Study MGYS00000394 in the MGnify database (Richardson *et al.* 2023). These samples originated from 12 individuals, each undergoing both Korean and Western dietary interventions, resulting in paired samples for comparative metabolic analysis. MetaDAG revealed distinct metabolic signatures associated with each diet, highlighting how dietary composition influences microbial community function. For instance, the Korean diet was linked to metabolic pathways related to fiber fermentation and short-chain fatty acid production, beneficial for gut health. Conversely, the Western diet showed enrichment in pathways tied to lipid metabolism and amino acid degradation, potentially associated with pro-inflammatory states.

The output datasets generated by MetaDAG, which MetaCube accepts, include core metabolic pathways (shared between diet types), pan-metabolic pathways (encompassing all samples), and individual sample datasets. The complexity of these metabolic maps makes them challenging to visualize and compare effectively in traditional formats. This is precisely where the pilot CIESM 3D visualization tool, realized as the MetaCube prototype, proves invaluable.

MetaCube serves as a dedicated 3D viewer for metabolic reconstructions, accepting reaction data obtained from MetaDAG. It functions as a dynamic reference map where each layer corresponds to a specific metabolic pathway, connecting enzymes involved in different pathways with straight vertical lines. Its fully navigable interface offers easy-to-learn, highly interactive 3D kinetics, leveraging the Three.js Javascript library and other open-source utilities under the M.I.T license.

As shown in Figure 6, we used MetaCube to visualize the core metabolism shared between Western and Korean diet samples. This CIESM-powered prototype allows us to clearly compare both core (left) and pan-metabolic (right) representations. For example, elements of pathway 00100 (steroid biosynthesis) are not commonly shared between the two datasets, indicating they are an accessory gene set.

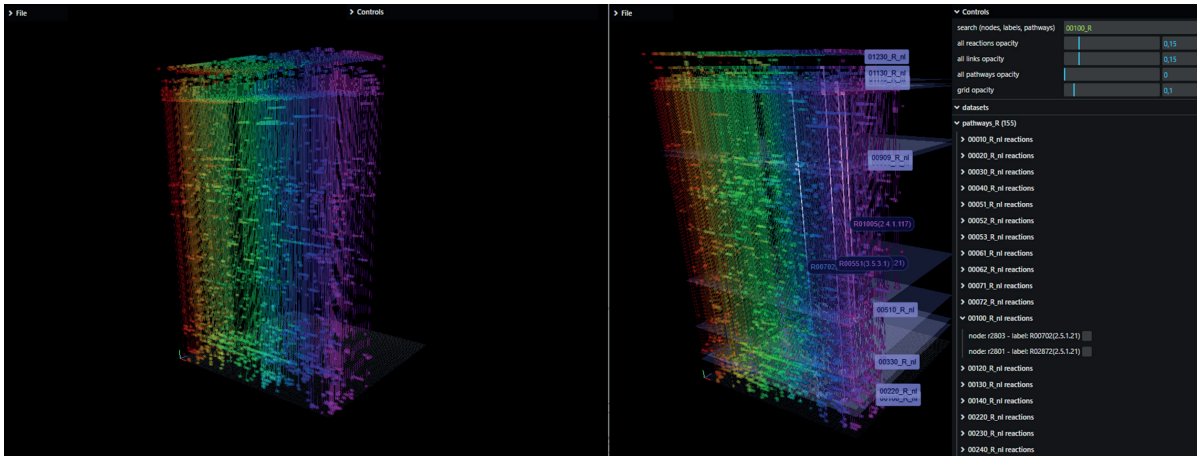


Figure 6. Core versus Pan metagenome representation.

The prototype MetaCube viewer also allows to identify differential pathways or gene features between two datasets. In the next depiction (Figure 7), we are considering the differences between the pan-metabolism of the Western (left) and Korean (right) diets.

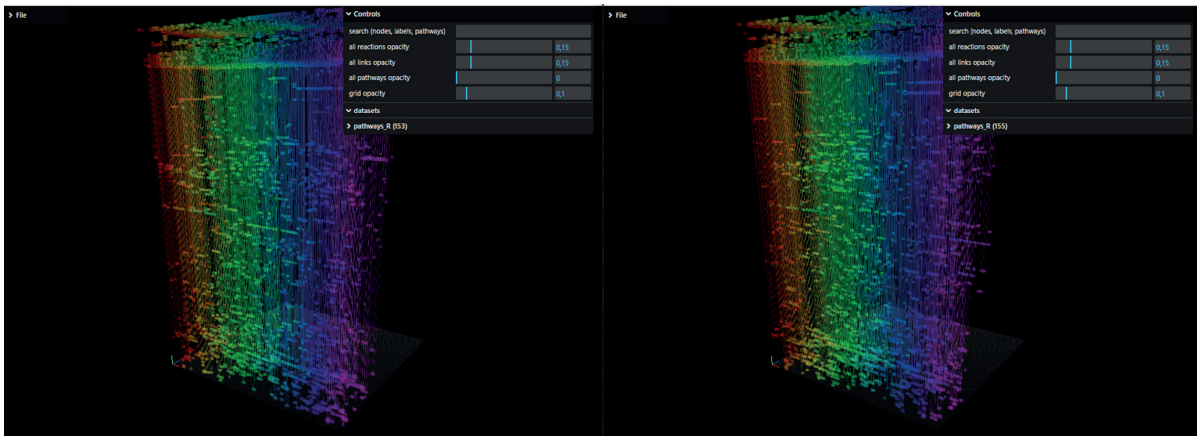


Figure 7. Pan-metabolisms from Western (left hand) and Korean (right hand) diets.

Even in its initial stages, the MetaCube interface, allows users to explore the intricate details of both datasets. This first version already enables highlighting differences in pathways or gene content by searching for specific features. The MetaCube viewer is a powerful visualization tool for metagenomics datasets, allowing us to identify common metabolic pathways or enzymes between samples. This tool, coupled with the reductive and comparative capabilities of MetaDAG, represents a comprehensive approach to metagenomics summarization and visualization.

Historically, the complexity of metagenomics results has made it difficult to establish general exploratory protocols. MetaCube, for the first time, offers a robust framework, to compare metagenomics datasets using a flexible reference platform. This platform allows for mapping every metagenomic result, either through a core or pan-metagenomic context.

MetaCube is currently hosted at metacube.ciesm.org.

6. VALUE OF OCEAN DATA

6.1. Legal implications and value chain of stakeholders

With the advancement of new technologies in visualization, modelling, and prediction, our society is entering a transformative era where ocean data play a critical role in the planning and safety of maritime activities, as well as in the design and implementation of environmental protection measures. In the face of current global crises—biodiversity loss, climate change, and pollution—access to high-quality ocean data and information is more vital than ever. These data are essential to advancing scientific understanding, particularly for complex marine phenomena such as ocean gradients.

However, significant barriers remain within the governance framework of ocean data collection, sharing, and use, which continue to impede scientific progress. Three main challenges can be identified, especially in relation to the use of cutting-edge technological tools such as 3D mapping, visualization, modelling, and predictive analysis for complex ocean phenomenon:

1. The nature of data collection activities
2. The modalities through which data are shared and used
3. The governance framework overseeing these processes

Despite substantial efforts made by States to enhance cooperation in sharing marine scientific data and information, the international legal framework still contains many ambiguities and loopholes. These ambiguities and loopholes complicate access to, and the collection, sharing and use of, ocean data & information. The marine scientific research (MSR) regime under the United Nations Convention on the Law of the Sea (UNCLOS) provides a foundational legal framework for conducting marine research and sharing related data & information. However, the difficulties encountered during the negotiation of UNCLOS left a key gap: the absence of a clear definition of "marine scientific research". This gap has led to divergent interpretations among Coastal States regarding activities conducted within their national jurisdictions. As a result, many activities involving the collection and sharing of marine data face legal and administrative challenges, due to the divergent interpretation of States of what constitutes MSR.

Among the diverse range of ocean data activities, operational oceanography stands out as a particularly illustrative case. It has been the subject of extensive discussions within the Advisory Body of Experts on the Law of the Sea (ABE-LOS) of the Intergovernmental Oceanographic Commission (IOC) of UNESCO. While ongoing disagreements persist among States on the nature of operational oceanography, these discussions ultimately led to the development of an ad hoc legal framework governing the operation of Argo floats. Other emerging technologies used within national jurisdiction—such as satellite-derived data, ocean observations from submarine cables (*e.g.*, SMART systems), as well as drones or autonomous unmanned vehicles—are also currently facing important legal uncertainties both within and outside the European Union. These uncertainties often translate into challenges in obtaining operational authorizations or in collecting, sharing and using the data collected by such technologies. Coastal States, seeking to protect their sovereignty and sovereign rights (including knowledge of resources, activities, and geographic features, including for defence and security purposes), have imposed various limitations. These may include restrictions on data resolution, geographic coverage, or the sharing and final use of the collected data, as well as the ban of certain activities. As a result, a wide range of valuable ocean data which could be collected by advanced technologies and

infrastructures remains outside the scope of the freedom of marine scientific research.

In addition to issues related to data collection, broader challenges also arise concerning the modalities for sharing and using scientific data, particularly when it involves advanced technological tools that could incorporate artificial intelligence. These tools often rely on aggregated data from multiple sources and stakeholders (including private) for visualization, modelling and predictive analysis. A notable example is the Digital Twin of the Ocean, developed under the leadership of the European Union. These technologies introduce a complex set of questions, including:

- Data ownership: Who owns the derivative products generated by the raw data? How to navigate the initial ownership or co-ownership of the raw data (from the Coastal State and/or research operator depending on the collection point) and the new potential ownership of derivative products?
- Model ownership: Who holds rights over AI-generated models or digital twins?
- Liability and responsibility: Who is accountable for data accuracy, errors in prediction, or unintended outcomes - especially when prediction is designed for policy decision-making?
- Traceability and auditability: How can data usage and decision-making processes be transparently tracked across the data chain of stakeholders?
- Jurisdictional and Regulatory Applicability: What legal issues are triggered by the use of the Digital Twin for Ocean by non-EU Members? How can the non-applicability of EU regulations (such as the EU AI Act or Open Data Directive) be navigated, requiring data agreements to be negotiated with non-EU Member States (including CIESM Member States)?

Moreover, the dual-use of ocean data—where scientific information may also serve commercial, security, or strategic purposes—and the increasing value of data when layered or integrated (*e.g.*, through AI or multi-source modelling) are creating new international legal and ethical dilemmas. These issues underscore the urgent need for States to refine existing governance mechanisms to prevent further conflicts over data collection and use, especially in contexts of transboundary study and cooperation.

At the governance level, the ocean data landscape is becoming increasingly complex due to the growing number of public and private stakeholders involved across the data value chain. These include States (coastal and flag States), their national research institutes various sectors, including defense, private sector, data repositories (public and private), operators developing modelling and predictive systems (public and private), and end-users such as policymakers or the private sector. As this ecosystem expands, governing the collection, sharing, and usage of ocean data becomes more challenging.

Since the legal ambiguities and loopholes within the United Nations Convention on the Law of the Sea (UNCLOS) and related international law are unlikely to be resolved in the short to medium term, Member States of the Mediterranean Science Commission (CIESM) — gathering States Members and not Members of the European Union[1] — could explore practical options

1 This distinction is of importance as different legal and regulatory frameworks are applicable to these Member States of the CIESM. Notably, the recent AI Act of the European Union (first ever AI legal framework) is not applicable to non-EU Members, posing potential issues of various standards and protections regarding data quality, bias mitigation and liability, especially in the context of transboundary data sharing and multi-stakeholders systems.

to address current challenges and promote cooperative frameworks for data collection and exchange to improve knowledge of ocean gradients.

Several avenues could be considered, requiring further in-depth study. One pragmatic option would be to leverage the CIESM forum—a direct subject of international law and mandated to foster cooperation, promote scientific research, and facilitate data exchange. Within this context, CIESM could serve as a platform for its Member States to engage in dialogue on ocean data governance issues, with the aim of developing a CIESM Ocean Data Governance Policy. Such a framework could help harmonize approaches among Member States while leveraging the input of the Scientific Council and scientific experts.

In the longer term, Member States might also consider negotiating a specific agreement to address the current legal uncertainties in ocean data governance related to the mandate and scope of work of the CIESM. Such an agreement, which could be limited in scope, could support and strengthen scientific cooperation, facilitate data exchange and use, and promote mutual understanding across Member of the CIESM.

By adopting either approach, States would not only fulfil their legal obligations under international law—including the duties to cooperate, to exchange scientific data and information, and to protect and preserve the marine environment—but also contribute to a forward-looking and innovative model of regional ocean governance; adapted to the diversity of profiles of CIESM Member States. This would demonstrate the Mediterranean region's commitment to modern, science-based cooperation in marine affairs, reinforcing CIESM's role as a key facilitator in this evolving landscape.

6.2. Data Archives and Bioinformatics Resources

The generation of data from biological samples is often the result of costly experiments and sampling campaigns. In the case of marine data derived from either long-term and/or large-scale ocean sampling missions that require substantial financial investment and extensive planning, this is particularly true. Maximizing the value extracted from these samples and their corresponding datasets, which cannot easily be recaptured, is essential to justify the investment.

A key strategy for achieving this maximised value is to ensure that data are made openly available, following any appropriate embargo period to allow the primary researchers to publish their findings. Data should be deposited in public archives accompanied by comprehensive metadata about the sample and experimental protocols that provides essential context. For nucleotide sequencing data, the International Nucleotide Sequence Database Collaboration (INSDC) serves as the central repository, supporting data sharing across its member organizations, National Library of Medicine, National Center for Biotechnology Information (NLM-NCBI) (Robinson *et al.* 2024), European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) (Liang *et al.* 2023), and Research Organization of Information and Systems, National Institute of Genetics (ROIS-NIG) (Quinton *et al.* 2024). These recognized repositories offer long-term data preservation and promote best practices in data standards, metadata capture, data linking, and provenance, thereby facilitating future data reuse and integration into meta-analyses.

While sequence data repositories are well established and widely known, many other resources exist to archive diverse data types. For example, PRoteomics IDEntifications Database (PRIDE) (Gohl *et al.* 2024) for proteomics, MetaboLights (Jeffryes *et al.* 2024) for metabolomics,

BioImage Archive and Electron Microscopy Public Image Archive (EMPIAR) for imaging (Gohl *et al.* 2024; Robinson *et al.* 2024), and Marine Environment Data and Information Network (MEDIN) data archive centre (DAC) (Marsh *et al.* 2024) for geology, geophysics and backscatter. Researchers are encouraged to deposit data in well-maintained public archives whenever possible to ensure data longevity and, critically, to enable linkage with associated datasets. The correct association of multi-modal datasets across resources is fundamental for scalable multi-omic analysis, and provides a basis for ocean modelling. However, as these various resources are independent of each other and therefore unaware of other associated omics datasets, this linking is most often achieved through the use of persistent sample identifiers. Resources such as BioSamples (Harris *et al.* 2024) provide stable identifiers that can be associated with data in other repositories, enabling the connection of diverse analyses performed on a single sample. As the BioSample record also contains sample metadata, such as geolocation and sampling data and time, it supports the linking of datasets in, for example, longitudinal studies.

In addition to describing raw data and samples, it is equally important to document any bioinformatic processing or analysis applied to a dataset. For example, metagenomic sequence data deposited in INSDC can be linked with taxonomic and functional profiles generated through specific bioinformatic pipelines. The interpretability of these profiles relies on transparent documentation of the analysis pipeline, including versioning and accessibility. Resources such as MGnify (Finn *et al.* 2024), Qiita (Zaneveld *et al.* 2018), and Integrated Microbial Genomes & Microbiomes (IMG/M) (Wilke *et al.* 2023) provide versioned, openly documented pipelines, ensuring full provenance for downstream users.

A wealth of marine data is available in public archives and is suitable for the types of studies and meta-analyses discussed during the workshop. However, heterogeneity in the associated metadata can limit data utility. Despite the existence of standards and guidelines-such as those from the Genomic Standards Consortium (GSC) (Filed *et al.* 2011; Gouldsbury *et al.* 2024) and the FAIR (Findable, Accessible, Interoperable, Reusable) data principles (Wilkinson *et al.* 2016), gaps in contextual metadata remain. It is essential to adopt and promote these best practices, capturing rich sample metadata to enhance dataset utility for future research. There are numerous large-scale initiatives underway, including DTO-BioFlow (GA:101112823), Blue-Cloud2026 (GA:101094227), BIOcean5D (GA:101059915) and BlueRemediomics (GA:101082304) to either generate or aggregate standardized, well-described marine data. These efforts offer substantial, high-quality datasets that can be expanded and leveraged to contribute to the aims of this workshop.

7. CONCLUDING NOTES

The workshop “An Ocean of Gradients - Towards a 3D Mapping and Visualization of High-Resolution Marine Data” highlighted the critical need to understand marine environmental heterogeneity and the complex dynamics of ocean ecosystems. Defining gradients as changes across various distances (spatial, temporal, and spatiotemporal) is fundamental to this understanding. Analyzing the growing volume of high-resolution marine data requires sophisticated visualization and quantification tools. The pilot CIESM 3D visualization tool marks a significant advancement, offering a versatile and user-friendly platform for visualizing and interpreting diverse marine data, including biological, physical, chemical, and even genetic parameters. By overlaying datasets onto 3D bathymetric maps, this tool enables volumetric

visualization of nutrient hotspots and enhances the detection of biogeochemical gradients, which is crucial for understanding the intricate interactions and environmental changes occurring throughout the Mediterranean Sea.

The pilot CIESM 3D visualization tool holds immense future potential. Its ability to easily incorporate and visualize data from extensive public repositories like GEBCO, EMODNET, MGnify, TARA, and ODATIS will lead to a more comprehensive and integrated approach to marine data mapping. This interoperability facilitates advanced visual exploration and analysis of generated niche profiles and associated data. Furthermore, the tool's capacity to develop detailed censuses of marine biological niches and contribute to a Microbiome MetaCube (MMC) promises to unify geological, physicochemical, and biological data into a globally accessible format. Crucially, the pilot CIESM 3D visualization tool is being developed with an emphasis on free access, promoting open science principles and maximizing the value of costly experimental and sampling campaigns.

Beyond the tool itself, the workshop also stressed the importance of improving ocean data governance. Significant barriers to data collection, sharing, and use, including legal ambiguities within the UNCLOS framework, continue to impede scientific progress. The dual-use nature of ocean data (for both scientific and commercial/security purposes) and the increasing value of integrated data also create new legal and ethical dilemmas. Addressing these challenges through cooperative frameworks, potentially leveraging CIESM as a platform for dialogue and policy development, is essential for informed conservation and management efforts. By ensuring open access and promoting best practices in data archiving and metadata capture, the scientific community can enhance data utility for future research and contribute to a forward-looking model of regional ocean governance.

References

Field, D., Amaral-Zettler, L., Cochrane, G. *et al.* (2011). The Genomic Standards Consortium. *Public Library of Science Biology*, 9(6), e1001088.

Finn, R.D., Abarca, V., Cottingham, M.G. *et al.* (2024). The MGnify database and analysis resource in 2023. *Nucleic Acids Research*, 52(D1), D708-D714.

Gohl, S., Roos, F., Branje, C. *et al.* (2024). BioImage Archive and EMPIAR: data resources for life sciences. *Nucleic Acids Research*, 52(D1), D731-D737.

Gouldsbury, H.J., Morales, S.E., Mardis, E.R. *et al.* (2024). The Genomic Standards Consortium: updates to the GSC guide to best practices in metadata reporting and recommendations for new types of environmental data. *Nucleic Acids Research*, 52(D1), D738-D745.

Harris, M.A., Bekhoff, A., Hassani, M. *et al.* (2024). BioSamples: a community resource for sample metadata and identifiers. *Nucleic Acids Research*, 52(D1), D715-D721.

Jeffryes, M.N., Gonzalez, A., Barker, K.R. *et al.* (2024). MetaboLights: a public repository for metabolomics data. *Nucleic Acids Research*, 52(D1), D722-D730.

Kostylev, V.E., Todd, B.J., Fader, G.B.J., Courtney, R.C., Dickinson, J.T. and Spagnolo, M. (2001). Benthic habitat mapping on the Scotian Shelf based on multibeam bathymetry, surficial geology and hydrography. *Marine Ecology Progress Series*, 219, 121-137.

Liang, J., Chen, C., Shah, K. *et al.* (2023). The European Bioinformatics Institute's data resources (2023). *Nucleic Acids Research*, 51(D1), D1-D10.

Lindeløv, J.K. (2020). mcp: An R package for regression with multiple change points. *OSF Preprints* doi:10.31219/osf.io/fzqxv.

La Cono, V., Smedile, F. and Bortoluzzi, G. (2011) Unveiling microbial life in new deep-sea hypersaline Lake Thetis. Part I: Prokaryotes and environmental settings. *Environmental Microbiology* 13(8), 2250-68.

Marsh, D., Badaoui, A., Bowler, C. *et al.* (2024). MEDIN: the Marine Environment Data and Information Network, data archive centres (DAC). Available from: https://medin.org.uk/data-archive-centres.

Merriam–Webster (n.d.) Gradient. In: *Merriam-Webster.com dictionary*. URL https://www.merriam-webster.com/dictionary/gradient.

Nezlin, N.P. (2010). Oceanographic characteristics and spatial distribution of pelagic communities in the eastern part of the Bering Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, 57(17-18), 1639-1653.

Paillard, D (2001) Glacial cycles: Toward a new paradigm. *Reviews of Geophysics* 39(3): 325–346. doi:10.1029/2000RG000091.

Palmer-Rodríguez, P., Alberich, R., Reyes-Prieto, M. *et al.* (2025) Metadag: a web tool to generate and analyse metabolic networks. *BMC Bioinformatics* 26(1), 31.

Quinton, M., Elgebaly, G., Ding, D. *et al.* (2024). The European Nucleotide Archive in 2024. *Nucleic Acids Research*, 52(D1), D69-D75.

Richardson, L., Allen, B., Baldi, G. *et al.* (2023) MGnify: the microbiome sequence data

analysis resource in 2023. *Nucleic Acids Research* 51(D1), D753-D759.

Robinson, C.J., Morales, S.E., Mardis, E.R. *et al.* (2024). NIH and NCBI: data and services for the biomedical research community. *Nucleic Acids Research*, 52(D1), D3-D9.

Tufekci, V., Auster, P.J., Rhodehamel, J.L. and Valentine, P.C. (2010). Spatial analysis of benthic habitat complexity using multibeam bathymetry and Geographic Information Systems (GIS). *Marine Geodesy*, 33(sup1), 272-294.

Wilke, A., Sukumaran, J., Wang, M. *et al.* (2023). The IMG/M system for genome, metagenome, and metatranscriptome analysis. *Nucleic Acids Research*, 51(D1), D772-D783.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J. *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018.

Yakimov, M.M., La Spada, G. and La Cono, V. (2014) Microbial community of the deep-sea brine Lake Kryos seawater-brine interface is active below the chaotropicity limit of life as revealed by recovery of mRNA. *Environmental Microbiology* 17(2), 364-382.

Zaneveld, J.R., Chase, J., Weber, P.M. *et al.* (2018). Qiita: A qPCR-based approach to the human gut microbiome. *PLoS One*, 13(9), e0204732.

WORKSHOP COMMUNICATIONS

A) GEOLOGICAL AND BIOGEOCHEMICAL MODELS

Seabed Representation of the Seafloor Using EMODnet Bathymetry

Thierry Schmitt

European Marine Observation and Data Network (EMODnet), Service Hydrographique et Océanographique de la Marine (SHOM), France.

Introduction

Bathymetric information is a key variable for numerous scientific, economic and environmental usages. EMODnet Bathymetry collaborators are gathering data, mainly collected for hydrographic and/or marine geophysical purposes. Through a set of tools and methods the EMODnet consortium has normalized the metadata and data discovery processes alongside with their synthesis into a Digital Terrain Model and derived product. This contribution is aiming at describing the product and services that EMODnet Bathymetry provides to the community, with a special focus on the Mediterranean Sea.

Keywords: EMODnet bathymetry, seafloor mapping, Digital Terrain Model (DTM), bathymetric data, acoustic methods, marine ecosystems

Data and Metadata

Bathymetric data are mainly acquired using acoustic driven technologies. Multibeam echo-sounders (MBES) or single beam echo-sounders (SBES) are the most widely used technologies for both coastal and deep-sea environments. Horizontal positioning of the soundings is relying on the accurate positioning of the surveying ship which is equipped with a Global Navigation System along with Inertial Navigation System. While MBES are nowadays commonly used by the scientific and hydrogaphic communities, SBES is still in use and has been widely used prior to the nineties. For some areas, historic data from SBES or from leadline is sometime the only source of data available.

In the shallow environment, which ranges from the coast to a few tens of meters in water depth, acoustic methods remain possible, but they become less efficient and hazardous to put in practice (risk of grounding). In this environment, optical based methods can be an alternative. Airborne LiDAR is often used to collect both bathymetric and topographic data. Also, satellite (or airborne) mounted optical sensors using multi-spectral or hyper-spectral sensors are also an appropriate alternative for bathymetric estimation (Jawak *et al.* 2015).

In the deepest environment, where no measurement is available from conventional acoustic methods, the only valuable way to estimate bathymetric information is to estimate it from the inversion of the effect of the gravity field (depending on the underlying shape of the seafloor and its composition) on the free surface of the ocean. Relative height of the free surface is measured from altimeter equipped satellites such as GOCE, Jason, Altika, and possibly Sentinel 1. Such estimation has a horizontal precision of the order of few hundreds of meters to the km and a vertical precision of the order of few tenths to hundreds of meters. In the case of EMODnet, such information is extracted from the GEBCO grid, which in turns uses the estimated surface from SRTM15+ (Tozer *et al.* 2019).

Overall, as described above, multiple techniques are available to measure or estimate bathymetric information, yielding to multiple levels of quality/accuracy of the generated information. Moreover, as various communities are stakeholders in the field of data acquisition and processing, it can be easily understood that those communities may have different levels of expectations according to the use they will make of the bathymetric information. It is therefore of importance to describe in detail each dataset to be made available to the public, using dedicated harmonized metadata based on shared vocabularies. Two centralized catalogues are being used: survey centered (Common Data Index – CDI) or data product centered (Composite PRoDuct index – CPRD). Following the INSPIRE directive, and based on ISO19119/19135 standards, the main metadata elements are summarized in Table 1.

Variable →	Typical values/comments
ID →	A unique local identifier dataset at the distributing data centre to access metadata and dataset.
Data provider reference →	Originator or distributing Data Centre known as EDMO identifier
Instrument →	Depth: Single beam, multibeam, lidar, other. Position : Determines indirectly the accuracy of the positioning system
Creation of the dataset →	Date or year of the survey or gridding
Revision of the dataset →	Date of the year
Start Date →	Start date of the survey (single survey) or of the oldest survey used to produce the DTM
End date →	End date of the survey (single survey) or of the most recent survey used to produce the DTM
Bounding →	Bounding box, curve (track line) or surface (seafloor coverage) polygon
Horizontal CRS →	CRS (preferably the EPSG code
Vertical CRS →	Vertical datum used for depth (e.g. Mean sea level, Chart datum, Lowest astronomical tide, not applicable, ellipsoid)Section Break (Continuous)
Sampling → and → gridding method	Sampling and interpolation method used with processing parameters Section Break (Continuous)
Resolution →	Spatial resolution
Quality →	Quality indicator (see description further below)
Licence →	Usage licence (CC-BY preferred)

Table 1. Main metadata attributes.

At the end of 2025, the CDI catalogue held 45470 entries, while the CPRD catalogue had 301 entries, originating from 47 European datacenters, with datasets widespread in European waters and globally. Figure 1 shows a snapshot of the survey coverage (trackline and surfaces). It also shows that each geometry is associated with a harmonized metadata set of descriptors (accessible in XML).

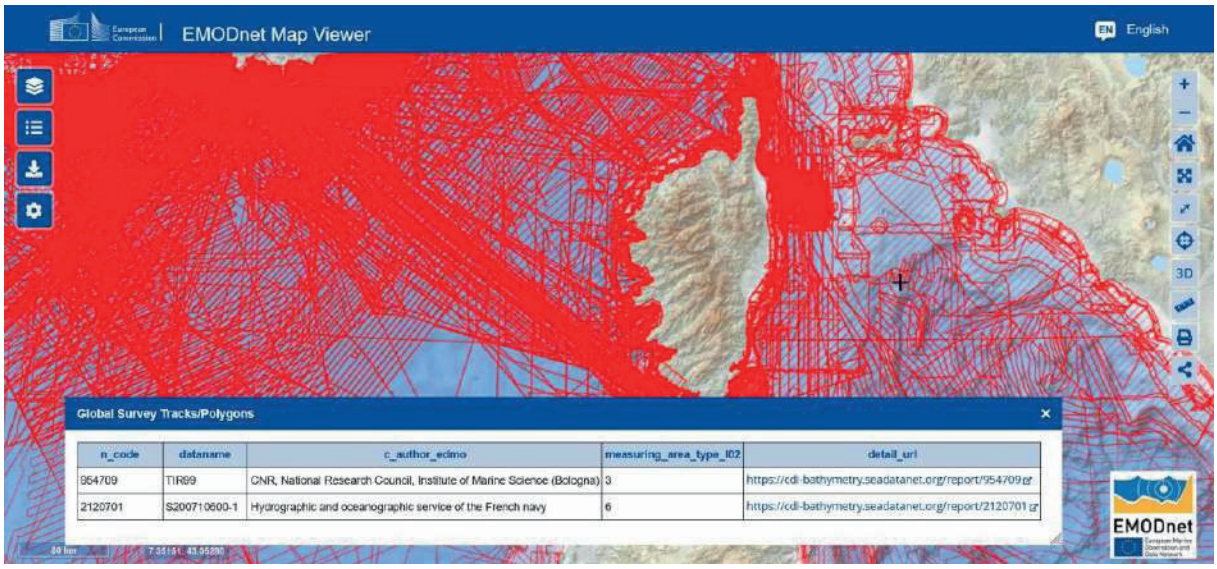


Figure 1. Non-exhaustive examples of potential sites of subsurface abiotic energy production in the Mediterranean area as suggested by onshore or offshore geology. Background from Google Maps.

Bathymetric Digital Elevation Model

As introduced earlier, EMODnet Bathymetry deals with heterogeneous bathymetric data. Moreover, bathymetric data are generally provided as point clouds with high variability in terms of spatial distribution and density (often related to the acquisition device in use). While users can be overwhelmed with such diversity, generating a continuous structured digital elevation model is often a valuable approach for an easier use of bathymetric information. For practical reasons and compatibility with tools provided by the GIS community, the data structure of the EMODnet DTM is a multi-layer raster grid, with statistically computed depth values centered in the middle of the grid node.

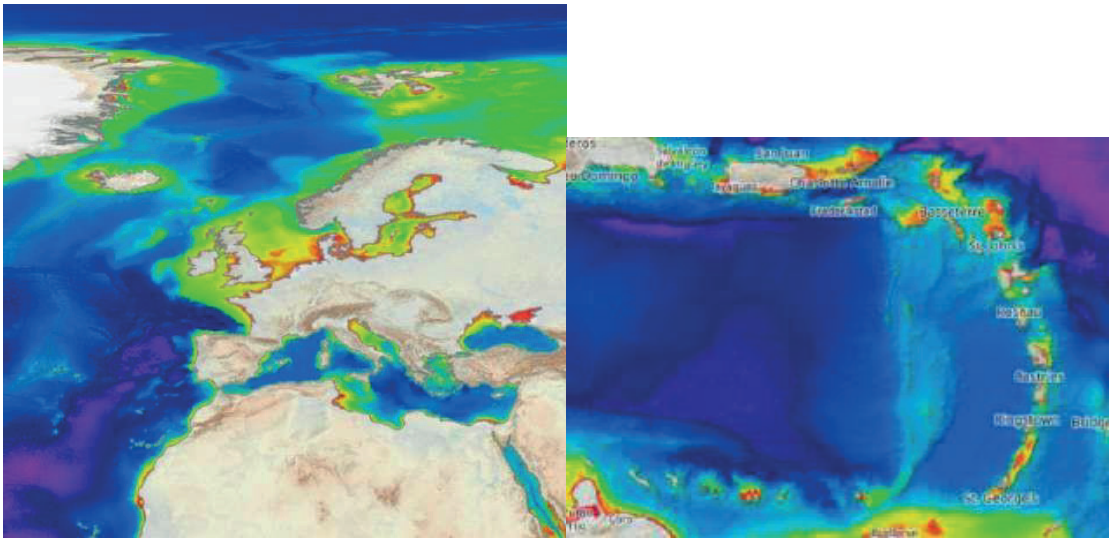


Figure 2. EMODnet Bathymetry coverage around Europe, also including the Caribbean area.

The full EMODnet Bathymetry processing chain is structured on 1) a harmonized sampling and bathymetric gridding method (harmonization of the origin, pyramiding resolution); 2) Selection tools, when datasets are overlapping, filtering/ setting priority of one dataset over another based on relevant metadata associated attributes 3) Complementary roles defined from the data holders to data integrator, through basin coordinators, with the objective of efficient quality assessment and manageable feedback loops. For this purpose, data are manipulated using the 3D viewing GLOBE software (Poncelet *et al.* 2020), developed for the EMODnet community and widely distributed.

As earlier mentioned, where no bathymetric data are available, the EMODnet bathymetric DTM benefits from GEBCO's bathymetric information extrapolated from the inversion of signal measured from altimetric sensors. Results of the collaboration between GEBCO and EMODnet, and the use of the remove-restore techniques (Weatherall *et al.* 2015), lead to the convergence of both bathymetric products for the European waters, hence minimizing the confusion for the users and smoothing the interface between the areas of pure bathymetric data and the ones issued from altimetric extrapolation, in both products. Note that despite the fact that the releases of GEBCO and EMODnet DTM products can't really be synchronized, each product uses the most recent and curated product from each other part.

Figure 2 describes the overall coverage of the EMODnet Bathymetry DTM. Note that since 2022, the Caribbean area has been added as an area of interest. This raster grid has a 1/16th of arcminute resolution (corresponding to 115 m grid resolution at the equator). Along with the average depth per grid node, the minimum and maximum depth, standard deviation, number of soundings per grid cell is also available. An important information which is also provided per grid node, is the identifier of the most prevalent data source contributor for the node. This last element allows you to refer directly to the corresponding metadata descriptor (either from the CDI or CPRD catalogues). This facility is used to generate quality layers, providing to the user confidence maps associated to the DTM (see Figure 3).

Conceptually the quality of a bathymetric dataset can be described using the main following parameters:

- The accuracy of the survey, which is a function of both the vertical and horizontal precision of the sounding measurement. This, in turn, can be broken apart in from a description of the vertical positioning system and the horizontal positioning system.
- The temporal representativity or in other terms the consistency between the measurement (at the time when it was acquired) and the actual morphology of the seabed (*e.g.* bathymetric measurements cannot be considered accurate if they were surveyed, say ten years ago in a highly dynamic area, such as for sand dunes for example)
- The completeness or the sampling of the seabed, which provides some forms of confidence in the sounding measurement (as of several soundings by unit of space). This is often related to the survey conditions (speed of the boat, overlapping between adjacent lines, sea states, …)

Four proxy indicators summarizing the above parameters are easily available from the knowledge of the data providers. These indicators are included as part of the metadata attributes (see above). They are known as "Quality Indicator":

- Horizontal accuracy (QI Horizontal)
- Vertical accuracy (QI Vertical)
- Purpose of the survey (QI Purpose)
- Age of the survey: the age of the survey is derived from the « start date » field of the metadata. (QI Age)

QI Horizontal	QI Vertical	QI Age	QI Purpose
-1: Multisources – unable to assess	-1: Multisources – unable to assess	-1: Unable to assess	-1: unable to assess
0: Unknown or >500m (grossly equivalent to TACAN, OMEGA system)	0: Unknown or plummet, leadline	0: >30 years	0: Historical survey with no associated information
1: Between 500 and 50m (Grossly equivalent to LORAN, DECCA)	1: SBES Low Frequency, SDB (similar than 2 + 5% water depth)	1: 10-30 years	1: Transit or opportunity survey
2: Between 50 and 20m (grossly equivalent to natural GPS)	2: MBES Low Frequency (similar than 1+2% water depth)	2: 5-10 years	2: Bathymetric / morphologic survey
3: ≤ 20 m (GPS with corrections – DGPS, RTK, …)	3: LiDAR, SBES high Frequency	3: 0-5 years	3: Hydrographic survey or survey compatible with recent international standards
	4: MBES High Frequency (higher than 100kHz) (similar than 1 + 0.5% water depth)		

Table 2. Description of the classes for each of the Quality Indicator.

Figure 3 illustrates the distribution of the "Age" Quality Indicator for the Mediterranean and Black Sea area. Note that the "undefined" value mostly relates to the altimetry derived bathymetry originating from the GEBCO contribution. The other quality indicators can be visualized the same way.

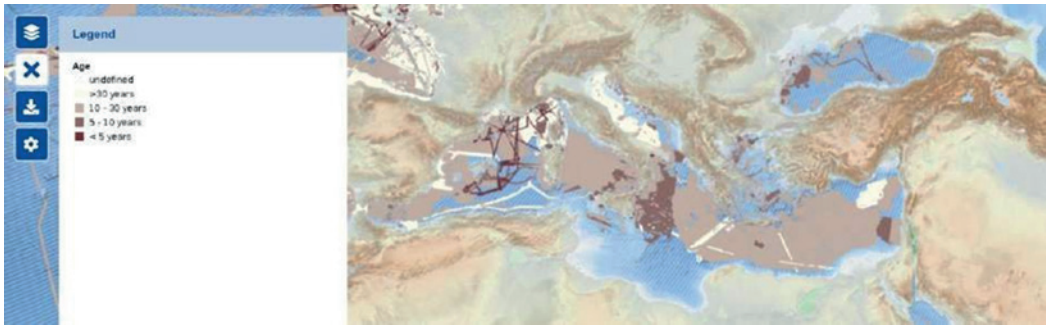


Figure 3. Display of the Quality indicator "Age" for the Mediterranean and Black Sea areas.

Web Services and Data Visualization

Users can access the DTM and the metadata either for viewing or for downloading trough the EMODnet central portal (https://emodnet.ec.europa.eu/en). Integration with other GIS platforms is also possible using Open Geospatial Consortium (OGC) web services, in particular:

Web Map Services (WMS), Web Feature Services (WFS), Web Coverage Services (WCS), and the Catalogue Service for the Web (CSW). Moreover, these are all supported by the underlying metadata standards such as ISO- 19115 and ISO-19139 described earlier, and well-known raster grid format (netcdf, ESRI Ascii, geotiff, *etc.*).

Layer	Web service endpoint	Content
Bathymetry Global Base layer	https://tiles.emodnet-bathymetry.eu/wmts/1.0.0/WMTSCapabilities.xml	pre-rendered georeferenced map tiles with a fixed geographic extent for different zoom levels.
Bathymetry WMS layer:	https://ows.emodnet-bathymetry.eu/wms?SERVICE=WMS&REQUEST=GetCapabilities&VERSION=1.3.0	Pre-rendered geo-registered map image.
Bathymetry WFS layer:	https://ows.emodnet-bathymetry.eu/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0	Coverage polygon and associated attributes.
Bathymetry WCS layer:	https://ows.emodnet-bathymetry.eu/wcs?SERVICE=WCS&REQUEST=GetCapabilities&VERSION=2.0.1	Web Coverage Services (WCS) to support requests for coverage data (rasters) extraction
CDI Data Discovery & Access service:	https://geo-service.maris.nl/emodnet_bathymetry/wms?SERVICE=WMS&REQUEST=GetCapabilities&VERSION=1.3.0	Web service for marine bathymetric datasets used from the Common Data Index (CDI) Data Discovery and Access service
CPRD Data Discovery & Access service	https://sextant.ifremer.fr/geonetwork/EMODNET_HYDROGRAPHY_PRODUCT_TILES/eng/csw?SERVICE=CSW&REQUEST=GetCapabilities	Web services for marine bathymetric datasets used from the CPRD catalogue
Full Bathymetric gridded product - ERDDAP	https://erddap.emodnet.eu/erddap/griddap/bathymetry_2022.html	Gridding extraction ERDDAP facility
REST API services	https://rest.emodnet-bathymetry.eu/	REST facility in use for single point or profile extraction.

Table 3. EMODnet Bathymetry Web services endpoints.

Note that recent efforts have been brought to further ease machine-to-machine processes, notably using tools, such as Environmental Research Division's Data Access Program (ERDDAP), that allows scientists to work in their format of choice and make the resulting data available through interoperable formats, such as NetCDF, geotiff, ASCII …, without an added burden on the user side. The following table provides endpoints to the corresponding EMODnet Bathymetry services. An associated documentation is provided here https://emodnet.ec.europa.eu/en/emodnet-web-service- documentation.

Recognizing that 3D display is particularly adapted and useful in the geophysical/oceanographic realm, efforts have been made to enable a web-based 3D rendering. In order to provide such a rendering in adequate timing, a method has been developed to create tiled Level Of Details (LOD) filtering using surface simplification out of large-scale regularly gridded terrain data. Campos *et al.* (2020) reviewed and updated simplification methods that are well known from the state of the art, to efficiently work at the tile level (greedy insertion, edge collapse simplification, and point set simplification). The developed framework has served to obtain

a qualitative and quantitative comparison between several simplification methods applied to the problem of terrain visualization and validated on a European- scale dataset (see Figure 4). From the results obtained, we can conclude that the greedy insertion method obtains the best results in terms of processing time and mean error, with respect to the original data. While the main advantage of the proposed solution is to be able to process each tile independently, the main drawback resides in the fact that the order in which the tiles in a given LOD are processed has an impact on the results. For the full EMODnet Bathymetry area, processing a large- scale terrain can take several days of processing even if we parallelize the creation of each tile.

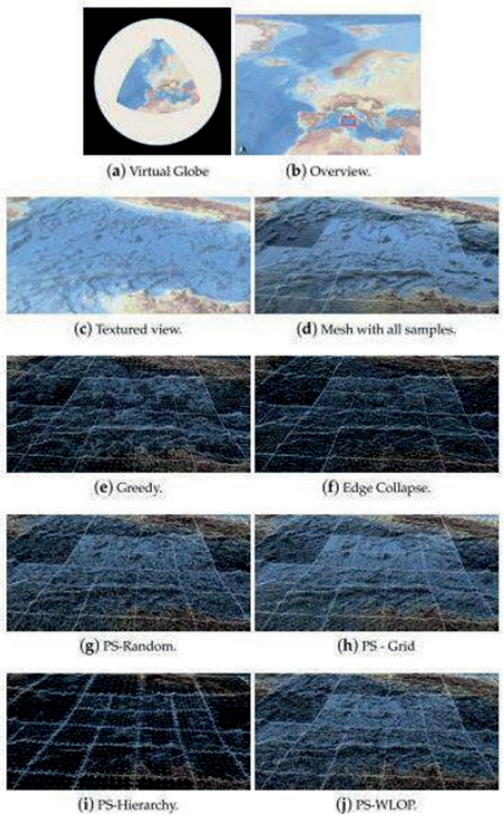


Figure 4. Examples of simplification methods used in the benefit of fast and efficient 3D rendering (image extracted from Campos *et al.* 2020).

Conclusive Remarks

The EMODnet Bathymetry consortium builds up on the gathering of a large community of experts with knowledge in bathymetric data acquisition, processing and rendering. Thanks to this community, EMODnet bathymetry provides a unique inventory of bathymetric data describing the European waters. Thanks to the harmonization of both metadata and data sampling mechanisms catalogues of existing data (CDI and CPRD) are available to the public along with the 1/16 arc minute resolution regional DTM. These enable any user to fulfil most of his usage either by a direct use of the DTM or by requesting source data directly from the provider. Web services allow automating machine to machine data transfer.

References

Campos, R., Quintana, J., Garcia, R., *et al*. (2020). 3D simplification methods and large-scale terrain tiling. *Remote Sensing*, 12 (3): 1–24.

Jawak, S.D., Vadlamani, S.S., Luis, A.J. (2015). A Synoptic Review on Deriving Bathymetry Information Using Remote Sensing Technologies: Models, Methods and Comparisons. *Advances in Remote Sensing*, 4 (2): 147–162.

Poncelet, C., Billant, G., Corre, M.P. (2020). GLOBE (Global Oceanographic Bathymetry Explorer) Software. www.seanoe.org.

Tozer, B., Sandwell, D.T., Smith, W.H.F., *et al*. (2019). Global Bathymetry and Topography at 15 Arc Sec: SRTM15+. *Earth and Space Science*, 6 (10): 1847–1864.

Weatherall, P., *et al*. (2015). A new digital bathymetric model of the world's oceans. *Earth and Space Science*, 2 (8): 331–345.

Mediterranean Subsurface Abiotic Energy

Alberto Vitale Brovarone

Department of Biological, Geological, and Environmental Sciences, Alma Mater Studiorum, Università di Bologna, Bologna, Italy.

Abstract

Life in the subsurface requires energy sources capable of sustaining chemosynthesis. Among these energy sources, H_2 and abiotic CH_4 are known to play a fundamental role in the sustainment of biological communities living at extreme conditions at seafloor hydrothermal vents, or microbial communities inhabiting the subsurface biosphere. Geological processes capable of producing these abiotic energy sources are identified in several settings, from mid-ocean ridges to convergent plate margins con continental basement rocks. Several of these settings –or analogous conditions– are present in the Mediterranean area. However, their identification, spatial distribution, and potential energy fluxes are challenging owing to the lack of an integrated geological and biological dataset for Mediterranean marine environments. The combination of geological and biological data into a single mapping and visualization resource may profoundly enhance our knowledge on the topic.

Keywords: Abiotic energy sources, serpentinization, chemosynthesis, Mediterranean marine environments, subsurface biosphere, geological processes, data integration

Main

Over the past two decades, an increasing body of scientific literature has documented the geologic production of natural energy forms such as molecular hydrogen and light hydrocarbons through abiotic processes (Figure 1) (Etiope and Sherwood Lollar 2013; Kelley, 2005; Sherwood Lollar *et al.* 2014; Truche *et al.* 2024). A large fraction of these processes is mediated by fluid-rock interactions taking place in the Earth's crust and mantle. Submarine environments host one of the most striking processes, called serpentinization (Cannat *et al.* 2010; Kelley 2005; Merdith *et al.*, 2020). It is the process whereby mineral olivine –the most abundant mineral in the lithosphere– reacts with the most abundant geologic fluid –water– to produce rocks called serpentinites. During serpentinization, the byproducts of serpentinization are magnetite—a magnetic mineral—and molecular hydrogen (H_2H)—a powerful energy source. As carbon is omnipresent in geologic fluids and seawaters, H_2-carbon interactions lead to the formation of another key serpentinization byproduct, abiotic methane (CH_4) (Abrajano *et al.* 1990; Klein *et al.* 2019; McCollom 2013). Molecular hydrogen represents a key natural resource for the energy transition: its combustion produces water, with no carbon emissions. Abiotic CH_4 also is a powerful energy source like other forms of light hydrocarbons but, compared to H_2, represents an unquantified form of greenhouse emission with high global warming potential. Besides potential energetic applications for modern society, natural H_2 and abiotic CH_4 also have an important role in biology as they may serve as energy source for chemosynthetic life in the subsurface biosphere (Brazelton *et al.* 2006; Kelley 2005). For this reason, the role of serpentinization and related H_2 and abiotic CH_4 has deserved increasing attention within the geo-biological scientific community.

Serpentinization, however, is not the only geological process capable of producing H^2 and abiotic CH_4 (Figure 1). As an example, Precambrian crystalline basements have been shown to represent promising sources of radiolysis mediated H_2 (Sherwood Lollar *et al.* 2014). Although the inventory of potential settings for $H_2$and abiotic CH_4 has substantially increased over the last two decades, recent discoveries of H_2 reservoirs in variable crustal settings suggest that H_2 (and related abiotic CH_4) production and accumulation may be much more common and widespread than previously known (Frery *et al.* 2021; Truche *et al.* 2024).

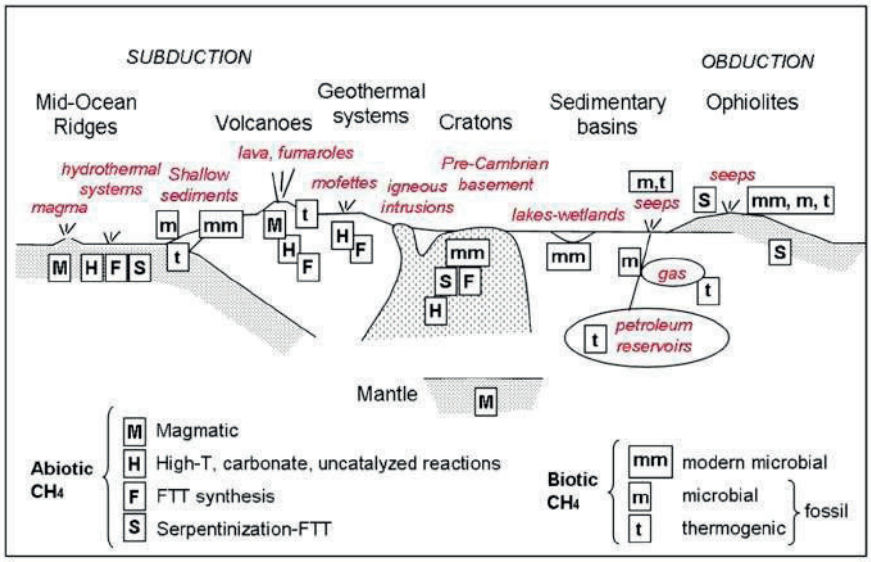


Figure 1. Sketch showing the range and geologic settings of abiotic CH_4 production. From Etiope and Sherwood-Lollar (2013).

The Mediterranean area hosts a large variety of marine geologic settings, from convergent to divergent, volcanic, and sedimentary. This results in an equally large variety of interconnected geological and biological processes producing or converting energy sources, from purely biotic to purely biogenic endmembers. The identification of potential areas of deep energy production is mandatory in order to expand our knowledge on the distribution of natural energy resources in the Mediterranean area and their impact on Mediterranean geo-biodiversity.

Both coastal and submarine manifestation of such natural energy production or key driving conditions (rocks, fluid/rock interactions…) are documented in the Mediterranean area (Figure 2). Examples include the Chimaeara Mount, Turkey (Figure 3) (Etiope 2023) the Acquasanta alkaline springs (Boschetti *et al.* 2013), Ligurian Alps, Italy, the Calabrian arc subduction in the Ionian sea (Polonia *et al.* 2017) (Figure 4), the Monte Maggiore massif in the Tyrrhenian sea, the Beni Bousera massif in the Alborean sea (Varas-Reus *et al.* 2016), ODP site 651 (Bonatti *et al.* 1990) and many, many others. These are just a few examples related to the process of serpentinization only.

Figure 2. Potential sites for subsurface abiotic energy production in the Mediterranean, identified through onshore and offshore geological data. (Background: Google Maps).

Figure 3. Example of underground natural energy production: Natural H_2 and abiotic CH_4 emissions from serpentinites at Chimaera Mount, Turkey (Etiope 2023). Coastline in the background. Photo credit: Martin Siepmann.

The cross-disciplinary characterization of geologic processes producing natural H_2 and abiotic CH_4 like serpentinization from divergent to convergent plate margins and onland has substantially increased over the last decades. This effort, managing and combining several types of petrologic, geochemical and modeling datasets, has provided valuable insights on the potential for both active and inactive specific geologic systems to produce and release abiotic energy sources, from single natural samples to global flux estimates.

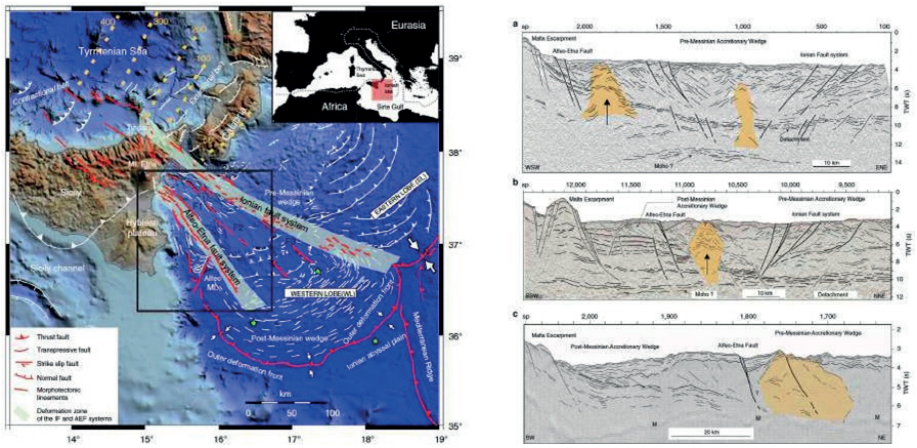


Figure 4. Serpentinite diapirs in the Calabrian arc region. These structures may represent preferential pathways for the upward migration of deep energy sources across the subsurface biosphere.
From Polonia *et al.* (2017).

For example, previous and current work by my research group have identified H_2 and abiotic CH_4 trapped in rocks in multiple localities worldwide. These gases are trapped inside small inclusions in rocks and can be released in response to tectonic or alteration processes.
The characterization of the host rocks and minerals and the included gases – through extraction and analytical protocols developed in-house– have provided unprecedented data on the abundance and chemical fingerprints of purely abiotic compounds.
The knowledge gain my research group is targeting requires the integration of data from petrologic, geochemical, and geophysical investigation from onland settings, and the definition of chemical fingerprints and flux estimates at various settings, from single geologic bodies to the global scale in the present and in the past (Figure 5).

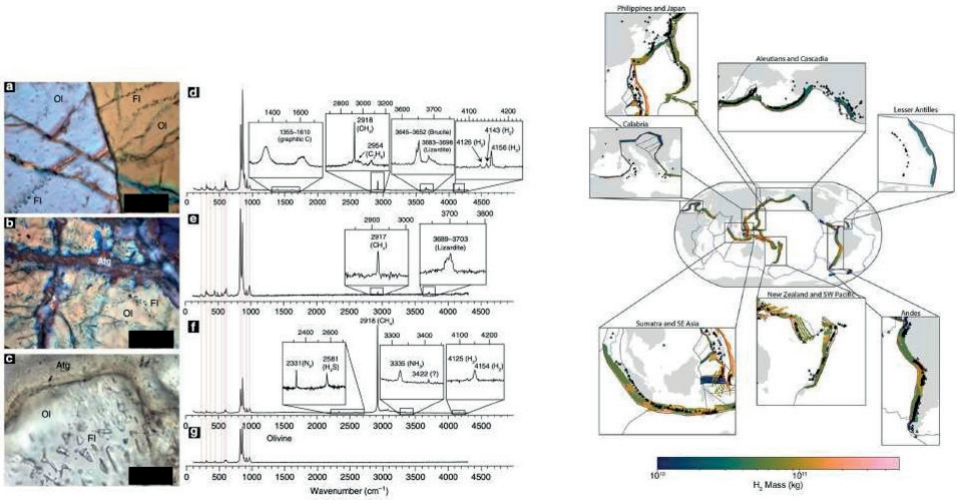


Figure 5. Left: Natural H_2 and abiotic CH_4 trapped within ultramafic rocks from the Mediterranean. Raman spectra confirm the presence of H_2 and CH_4 in small fluid inclusions (Vitale Brovarone *et al.* 2020). Right: Global estimates of H_2 production via serpentinization at convergent margins (Merdith *et al.* 2023). This process likely occurs along the Mediterranean's convergent plate margins.

The case study of Monte Maggiore massif, Corsica, France, represent a nice example of onland H_2 and abiotc CH_4 source extending offshore. We have managed datasets spanning field observations, petrography and petrology, geochemistry (from bulk to in-situ major, minor, trace elements and stable isotopes), geophysics (magnetism and gravimetry), and modeling to derive fluxes of H_2 and abiotic methane in the past and in the future for a single massif.

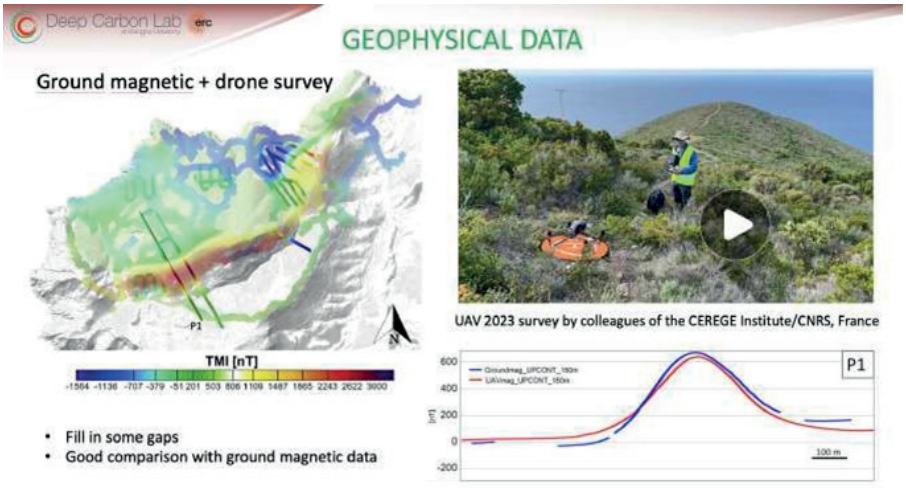


Figure 6. Unpublished magnetic geophysical data from the Monte Maggiore massif, Corsica, France. Left: Ground magnetic data, where signal intensity indicates the degree of serpentinization and, consequently, the cumulative flux of natural H_2 and abiotic CH_4 from the massif. Right: Field image showing drone magnetic data acquisition. Data property of Dr Zeudia Pastore (University of Bologna, NTNU, CEREGE-CNRS).

This knowledge can be extended to marine environments provided that a consistent database on direct (drill cores, *etc.*) and indirect (geophysical data) Mediterranean marine geology. Although large uncertainty exists on the interpretation of seismic data and a large-scale mapping approach may be affected by major biases, the integration of a broad range of direct and indirect (seismic, magnetic, gravimetric…) geological and biological data into a single mapping and visualization resource may provide an unprecedented means to assess the distribution and fluxes of natural abiotic energy source in the Mediterranean area and their control on the distribution of chemosynthetic life in the subsurface and on the distribution of abiotic vs. abiotic processes.

Acknowledgements

This work is part of the project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program to Vitale Brovarone (grant 864045, acronym DeepSeep). A MUR FARE (Italian Ministry of University Framework for Attracting and Strengthening Research Excellence in Italy; grant R20ZJYMPAR, acronym DRYNK) grant and a MUR PRIN2022 (grant 20224YR3AZ; acronym HYDECARB) to Vitale Brovarone are also acknowledged.

References

Abrajano, T.A., Sturchio, N.C., Bohlke, J.K. *et al.* (1990). Methane-hydrogen gas seeps in the Philippines: Evidence for abiotic origin. Chemical Geology, 85(1-2), 89-94.

Bonatti, E., Seyler, M., Channell, J., *et al*. (1990). Peridotites drilled from the Tyrrhenian Sea, ODP LEG 107. In: *Proceedings of the Ocean Drilling Program, Scientific Results*, pp. 37–47.

Boschetti, T., Etiope, G., Toscani, L. (2013). Abiotic Methane in the Hyperalkaline Springs of Genova, Italy. *Procedia Earth and Planetary Science*, 7: 248–251.

Brazelton, W.J., Schrenk, M.O., Kelley, D.S., *et al*. (2006). Methane- and Sulfur-Metabolizing Microbial Communities Dominate the Lost City Hydrothermal Field Ecosystem. *Applied and Environmental Microbiology*, 72 (9): 6257–6270.

Cannat, M., Fontaine, F., Escartín, J. (2010). Serpentinization and associated hydrogen and methane fluxes at slow spreading ridges. In: *American Geophysical Union Monograph Series*, Vol. 191, American Geophysical Union: Washington, D.C., pp. 241–264.

Etiope, G. (2023). Massive release of natural hydrogen from a geological seep (Chimaera, Turkey): Gas advection as a proxy of subsurface gas migration and pressurised accumulations. *International Journal of Hydrogen Energy*, 48 (17): 9172–9184.

Etiope, G. and Sherwood Lollar, B. (2013). Abiotic methane on Earth. *Reviews of Geophysics*, 51 (2): 276–299.

Frery, E., Langhi, L., Moretti, I. (2021). Natural hydrogen seeps identified in the North Perth Basin, Western Australia. *International Journal of Hydrogen Energy*, 46 (65): 31158–31173.

Kelley, D.S. (2005). A Serpentinite-Hosted Ecosystem: The Lost City Hydrothermal Field. *Science*, 307 (5714): 1428–1434.

Klein, F., Grozeva, N.G., Seewald, J.S. (2019). Abiotic methane synthesis and serpentinization in olivine-hosted fluid inclusions. *Proceedings of the National Academy of Sciences*, 116 (35): 17666–17672.

McCollom, T.M. (2013). Laboratory Simulations of Abiotic Hydrocarbon Formation in Earth's Deep Subsurface. *Reviews in Mineralogy and Geochemistry*, 75 (1): 467–494.

Merdith, A.S., Daniel, I., Sverjensky, D., *et al*. (2023). Global Hydrogen Production During High-Pressure Serpentinization of Subducting Slabs. *Geochemistry, Geophysics, Geosystems*, 24 (9): e2023GC010947.

Merdith, A.S., Real, P.G., Daniel, I., *et al*. (2020). Pulsated Global Hydrogen and Methane Flux at Mid-Ocean Ridges Driven by Pangea Breakup. *Geochemistry, Geophysics, Geosystems*, 21 (3): e2019GC008869.

Polonia, A., Torelli, L., Gasperini, L., *et al*. (2017). Lower plate serpentinite diapirism in the Calabrian Arc subduction complex. *Nature Communications*, 8 (1): 2172.

Sherwood Lollar, B., Onstott, T.C., Lacrampe-Couloume, G., *et al*. (2014). The contribution of the Precambrian continental lithosphere to global H2 production. *Nature*, 516 (7530): 379–382.

Truche, L., Donzé, F.V., Goskolli, E., *et al*. (2024). A deep reservoir for hydrogen drives intense degassing in the Bulqizë ophiolite. *Science*, 383 (6683): 618–621.

Varas-Reus, M.I., Garrido, C.J., Marchesi, C., *et al*. (2016). Refertilization Processes in the Subcontinental Lithospheric Mantle: the Record of the Beni Bousera Orogenic Peridotite (Rif Belt, Northern Morocco). *Journal of Petrology*, 57 (11): 2251–2270.

Vitale Brovarone, A., Sverjensky, D.A., Piccoli, F., *et al*. (2020). Subduction hides high-pressure sources of energy that may feed the deep subsurface biosphere. *Nature Communications*, 11 (1): 1–11.

Nutrients, Trace Element in Western and Eastern Mediterranean Sea Surface Sediment: Environmental Variability and Anthropogenic Footprint

Noureddine Zaaboub[1], Micha Rijkenberg [2], Lamia Trabelsi[1], Loes Gerringa[2], Monia El Bour[1]

[1]National Institute of Marine Sciences and Technologies, Marine science laboratory, Tunisia
[2]Royal Netherlands Institute for Sea Research, NIOZ, Netherlands

Abstract

During the GEOTRACES Med Black Sea cruise, our work focused on measuring physicochemical parameters, collecting water samples, and taking short sediment cores. Trace element analysis requires sampling with a trace metal-clean CTD system. Our analysis assesses nutrients, trace elements (Fe, Pb, Cd, Zn, Co, Mo, Cu, and Ni), and trace element fractionation, carried out on surface sediment from both eastern and western areas. X-ray diffraction was applied to the clay fraction. Salinity section plots reveal clearly recognizable water masses. Clay mineral assemblages originate from distinctive sources, and their dispersal reflects different transport agents in the eastern Mediterranean Sea. Nutrients indicate more oligotrophic conditions in the eastern area. The most important sources of dissolved silicate in the Mediterranean Sea originate from continental fluvial systems and groundwater discharges. Electron microscopy revealed the dominance of diatoms, which play an important role in organic matter export to the deep sea. Trace element fractionation differentiated five fractions: the first four constitute the bioavailable fractions, which were compared to deep water (near the sediment-water interface). This comparison primarily highlights the importance of surface sediment as a potential pump for trace elements into the water column. It also demonstrates the significant influence of continental discharges on trace element accumulation in both surface sediment and deep water, especially for Fe, Cu, and Co.

Keywords: bioavailability. Mediterranean Sea, nutrient, surface sediment, trace éléments

Introduction

Marine biogeochemical cycling refers to the distribution and concentration of nutrients and bio-essential elements, which are controlled by their uptake by phytoplankton in surface waters, the sinking and remineralization of organic remains in deeper waters, and subsequent redistribution by thermohaline circulation (De Baar *et al.* 2018). The Mediterranean Sea is an important global water body and may be considered a small-scale ocean system where several spatial and temporal processes occur (El-Geziry and Bryden 2010). Specifically, the deep waters within the Mediterranean basin may be classified as: Aegean Deep Water and Adriatic Deep Water, which form the Eastern Mediterranean Deep Water (EMDW); and Tyrrhenian Deep Water and Gulf of Lions Deep Water, which form the Western Mediterranean Deep Water (WMDW) (Bergamasco and Malanotte-Rizzoli 2010). The Mediterranean Sea is an oligotrophic sea because of its general water circulation (Schroeder *et al.* 2010).

Regarding nutrients, the demand for silicon by primary producers is not as great as it is for nitrogen or phosphorus. Nevertheless, the silicon cycle has acquired significant importance due to its role in marine primary production, particularly for organisms like diatoms which play an

important role in organic matter export to the deep sea and require silicon, phosphorus, and nitrogen for their growth (Buesseler 1998). Trace elements in marine systems are primarily attributed to continental margins, which receive natural and anthropogenic trace elements (TEs) from direct atmospheric deposition of aerosols onto the sea surface and from advection of riverine suspended particles and/or resuspended sediments from the continental shelf/slope (Cossa *et al.* 2014).
In specific areas, increased concentrations of trace elements are observed with increasing depth, which prompted an investigation into the interaction processes at the solid-liquid interface between volcanic ash and seawater throughout the water column. The transfer of trace elements from marine sediment to the water column has in most previous works been considered limited. Given high anthropogenic activity, and as explained by Béthoux *et al.* (1990), while the deep ocean's response time to perturbations is on the order of 1000 years, the Mediterranean's response to environmental disturbances is perceptible within two decades. Comparison of surface sediment metal bioavailability with deep water metal concentrations allows for an estimation of the possible increase in concentrations of dissolvable bio-essential elements (Fe, Ni, Co, Cd, Cu, Zn, Pb) originating from sediment as dissolved bioavailable fractions.

Material and Methods

Samples were collected aboard the Dutch R/V Pelagia along the GEOTRACES-A04N section (Figure 1), extending from the Northeast Atlantic (14.2°W, 39.7°N) to the Marmara Sea (27.5°E, 40.8°N), during cruises (64PE370) and (64PE374) in 2013. All samples were collected using a TITAN conductivity-temperature-depth (CTD) system, equipped with 24 ultra-clean 24 L PRISTINE sampling bottles made of polyvinylidene (PVDF) and titanium (Rijkenberg *et al.*, 2015). After deployment, the TITAN system was moved to a Class 100 container for subsampling (De Baar *et al.* 2008). Here, all 100 samples were collected using inline filtration under N_2 pressure (0.7 atm, with 99.99% filtered N_2).

The surface sediment core samples were sliced under N_2 gas during the Med Black cruise; the profiles did not exceed 22 cm in depth and focused on processes within the first centimeters. Trace element fractionation was carried out on these surface sediment samples. For trace elements, ICP-MS and ICP-OES were used for analysis, depending on whether they were solution or sediment samples.

Chemical fractionation was carried out via sequential extraction according to Tessier *et al.* (1979), with the extraction of five fractions as follows:

- (I) exchangeable fraction
- (II) carbonate-bound fraction
- (III) reducible or Fe/Mn oxide-bound fraction
- (IV) organic matter-bound and/or sulfide fraction
- (V) residual fraction

C, H, and N analysis in sediment was carried out using LECO elemental analysis. For sulfur in sediment, an elemental SC632 (S, C) analyzer was used.

Results and Discussion

Biogeochemical processes at the sediment-water interface have resulted in a number of environmental issues, including the enrichment of nutrients and chemicals, which affect their bioavailability. Most marine deposits originate from land-based sources, primarily continental alteration products, as well as industrial, agricultural, and urban waste.

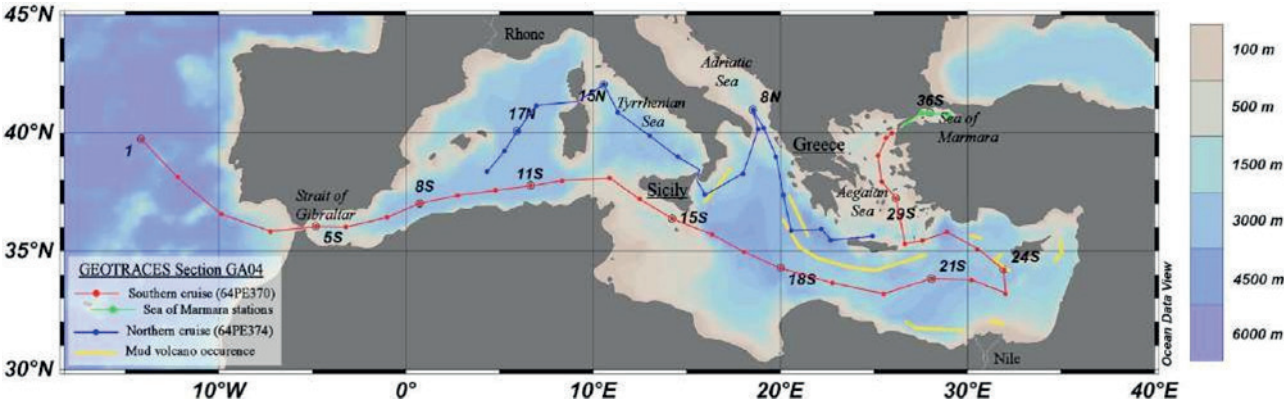


Figure 1. Sampling sites in the Mediterranean Sea, Red color southern Leg (S), blue Core northern Leg (L).

In EMS the unusual nutrient ratio (Figure 2) is due to high N:P values in all the external nutrient inputs to the Eastern Mediterranean Sea (EMS), coupled to low denitrification rates within the ultra-oligotrophic basin. P limitation in the EMS is due to regionally high rates of diazotrophic N_2 fixation (Van Cappellen *et al.* 2014).

Nitrate distribution shows a distinct biogeochemical difference between the Western Mediterranean Sea (WMS) and Eastern Mediterranean Sea (EMS) (Powley *et al.*, 2017). The main source of P and N to the Mediterranean Sea is the inflow of surface water from the Atlantic Ocean via the Strait of Gibraltar, rather than land-derived sources *ibid*. The nutrient budget demonstrates that N removal from the Eastern Mediterranean Sea (EMS) is balanced by N inputs from rivers and atmospheric deposition, without the need to invoke additional N_2 fixation (Figure 3).

Silica budgets in the Southern Mediterranean Sea (SMS), particularly in the eastern area, include the dissolution of external biogenic silica and diagenetic clay weathering as important sources of silicate to the basin (Krom *et al.* 2014). Another potentially important process to consider in this budget is silica cycling in coastal areas (Figure 4). To further improve the budget, it is necessary to better quantify the annual biogenic-Si in the surface inflowing waters at the Straits of Sicily, as well as contributions from rivers and from sediments in both offshore and coastal regions.

Mineralogy of surface sediment shows 34% carbonates and approximately 60% phyllosilicates. Based on the abundances and distributional patterns, different clay mineral assemblages are distinguished. Clay mineral assemblages have distinctive sources, and their dispersal reflects different transport agents in the eastern Mediterranean Sea.

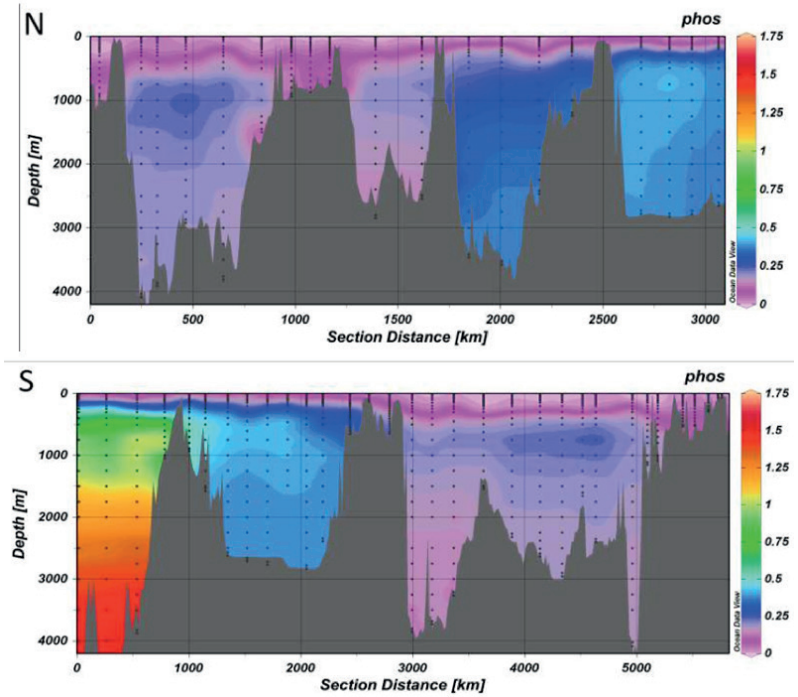


Figure 2. Phosphate (µmol/L) in southern Leg (S) and northern Leg (N) in the Mediterranean Sea.

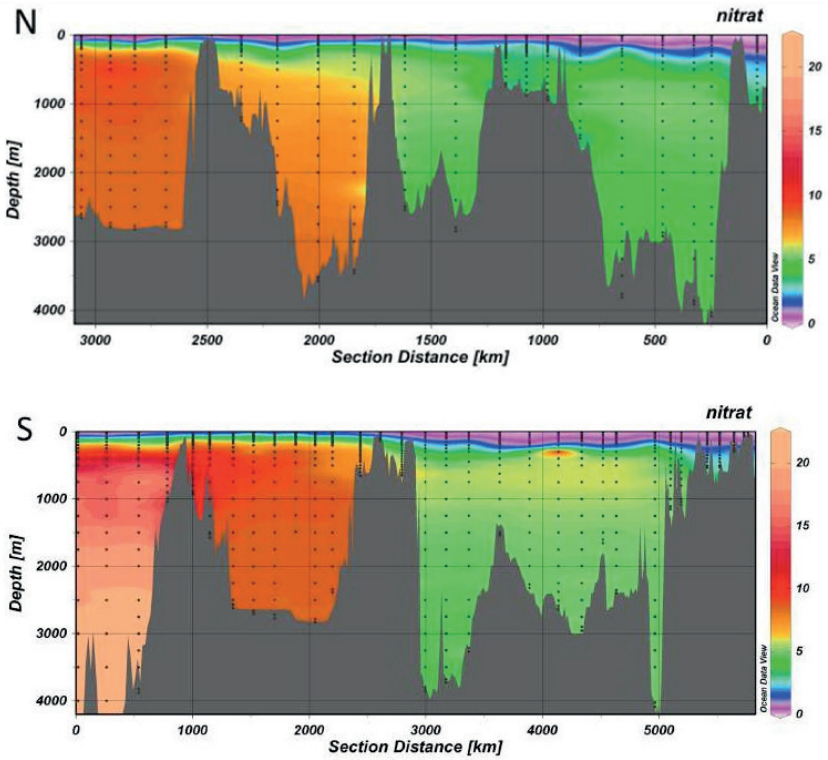


Figure 3. Nitrate (µmol/L) in southern Leg (S) and northern Leg (N) in the Mediterranean Sea.

Within the first few centimeters of the deep sediment layers, there is a gradual disappearance of smectite, indicating diagenetic processes (Figure 5, S3 and S19). These assemblages have distinctive sources, and their dispersal reflects different agents of transport in the eastern Mediterranean Sea. A Nile assemblage with smectite (> 15%) and 15–25% kaolinite is found on the eastern Nile cone and within the eastern Levantine Basin. The distribution of chlorite in the eastern Mediterranean Sea has resulted from its dispersal by the eastward-directed surface water currents, which form part of the counter-clockwise gyre in the eastern Mediterranean Sea. The Levantine Intermediate Water has transported an assemblage characteristic of the southeast Aegean Sea, characterized by 40–60% well-crystallized smectite and higher contents of chlorite and illite than in the Nile assemblage, onto the Mediterranean Ridge southeast of Crete (Figure 5, S19 and S31).

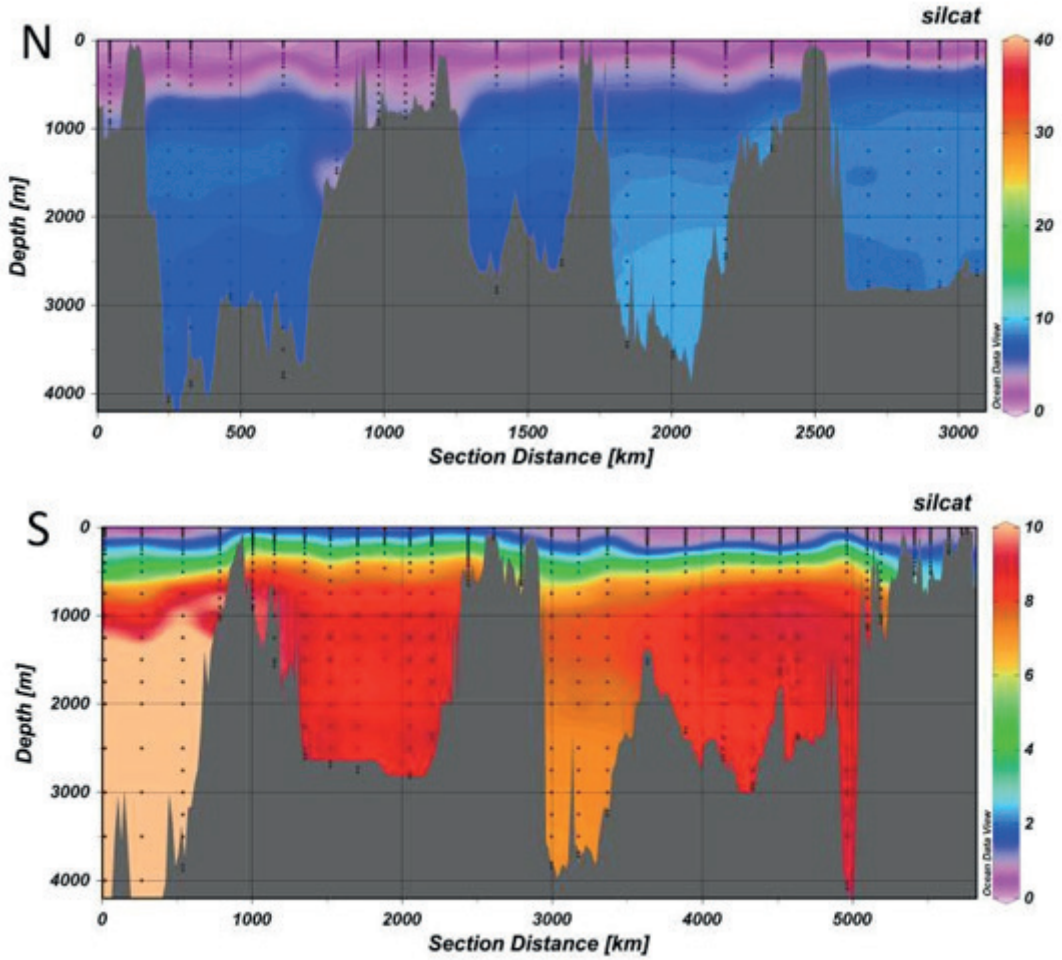


Figure 4. Silicate (µmol/L) in southern Leg (S) and northern Leg (N) in the Mediterranean Sea.

A kaolinite-rich assemblage (20%–30% kaolinite), coinciding with high carbonate values, occurs on the western section of the Mediterranean ridge and to some extent in the western Nile cone, because of transport by wind and rivers from North Africa (Figure 5, S19).

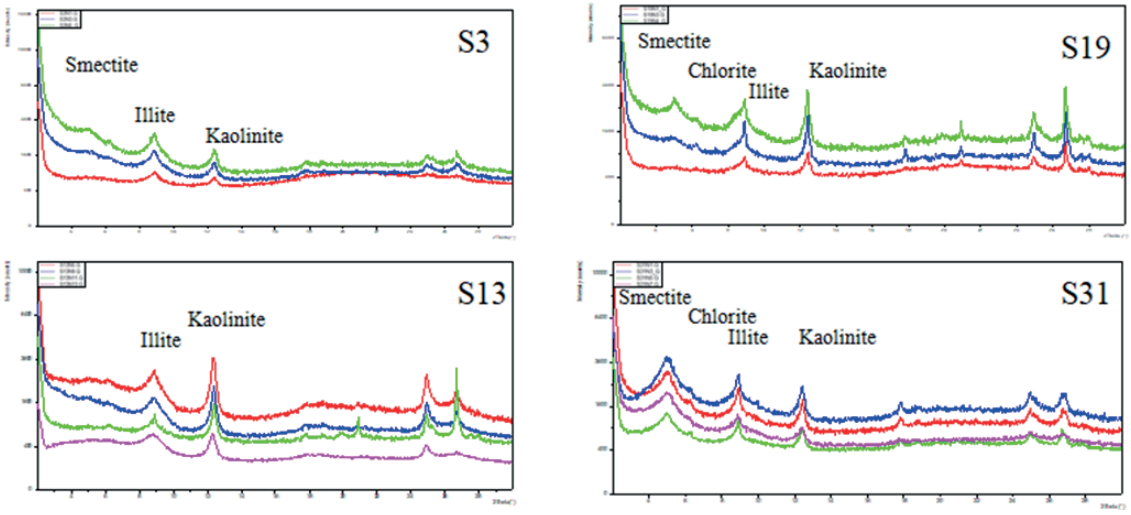


Figure 5. X-ray diffraction of the clay fraction (less than 2µm in size) in the surface core sediments reveals different assemblages in the Mediterranean Sea.

The restriction of Kithira and Messina assemblages (illite- and chlorite-rich) to deeper parts of the Ionian Basin is chiefly due to water movements involved in deep circulation. A Sicilian assemblage, with 30–50% kaolinite and 15–20% smectite, found in the westernmost part of the Ionian Basin south of Sicily, has mainly resulted from dispersal by eastward-moving surface waters from the Western Mediterranean Sea.

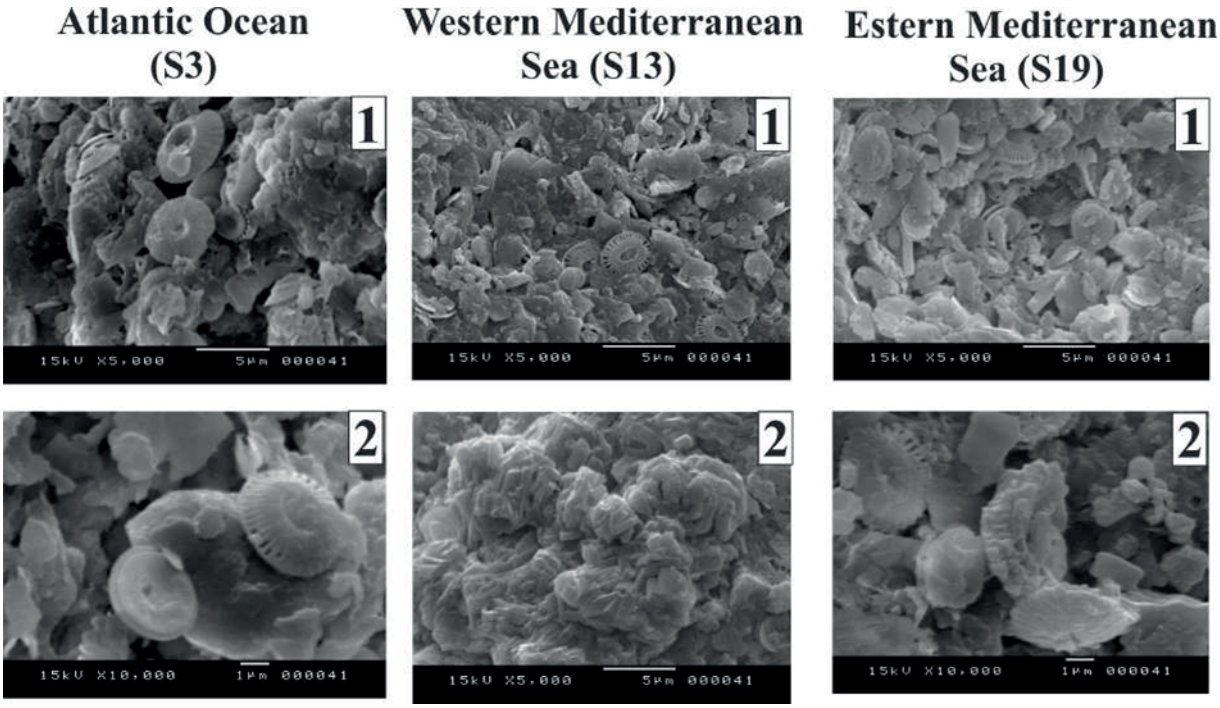


Figure 6. X-ray diffraction of the clay fraction (less than 2µm in size) in the surface core sediments reveals different assemblages in the Mediterranean Sea.

Mineral assemblage is the main controlling process for diagenetic reactions at the sediment-water interface. Modelling can illustrate the geological relationships between marine water and sedimentary basins, typically generating authigenic clays. Kaolinite morphology and distribution are identified as flux- or diffusion-controlled.

Silicates are abundant in surface sediments, as observed by electron microscopy (Figure 6). Diatoms, which are the most abundant taxonomic group of phytoplankton found in the surface sediment matrix, play an important role in organic matter export from the water column to the deep sea, requiring silicon, nitrogen, and phosphorus for their development (Buesseler 1998). The present results highlighted active aerobic and anaerobic heterotrophic bacteria in oligotrophic conditions across different deep-sea areas of the Mediterranean Sea (El Bour *et al*. 2021).

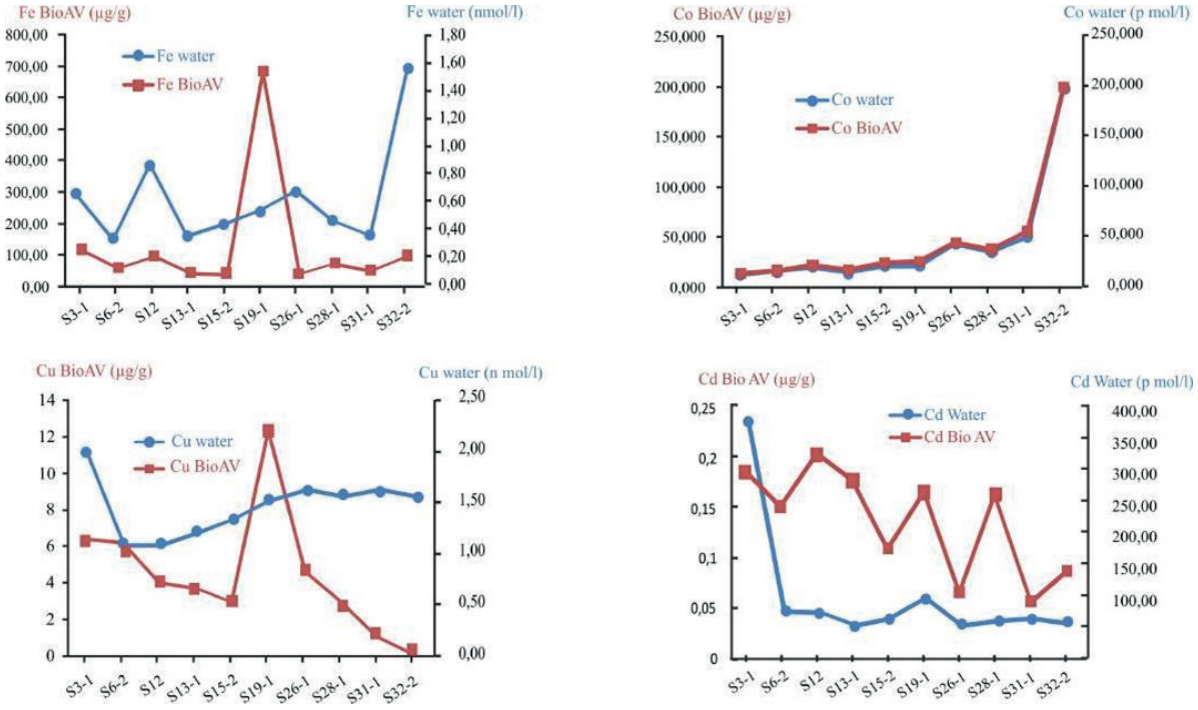


Figure 6. trace element in deep water and surface sediment bioavailable fraction in Mediterranean Sea.

Distributions of trace elements (Cu, Ni, Pb, Zn, Co, Mo, and Fe) in the Southern Mediterranean Sea's deep water near the sediment-water interface, when compared to the bioavailable fractions (Phases I, II, III, IV) from surface sediment, reveal a strong similarity in distribution. This highlights distinct footprints of local patches where high concentrations of bioavailable trace elements, such as Fe and Cu near the Nile Canyon, are well developed (Figure 7).

The surface morphology confirms their terrigenous origin in association with clay minerals, whereas Cd is more associated with organic matter (OM). All these trace elements are anthropogenically enriched from dust input and from sediments in submarine canyons (Cossa *et al.* 2014). Anthropogenic influence remains clearly discernible.

Dissolved Co concentrations decreased closer to the sediment-water interface; the bioavailable Co fraction (Co BioAV) typically mirrors that of deep water, which we attribute to sediment resuspension (Figure 7) that increases Co exchange. This exchange can then be utilized by

organisms as an important element for phytoplankton growth, productivity, and diversity (Dulaquais *et al.*, 2017).

The bioavailable fraction of Fe shows local patches (Figure 7). The occurrence of discrete zones in deep waters characterized by elevated dissolved iron levels and bioavailable ones could be elucidated through a combination of specific physical mechanisms and localized sources. The boundaries of these high bioavailable Fe patches in deep waters suggest lateral movement, facilitated by phenomena like mesoscale eddies originating from deep-sea sources. These sources are likely diverse and could include mud volcanoes, continental inputs, and deep-sea mountains.

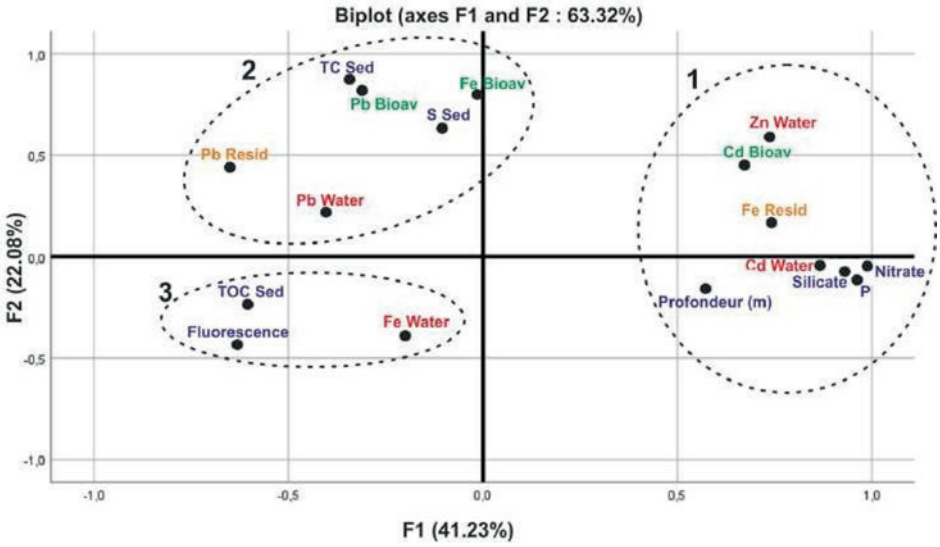


Figure 8. The PCA biplot of geochemical parameters at water sediment interface in the Mediterranean Sea.

The PCA biplot (Figure 8) effectively illustrates the relationships among physicochemical variables, depth, fluorescence, and trace element concentrations (Fe, Zn, Cd, and Pb) in the water column. It also details the bioavailable and residual fractions of these trace elements within the sediment. Our observations reveal a strong correlation between fluorescence, total organic compound levels in sediment, and waterborne iron concentrations. Furthermore, these findings suggest that the concentration of iron in water is a crucial factor influencing the growth and development of certain marine organisms, such as corals, algae, and various fish species.
The analysis also indicates correlations between bioavailable concentrations of iron, Pb, and Cd, and total carbon. This can be attributed to deep water circulation, primarily driven by the dispersal of westerly moving surface water from the Western Mediterranean Sea. The availability of these trace elements plays a significant role in controlling diagenetic reactions at the sediment-water interface.

The observed relatively elevated concentrations of dissolved trace elements in some Mediterranean deep waters. Sediment resuspension during episodes of deep water formation appears to be the most likely source of these additional dissolved trace elements from the bioavailable fraction (Figure 4). Intense sediment resuspension has been extensively documented during periods when dense shelf water cascades through submarine canyons carved into continental shelves at multiple locations across the Mediterranean Sea (*e.g.,* Puig *et al.* 2014).

Conclusion

Nutrient distribution throughout the Eastern and Western Mediterranean Sea is profoundly influenced by Atlantic water flowing through the Straits of Gibraltar. Mineralogical distribution reveals distinct assemblages originating from various sources, with their dispersal reflecting different transport agents, primarily in the Eastern Mediterranean Sea. Furthermore, current effects near the Strait of Gibraltar and the Strait of Sicily significantly impact local biogeochemical processes.

The Mediterranean Sea exhibits substantial exchanges between deep water trace element concentrations (particularly near the sediment-water interface) and the bioavailable fraction in surface sediments. The presence of distinct patches of elevated trace element concentrations in deep waters can only be explained by a combination of specific physical processes and localized sources at particular depths, such as those observed in front of the Nile Canyon.

References

Berganasco, G. and Malanotte-Rizzoli, P. (2010). Numerical simulations of the circulation in the North Atlantic. Journal of Physical Oceanography, 40(7), 1547-1563.

Béthoux, J.P., Courau, P., Nicolas, E., *et al*. (1990). Trace metal pollution in the Mediterranean Sea. *Oceanologica Acta*, 13 (4): 481–488.

Buesseler, K.O. (1998). The decoupling of production and particulate export in the surface ocean. *Global Biogeochemical Cycles*, 12 (2): 297–310.

Cossa, D., Buscail, R., Puig, P., *et al*. (2014). Origin and accumulation of trace elements in sediments of the Northwestern Mediterranean margin. *Chemical Geology*, 380 (1-2): 61–73.

De Baar, H.J.W., van Heuven, S.M.A.C., Middag, R. (2018). Ocean Biochemical Cycling and Trace Elements. In: White, W.M. (Ed.), *Encyclopedia of Geochemistry*, Springer International Publishing AG: Cham, pp. 1023–1048.

De Baar, H.J.W., Timmermans, K.R., Laan, P., *et al*. (2008). Titan: A new facility for ultraclean sampling of trace elements and isotopes in the deep oceans in the international Geotraces program. *Marine Chemistry*, 111 (1): 4–21.

Dulaquais, G., Planquette, H., L'Helguen, S., *et al*. (2017). The biogeochemistry of cobalt in the Mediterranean Sea. *Global Biogeochemical Cycles*, 31 (7): 377–399.

El Bour, M., Rijkenberg, M., Saadi, A., *et al*. (2021). Characterization of Deep-Sea Sediment Microbial Communities from Different Mediterranean Sea Regions. In: Ksibi, M., Ghorbal, A. and Chakraborty, S. (Eds.), *Recent Advances in Environmental Science from the Euro-Mediterranean and Surrounding Regions (2nd Edition)*, Springer: Cham, pp. 3–5.

El-Geziry, M. and Bryden, I.G. (2010). The circulation pattern in the Mediterranean Sea: issues for modeller consideration. *Journal of Operational Oceanography*, 3 (2): 39–46.

Krom, M.D., Kress, N., Fanning, K. (2014). Silica cycling in the ultra-oligotrophic eastern Mediterranean Sea. *Biogeosciences*, 11 (15): 4211–4223.

Powley, H.R., Krom, M.D., Van Cappellen, P. (2017). Understanding the unique biogeochemistry of the Mediterranean Sea: Insights from a coupled phosphorus and nitrogen model. *Global Biogeochemical Cycles*, 31 (7): 1010–1031.

Puig, P., Palanques, A., Martín, J. (2014). Contemporary sediment-transport processes in submarine canyons. *Annual Review of Marine Science*, 6 (1): 53–77.

Rijkenberg, M.J.A., Middag, R., van Haren, H., *et al*. (2015). "PRISTINE", a new high-volume sampler for ultraclean sampling of trace metals and isotopes. *Marine Chemistry*, 177: 501–509.

Schroeder, K., Gasparini, G.P., Borghini, M., *et al*. (2010). Biogeochemical tracers and fluxes in the Western Mediterranean Sea, spring 2005. *Journal of Marine Systems*, 80 (1-2): 8–24.

Tessier, A., Campbell, P.G.C., Bisson, M. (1979). Sequential extraction procedure for the speciation of particulate trace elements. *Analytical Chemistry*, 51 (6): 844–851.

Van Cappellen, P., Powley, H.R., Emeis, K.C., *et al*. (2014). A biogeochemical model for phosphorus and nitrogen cycling in the Eastern Mediterranean Sea: Part 1. Model development, initialization and sensitivity. *Journal of Marine Systems*, 139: 460–471.

WORKSHOP COMMUNICATIONS

B) GEOMICROBIOLOGY

Beyond Biogeochemistry: Exploring Microbial Interactions with the Solid Earth in Marine Ecosystems

Donato Giovannelli

Department of Biology, University of Naples Federico II, Naples, Italy

Institute for Marine Biological Resources and Biotechnologies, Italian National Research Council, CNR-IRBIM, Ancona, Italy

Earth-Life Science Institute, ELSI, Tokyo Institute of Technology, Tokyo, Japan

Marine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA

Department of Marine and Coastal Science, Rutgers University, New Brunswick, NJ, USA

Abstract

Microorganisms are key players in global biogeochemical cycles essentially constituting the "bio" in global biogeochemistry. While their importance in the global functioning of the ocean and our planet is undisputed, the extent to which they contribute to planetary scale recycling of nutrients and elements is less constrained, especially when considering processes that lay at the interface with the solid Earth. In the ocean and subsurface ocean ecosystems the interaction between microorganisms and geological processes likely control key parameters connected to global volatile cycling and the long-term stability of our climate. While classic studies of microbial diversity drivers have been extremely useful to understand the controls structuring microbial ecosystems, a large portion of the diversity variance remains unexplained and functional redundancy seems to be a major feature of microbial communities. Interactions with geochemistry and geology at changing scales can provide an additional layer through which we can interpret microbial diversity. Shedding light into the complex nature of the interactions between the marine biosphere and the solid earth will require a strong interdisciplinary approach challenging the established discipline boundaries and changing the way we approach the study of ocean functioning at diverse scales.

Keywords: biogeochemistry, microbial diversity, geosphere-biosphere interactions, ecosystem functioning

Introduction

Microorganisms are key players in global biogeochemical cycles (Falkowski *et al.* 2008) and effectively control a large portion of the biological recycling of elements and nutrients on a global scale, essentially constituting the "bio" in biogeochemistry (Giovannelli 2023). In marine ecosystems, microorganisms, which include prokaryotes (Bacteria and Archaea), protist and unicellular algae, control a large swath of key ecosystem functions (Jørgensen 2000; Arrigo 2005; Madsen 2011). For example, they produce ~ 50 % of the oxygen released annually in the atmosphere, are fundamental in the transformation of di-nitrogen gas into forms of nitrogen usable by other organisms and are key players in the remineralization of organic matter. They are able to influence the (geo)chemical composition of entire ecosystems, and are for example key drivers of anoxic events, eutrophication, oxygen minimum zones, and complex feedback

dynamics in a number of other ecosystems (Ulloa *et al.* 2012; Pacton and Gorin 2014; Berg *et al.* 2022). They are also fundamental in primary productivity, not only in the photic zones as cyanobacteria, phytoplankton (Harris 2012; Sigman and Hain 2012; Winder and Sommer 2012) and the less studied anoxygenic phototrophs (Koblížek 2015), but also in the deep sea contributing to primary productivity in the dark water column, such as with the ammonia oxidising archaea and bacteria (Herndl and Reinthaler 2013; Ricci and Greening 2024), and in a number of other chemolithotrophic supported ecosystems, such as deep-sea hydrothermal vents, cold seeps, deep-sea hypersaline anoxic basins and subsurface sediments (Hügler and Sievert 2011).

Especially below the photic zones, microorganisms play a disproportionately big role in controlling biogeochemistry, mainly due to a reduced influence of primary productivity by multicellular eukaryotes, terrestrial runoff and reduced or absent influence of the light from our star (Ricci and Greening 2024). As we reach closer to the seafloor, deep into the sediments and oceanic crusts and in areas of increased geological activity the interaction of life, especially microbial life that in these areas dominates the biomass standing stock, with the geosphere is heightened (Magnabosco *et al.* 2019; Giovannelli *et al.* 2022). These biomes represent a conspicuous portion of our biosphere, consider for example the average depth of the oceans and the fact that they cover 70 % of our planet's surface, and have the potential to significantly influence the solid Earth and large-scale geological processes. For example, primary productivity in the water column and sediments influences, together with microbial organic matter degradation, the amount of organic carbon that is recycled back into subduction zones (Plank and Manning 2019).

Subduction zones are key areas of interactions between Earth surface and deep interior and effectively connect the biological and geological side of biogeochemistry in ways that are less conspicuous compared to other planetary processes (Bekaert *et al.* 2021). At subduction zones as one oceanic plate is recycled beneath a continent or another oceanic plates volatiles (*i.e.* water, carbon, nitrogen, sulfur and a myriad of other elements and molecules) are recycled into the mantle where they undergo transformation and mobilization before being recycled, often with a lag of a few to hundred million of year, to the surface or being sequestered in the depths of our planets for hundreds of millions, potentially billions of years (Kelemen and Manning 2015; Bekaert *et al.* 2021). Subduction zones are responsible, together with mid ocean ridges, for the movement of the plates that characterizes our planet tectonic, they control the assembly of continents (*i.e.,* Wilson cycle) and create several key oceanic features such as trenches, volcanic arcs and smaller features such as mud volcanoes and cold seeps on active margins, influencing oceanic and continental topography.

At subduction zones microorganisms directly or indirectly control some of the most important parameters, such as the sedimentary content of organic matter, the amount of carbonate going into subduction and the sedimentary content of pyrite, just to mention a few (Plank and Manning 2019; Li *et al.* 2020). These parameters are fundamental in controlling the redox state of the generated magmas, the volatile content of the primitive magma, the type of volcanism, and the efficiency of the subduction zone (Evans 2012; Galvez and Pubellier 2019). All these have in turn strong feedback on the functioning of our planet and directly influence biology, both on continents and in the ocean (Giovannelli *et al.* 2020, 2021; Vitale Brovarone *et al.* 2020).

This complex geosphere-biosphere interaction is a great example of complex feedbacks between our solid earth and the biosphere that have shaped the evolutionary trajectory of our planet (Giovannelli 2023). While a comprehensive list of possible geosphere-biosphere feedbacks is lacking in the literature, including examples of key co-evolutionary events at the interface

between geo and bio, it is possible to appreciate how widespread these types of feedback interactions are (Jelen *et al.* 2016; Colman *et al.* 2017; Moore *et al.* 2017; Giovannelli, 2023). Direct investigations of the feedback mechanisms between geosphere and biosphere have been studied for decades at deep-sea hydrothermal vents (Jannasch and Mottl 1985; Baquiran *et al.* 2016; McNichol *et al.* 2018) and more recently in shallow-water hydrothermal vents (Price *et al.* 2007; Giovannelli *et al.* 2013a; Barosa *et al.* 2023), although at a local scale. In these ecosystems the biology is tightly coupled to the geological and geochemical regimes created by the architecture of mid ocean ridges and volcanic areas (Price and Giovannelli 2017; Dick 2019). The activity of chemolithotrophic microorganisms in deep-sea hydrothermal vent subsurface chimneys and hydrothermal plumes plays a huge role in contributing to the mobility of colloidal and nano particulate metals that constitute one of the major sources of biologically essential trace metals to the global ocean, including photic zones (Yücel *et al.* 2011).

A more recent example of studied geo-bio interaction is the role of subsurface microorganisms in controlling the quality and quantity of volatile recycled to the surface in the forearc region of subduction zones (Giovannelli *et al.* 2020; Fullerton *et al.* 2021). In these systems low temperature geochemical reactions and microbial metabolism have been shown to sequester massive amounts of carbon in the crust, preventing its release to the atmosphere and reducing up to 40 % the carbon that was believed to be sequestered into the deep mantle (Barry *et al.* 2019; Fullerton *et al.* 2021; Rogers *et al.* 2022). While these constitute great examples, the extent to which the marine ecosystem is influenced and can influence the solid earth remains largely unknown (Giovannelli *et al.* 2022; Cloetingh *et al.* 2023). A major portion of marine microbial diversity is associated with marine sediment and the marine subsurface and remains largely underexplored at best or completely unknown.

Given the importance of the oceans for the functioning of our planet it is fundamental to understand how and to what extent the marine ecosystem is influenced and can influence geosphere processes, both to understand planetary evolution and to better forecast the complex feedback that controls the response to the ongoing climate crisis. Shedding light into the complex nature of the interactions between the marine biosphere and the solid earth will require a strong interdisciplinary approach challenging the established discipline boundaries and changing the way we approach the study of ocean functioning at diverse scales.

Microbial Communities and Volatile and Element Cycling in Marine Ecosystems

The role that microbes play in controlling volatile cycling has been extensively discussed at the planetary scale and in certain marine extreme environments, such as deep-sea hydrothermal vents and cold seeps (Falkowski *et al.* 2008; Taupp *et al.* 2010; Dick 2019). However, the larger role they play in the marine ecosystems when thinking about the interactions with the geosphere has been investigated to a lesser detail. This is despite the many key roles that microbial communities play.

Starting from the photic zone, microorganisms are key agents of primary productivity, both through oxygenic and anoxygenic photosynthesis in all ecosystems where light constitutes the primary energy source. As primary producers they do not only contribute to the sustaining of higher trophic level and marine food webs, but they also play a key role in sequestering carbon, nitrogen, and many other elements from the atmosphere and surface oceans contributing to their deposition into marine sediments (Azam and Malfatti 2007). This effectively redirects several biogeochemical cycles from dissolved or gas molecules to a solid, and therefore sinking,

state of matter that is better integrated into the rock cycle (Plank and Manning 2019). Primary productivity is not limited to the photic zone, and is widespread along the deep water column, in marine sediments and in many marine extreme environments where chemolithotrophy contributes to a large portion of the carbon then transferred to higher trophic levels. In this context, the link with the geosphere is even tighter, as the electron donor and acceptors, and in many cases the carbon itself, used by chemolithotrophic microbes from primary productivity is directly of geological origin and they are often released as results of water-rock interactions in the oceanic crust and underlying structures, such as subduction zones (Hügler and Sievert 2011; Sievert and Vetriani 2012; Yücel 2013). The different reduced compounds produced by water-rock interactions amount to alternative forms of energy to the sunlight provided by the solid earth that sustain extremely diverse microbial communities (Vitale Brovarone *et al.* 2020).

Aside from primary productivity marine microbes play key roles in a number of other element cycles. They are fundamental in the marine nitrogen cycle, contributing uniquely to nitrogen fixation in diverse marine ecosystems, to denitrification, anammox, ammonia oxidation and all the intermediate steps of nitrogen cycling which have huge impacts on climate (Ward et *al.* 2007). In marine sediments they are major players in the remineralization of organic carbon and in the consumption of sulphate, through sulphate reduction and the conversion of organics in hydrogen (Jørgensen *et al.* 2019), which is quickly recycled *in situ*, and methane, a major contributor to the global carbon cycle and long term climate engine (Valentine 2011; Whiticar 2020). They are themselves also involved in the oxidation of large amounts of methane and hydrogen, including those produced abiotically and thermogenically in hydrothermal sediments and deeper into the crust and mantle through serpentinization (Knittel and Boetius 2009; Vitale Brovarone *et al.* 2020; Greening and Grinter 2022; Lappan *et al.* 2023).

Beyond the classic CHNOPS elements, we need to consider the key role that microbes play in the cycling of trace elements, many of which are fundamental for all life on this planet and are key constituents of our technological societies (Giovannelli 2023; Hay Mele *et al.* 2023).

Microorganisms, and especially extremophiles, are in fact capable of interacting with a large swath of the periodic table, well beyond the few elements classically included in the list of elements of biological interest. For example, a large number of metals are directly used by life as cofactors to control several fundamental metabolic reactions including the pervasive redox chemistry which constitute the essential way by which Life interacts with energy (Hay Mele *et al.* 2023). The different metals play diverse roles in microbial metabolic reactions, and different steps in each biogeochemical cycle require a diverse set of metals to be completed. Additionally, these metals (*e.g.*, Fe, Mo, W, Zn, Cu, V, Mn, Mg, Ni and Co) are also key building blocks in minerals and rock and are therefore central in the interaction of life with the geosphere. Additionally, their availability and distribution have dramatically changed in the course of the evolution of the planet (Anbar 2008), suggesting possible strong geosphere-biosphere feedback at the crossway between trace metal availability and metabolic evolution (Giovannelli *et al.* 2023).

Changing Scales

Elucidating how marine microbial life and the solid earth interact requires a paradigm shift in the way we approach the study of microbial ecology. Classic approaches are typically focused on small scale variation of microbial diversity across transects of a few meters to a few hundred meters or associated with the comparison of contrasting conditions (Xie *et al.* 2011; Giovannelli *et al.* 2013a; Pasotti *et al.* 2014; Yakimov *et al.* 2015; Merlino *et al.* 2018). Studies investigating the variability of microbial communities at larger scales are less common and typically do not

connect with geological features. Large scale sampling across physical features is not a new approach, and it has been successfully applied by biological oceanographers working in close collaborations with physical and chemical oceanographers (Sunagawa *et al.* 2015; Tagliabue *et al.* 2017; Raes *et al.* 2018). Beyond the domain of the water column, and especially the surface oceans, interactions among disciplines are reduced, to the extent for example that research cruises to deep sea hydrothermal vents rarely are designed for extensive synoptic sampling of vents across the geological and biological domains.

Even when large scales are considered, marine microbial diversity is traditionally linked with changes in macro physico-chemical variables (like depth, pH, temperature and salinity), the availability of trophic resources (organic carbon, nutrients), or community interactions (predation, competition, cooperation) to explain the observed structuring of marine microbial communities.

For example, a global survey of 243 samples from 68 locations sampled in the epipelagic and mesopelagic waters across the globe carried off by the TARA Oceans consortia demonstrated that temperature and dissolved oxygen are the primary drivers of both taxonomic and functional diversity (Sunagawa *et al.* 2015). In the mesopelagic and bathypelagic waters, the Malaspina expedition showed that free-living or particle-attached lifestyles significatively drive functional differences in bathypelagic prokaryotic communities (Acinas *et al.* 2021), linking microbial diversity to trophic resources independently from their biogeography. Similarly, in marine sediments, microbial diversity has been primarily linked to the strong role of trophic resource availability (Giovannelli *et al.* 2013b) and to sediment lithology due to their different organic matter content (Hoshino *et al.* 2020). In extreme environments, such as deep-sea and shallow-water hydrothermal vents, changes in microbial diversity are primarily linked to the thermal and geochemical gradients (Giovannelli *et al.* 2013a; Price *et al.* 2013; Ren *et al.* 2018). Inter-domain interactions are instead identified as drivers of the bacterioplankton diversity in the Ross Sea surface waters (Cordone *et al.* 2022).

These approaches have helped tremendously to advance our understanding of the structuring of marine microbial communities; however, a large fraction of the observed variance remains unexplained and a lot of what we consider ecological redundancy escapes our current models. A deeper connection with the geological processes might bring new life to the study of marine ecosystem diversity and functions (Giovannelli *et al.* 2022). For example, consider shallow-water hydrothermal vents sampled across a thermal gradient or between vents located in diverse areas of the same volcanic system (Giovannelli *et al.* 2013a; Barosa *et al.* 2023). While it is well known, and by now considered textbook knowledge that the temperature, pH and geochemical regime will control microbial diversity, the reason these parameters are changing is often not directly brought into the discussion regarding biodiversity. For example, changes in the geochemical regime might be driven by diversity of the primary fluids interacting with the host rocks or changes in the trace elements composition of the hosting rock themself; changes in temperatures might be controlled simply by mixing, or reveal differences in the thermal structure of the volcanic region investigated; and changes in pH and volatile inventory can reveal geological processes happening at depth.

While all this might seem inconsequential for microbial diversity, a recent series of papers by diverse groups suggests instead that delving into the geological processes might provide a new and important lens through which we can interpret microbial diversity and functions. Examples include tight interactions with subduction parameters (Basili *et al.* 2022; Upin *et al.* 2023), host rock diversity (Savaglia *et al.* 2024), carbon sequestration (Barry *et al.* 2019; Rogers *et al.* 2022) and structural volcanisms (Fullerton *et al.* 2021). All these examples currently come

from terrestrial ecosystems, but it is easy to see how this approach could be applied to marine environments, especially on the seafloor and subsurface ecosystems. This can only be achieved with a new level of interactions among disciplines, that requires developing specific tools to integrate data across academic and cultural boundaries.

For example, a recent preprint has demonstrated that integrating metagenomic data with geological characteristics of the host rocks and clumped isotope data for the gases it is possible to track the origin and fate of methane across a 500 km segment of the Central American Volcanic Arc (Figure 1). This could reveal controls on the diversity of microbes and their role in the biogeochemistry of methane, that could not be appreciated unless diverse disciplines are integrated across large spatial scales (Selci *et al.* 2024).

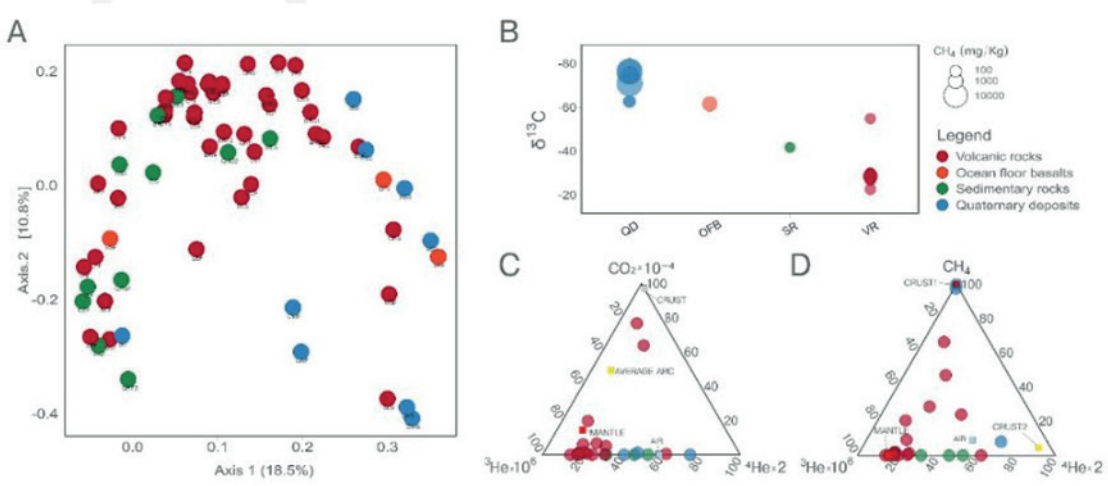


Figure 1. Example of integration of data across disciplines: microbial functional methane cycling gene diversity across the Central American Volcanic Zone subsurface. (A) Functional based Principal Coordinates Analysis (PCoA) based on weighted Jaccard similarity of the shotgun metagenome functional read assignment, subsetted for methane cycling; (B) δ^{13}C-CH4 distribution across different geologic units (QD: quaternary deposits; OFB: oceanic floor basalts; SR: sedimentary rocks; VR: volcanic rocks) with the size proportional to the CH_4 concentration (mg/kg); (C) Ternary plot of CO_2, ^{3}He, and ^{4}He and (D) CH^4, ^{3}He, and ^{4}He relative compositions. Reproduced from Selci *et al.* (2024).

Especially when moving from the water column to the sediments and the oceanic crusts, there are several scales at which biology and geology intersect. Recent studies have shown that individual microbes in the subsurface undergo total carbon turnover times spanning months to hundreds of years, meaning that an individual can survive for even longer, possibly operating on timescales beyond thousands of years (Lloyd 2021). Any interaction these microbes have with the surrounding fluids and solid mineral phases at depth should then be considered to occur over these long timescales over which the microbes operate (Giovannelli *et al.* 2022). Processes like microbial metabolic transformations of volatiles, the diagenesis of organic matter and the microbially-mediated dissolution and precipitation of minerals, might act at temporal and spatial scales that overlap with geological processes in the same area (Figure 2), influencing large scale volatile cycling in deep time.

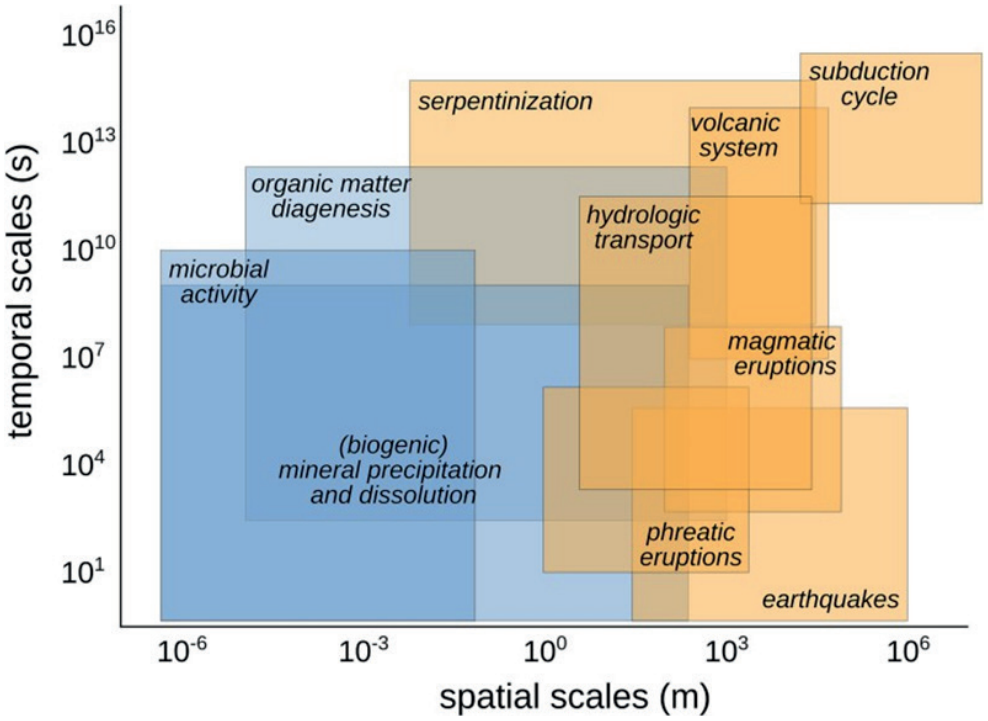


Figure 2. Conceptual diagram showing the relative spatial and temporal scales of biological (in blue) and geological (in orange) processes at convergent margins and their overlap in space–time. Similar conceptual diagrams showing the overlap between geological and biological processes in different tectonic settings can help to identify the relevant processes to track and the type of synoptic, co-located samples to collect. Reproduced from Giovannelli *et al.* (2022).

Conclusions

Microbes are much more than just food across the trophic chain. They can influence marine ecosystems in many other ways, directly through the interaction with higher organisms and indirectly through the interaction with nutrients and elements and solid Earth processes. The integration of data across disciplines and the adoption of interdisciplinary methodologies emerge as critical imperatives to bridge this knowledge gap. By embracing a holistic perspective that combines microbiology, geochemistry, geology, and oceanography, we can unravel the complex nature of the interactions between the marine biosphere and the solid Earth, shedding light on processes that lay at the heart of our planet's biogeochemical cycles. The need to integrate data and disciplines is particularly evident when considering the interaction between microbial life and geological processes at varying scales. For instance, the role of microorganisms in controlling the quality and quantity of volatiles recycled at subduction zones highlights the intricate links between biological and geological processes. Such interactions underscore the necessity of adopting an interdisciplinary approach that transcends traditional boundaries, enabling a comprehensive understanding of the mechanisms that drive geosphere-biosphere interactions in marine ecosystems.

The shift towards investigating marine microbial diversity and function at larger scales offers unprecedented opportunities to elucidate the underlying principles that govern these complex interactions. By leveraging advances in metagenomics, isotope geochemistry, and geophysical modelling, researchers can gain insights into the spatial and temporal dynamics of microbial communities and their role in geochemical cycles. This holistic approach not only enhances our

understanding of marine ecosystems but also paves the way for groundbreaking discoveries that can inform the development of sustainable solutions to environmental challenges to understand past changes and future directions.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 948972—COEVOLVE—ERC-2020-STG).

Acknowledgments

The author acknowledges Dr. M. Cascone for helpful discussion and suggestions.

Data Availibility Statement

No new data has been generated for the realisation of this article.

References

Acinas, S.G., Sánchez, P., Salazar, G., *et al*. (2021). Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Communications Biology*, 4 (1): 1–15.

Anbar, A.D. (2008). Elements and Evolution. *Science*, 322 (5909): 1481–1483.

Arrigo, K.R. (2005). Marine microorganisms and global nutrient cycles. *Nature*, 437 (7057): 349–355.

Azam, F. and Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nature Reviews Microbiology*, 5 (10): 782–791.

Baquiran, J.P.M., Ramírez, G.A., Haddad, A.G., *et al*. (2016). Temperature and Redox Effect on Mineral Colonization in Juan de Fuca Ridge Flank Subsurface Crustal Fluids. *Frontiers in Microbiology*, 7: 396.

Barosa, B., Ferrillo, A., Selci, M., *et al*. (2023). Mapping the microbial diversity associated with different geochemical regimes in the shallow-water hydrothermal vents of the Aeolian archipelago, Italy. *Frontiers in Microbiology*, 14: 1148455.

Barry, P.H., Moor, J.M., Giovannelli, D., *et al*. (2019). Forearc carbon sink reduces long-term volatile recycling into the mantle. *Nature*, 568 (7753): 487–492.

Basili, M., Timothy, J.R., Yucel, M., *et al*. (2022). Microbial community structure is shaped by physical characteristics of the Central America subduction zone. In preparation.

Bekaert, D.V., Turner, S.J., Broadley, M.W., *et al*. (2021). Subduction-Driven Volatile Recycling: A Global Mass Balance. *Annual Review of Earth and Planetary Sciences*, 49: 395–424.

Berg, J.S., Ahmerkamp, S., Pjevac, P., *et al*. (2022). How low can they go? Aerobic respiration by microorganisms under apparent anoxia. *FEMS Microbiology Reviews*, 46 (3): fuac006.

Cloetingh, S., Sternai, P., Koptev, A., *et al*. (2023). Coupled surface to deep Earth processes: Perspectives from TOPO-EUROPE with an emphasis on climate- and energy-related societal challenges. *Global and Planetary Change*, 226: 104140.

Colman, D.R., Poudel, S., Hamilton, T.L., *et al*. (2017). Geobiological feedbacks and the evolution of thermoacidophiles. *The ISME Journal*, 12 (1): 225–239.

Cordone, A., D'Errico, G., Magliulo, M., *et al*. (2022). Bacterioplankton Diversity and Distribution in Relation to Phytoplankton Community Structure in the Ross Sea Surface Waters. *Frontiers in Microbiology*, 13: 722900.

Dick, G.J. (2019). The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nature Reviews Microbiology*, 17 (5): 271–283.

Evans, K.A. (2012). The redox budget of subduction zones. *Earth-Science Reviews*, 113 (1-2): 11–32.

Falkowski, P.G., Fenchel, T., Delong, E.F. (2008). The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science*, 320 (5879): 1034–1039.

Fullerton, K.M., Schrenk, M.O., Yücel, M., *et al*. (2021). Effect of tectonic processes on biosphere–geosphere feedbacks across a convergent margin. *Nature Geoscience*, 14 (5): 301–306.

Galvez, M.E. and Pubellier, M. (2019). How Do Subduction Zones Regulate the Carbon Cycle? In: Orcutt, B.N., Daniel, I. and Dasgupta, R. (Eds.), *Deep Carbon Past Present*. Cambridge University Press: Cambridge, UK; New York, NY, pp. 276–312.

Giovannelli, D. (2023). Trace metal availability and the evolution of biogeochemistry. *Nature Reviews Earth & Environment*, 4 (1): 1–2.

Giovannelli, D., Barry, P.H., Bekaert, D.V., *et al*. (2021). Subsurface life can modify volatile cycling on a planetary scale. *Memorie della Società Astronomica Italiana*, 92: 60.

Giovannelli, D., Barry, P.H., de Moor, J.M., *et al*. (2022). Sampling across large-scale geological gradients to study geosphere–biosphere interactions. *Frontiers in Microbiology*, 13: 998133.

Giovannelli, D., Barry, P.H., de Moor, J.M., *et al*. (2020). Microbial Influences on Subduction Zone Carbon Cycling. *Eos, Transactions American Geophysical Union*, 101.

Giovannelli, D., d'Errico, G., Manini, E., *et al*. (2013a). Diversity and phylogenetic analyses of bacteria from a shallow-water hydrothermal vent in Milos island (Greece). *Frontiers in Microbiology*, 4: 184.

Giovannelli, D., Molari, M., d'Errico, G., *et al*. (2013b). Large-Scale Distribution and Activity of Prokaryotes in Deep-Sea Surface Sediments of the Mediterranean Sea and the Adjacent Atlantic Ocean. *PLoS ONE*, 8 (8): e72996.

Greening, C. and Grinter, R. (2022). Microbial oxidation of atmospheric trace gases. *Nature Reviews Microbiology*, 20 (3): 1–16.

Harris, G. (2012). *Phytoplankton Ecology: Structure, Function and Fluctuation*. Springer Science & Business Media: Heidelberg, Germany.

Hay Mele, B., Monticelli, M., Leone, S., *et al*. (2023). Oxidoreductases and metal cofactors in the functioning of the earth. *Essays in Biochemistry*, 67 (2): 1–18.

Herndl, G.J. and Reinthaler, T. (2013). Microbial control of the dark end of the biological pump. *Nature Geoscience*, 6 (9): 718–724.

Hoshino, T., Doi, H., Uramoto, G.I., *et al*. (2020). Global diversity of microbial communities in marine sediment. *Proceedings of the National Academy of Sciences of the United States of America*, 117 (44): 27587–27597.

Hügler, M. and Sievert, S.M. (2011). Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. *Annual Review of Marine Science*, 3: 261–289.

Jannasch, H.W. and Mottl, M.J. (1985). Geomicrobiology of Deep-Sea Hydrothermal Vents. *Science*, 229 (4714): 717–725.

Jelen, B.I., Giovannelli, D. and Falkowski, P.G. (2016). The Role of Microbial Electron Transfer in the Coevolution of the Biosphere and Geosphere. *Annual Review of Microbiology*, 70: 45–62.

Jørgensen, B.B. (2000). Bacteria and Marine Biogeochemistry. In: Schulz, H.D. and Zabel, M. (eds), *Marine Geochemistry*. Springer: Berlin Heidelberg, pp. 173–207.

Jørgensen, B.B., Findlay, A.J., Pellerin, A. (2019). The Biogeochemical Sulfur Cycle of Marine Sediments. *Frontiers in Microbiology*, 10: 2686.

Kelemen, P.B. and Manning, C.E. (2015). Reevaluating carbon fluxes in subduction zones, what goes down, mostly comes up. *Proceedings of the National Academy of Sciences of the United States of America*, 112 (32): E3997–E4006.

Knittel, K. and Boetius, A. (2009). Anaerobic Oxidation of Methane: Progress with an Unknown Process. *Annual Review of Microbiology*, 63: 311–334.

Koblížek, M. (2015). Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiology Reviews*, 39 (6): 854–870.

Lappan, R., Shelley, G., Islam, Z.F., *et al*. (2023). Molecular hydrogen in seawater supports growth of diverse marine bacteria. *Nature Microbiology*, 8 (4): 581–595.

Li, J.L., Schwarzenbach, E.M., John, T., *et al*. (2020). Uncovering and quantifying the subduction zone sulfur cycle from the slab perspective. *Nature Communications*, 11 (1): 514.

Lloyd, K.G. (2021). Time as a microbial resource. *Environmental Microbiology Reports*, 13 (1): 18–21.

Madsen, E.L. (2011). Microorganisms and their roles in fundamental biogeochemical cycles. *Current Opinion in Biotechnology*, 22 (3): 456–464.

Magnabosco, C., Biddle, J.F., Cockell, C.S., *et al*. (2019). Biogeography, Ecology, and Evolution of Deep Life. In: *Deep Carbon: Past to Present*. Cambridge University Press: Cambridge, UK; New York, NY, pp. 524–555.

McNichol, J., Stryhanyuk, H., Sylva, S.P., *et al*. (2018). Primary productivity below the seafloor at deep-sea hot springs. *Proceedings of the National Academy of Sciences of the United States of America*, 115 (26): 6756–6761.

Merlino, G., Barozzi, A., Michoud, G., *et al*. (2018). Microbial ecology of deep-sea hypersaline anoxic basins. *FEMS Microbiology Ecology*, 94 (11): fiy166.

Moore, E.K., Jelen, B.I., Giovannelli, D. *et al*. (2017). Metal availability and the expanding network of microbial metabolisms in the Archaean eon. *Nature Geoscience*, 10 (9): 629–636.

Pacton, M. and Gorin, G.E. (2014). Role of microorganisms in oceanic anoxic events (OAEs). *Geology Today*, 30 (5): 215–221.

Pasotti, F., Manini, E., Giovannelli, D., *et al*. (2014). Antarctic shallow water benthos in an area of recent rapid glacier retreat. *Marine Ecology*, 35 (4): 438–448.

Plank, T. and Manning, C.E. (2019). Subducting carbon. *Nature*, 574 (7777): 343–352.

Price, R.E., Amend, J.P. and Pichler, T. (2007). Enhanced geochemical gradients in a marine shallow-water hydrothermal system: Unusual arsenic speciation in horizontal and vertical pore water profiles. *Applied Geochemistry*, 22 (12): 2595–2605.

Price, R.E. and Giovannelli, D. (2017). A Review of the Geochemistry and Microbiology of Marine Shallow-Water Hydrothermal Vents. In: *Reference Module in Earth Systems and Environmental Sciences*. Elsevier: Oxford, UK.

Price, R.E., Lesniewski, R., Nitzsche, K., *et al*. (2013). Archaeal and bacterial diversity in an arsenic-rich shallow-sea hydrothermal system undergoing phase separation. *Frontiers in Microbiology*, 4: 158.

Raes, E.J., Bodrossy, L., van de Kamp, J., *et al*. (2018). Oceanographic boundaries constrain microbial diversity gradients in the South Pacific Ocean. *Proceedings of the National Academy of Sciences of the United States of America*, 115 (33): E8266–E8275.

Ren, G., Ma, A., Zhang, Y., *et al*. (2018). Electron acceptors for anaerobic oxidation of methane drive microbial community structure and diversity in mud volcanoes. *Environmental Microbiology*, 20 (6): 2370–2385.

Ricci, F. and Greening, C. (2024). Chemosynthesis: a neglected foundation of marine ecology and biogeochemistry. *Trends in Microbiology*, 32 (4): 282–293.

Rogers, T.J., Buongiorno, J., Jessen, G.L., *et al*. (2022). Chemolithoautotroph distributions across the subsurface of a convergent margin. *The ISME Journal*, 16 (1): 1–11.

Savaglia, V., Lambrechts, S., Tytgat, B., *et al*. (2024). Geology defines microbiome structure and composition in nunataks and valleys of the Sør Rondane Mountains, East Antarctica. *Frontiers in Microbiology*, 15: 1378378.

Selci, M., Cascone, M., Rogers, T., *et al*. (2024). Origin and fate of methane in the Central American convergent margin. Submitted.

Sievert, S. and Vetriani, C. (2012). Chemoautotrophy at Deep-Sea Vents: Past, Present, and Future. *Oceanography*, 25 (4): 218–233.

Sigman, D.M. and Hain, M.P. (2012). The biological productivity of the ocean. *Nature Education Knowledge*, 3 (10): 21.

Sunagawa, S., Coelho, L.P., Chaffron, S. *et al.* (2015). Structure and function of the global ocean microbiome. *Science*, 348(6237), 1261359.

Tagliabue, A., Aumont, O., and Arrigo, K.R. (2017). The impact of a realistic representation of ocean biogeochemistry on the simulation of global ocean productivity. *Global Biogeochemical Cycles*, 31(2), 246-267.

Taupp, M., Oren, A., Hiran, T. *et al.* (2010). Distribution and diversity of *Epsilonproteobacteria* in sediments of the deep-sea hydrothermal vent field Logatchev. *Environmental Microbiology*, 12(8), 2217–2229.

Ulloa, O., Canfield, D.E., DeLong, E.F., *et al*. (2012). Microbial oceanography of anoxic oxygen minimum zones. *Proceedings of the National Academy of Sciences of the United States of America*, 109 (40): 15996–16003.

Upin, H.E., Newell, D.L., Colman, D.R., *et al.* (2023). Tectonic settings influence the geochemical and microbial diversity of Peru hot springs. *Communications Earth & Environment*, 4 (1): 1–8.

Valentine, D.L. (2011). Emerging Topics in Marine Methane Biogeochemistry. *Annual Review of Marine Science*, 3: 147–171.

Vitale Brovarone, A., Sverjensky, D.A., Piccoli, F., *et al*. (2020). Subduction hides high-pressure sources of energy that may feed the deep subsurface biosphere. *Nature Communications*, 11 (1): 3880.

Ward, B., Capone, D., Zehr, J. (2007). What's New in the Nitrogen Cycle? *Oceanography*, 20 (1): 101–109.

Whiticar, M.J. (2020). The biogeochemical methane cycle. In: *Hydrocarbons, Oils and Lipids: Diversity, Origin, Chemistry and Fate*. Springer International Publishing, pp. 669–746.

Winder, M. and Sommer, U. (2012). Phytoplankton response to a changing climate. *Hydrobiologia*, 698 (1): 5–16.

Xie, W., Wang, F., Guo, L., *et al*. (2011). Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *The ISME Journal*, 5 (3): 414–426.

Yakimov, M.M., Cono, V.L., Spada, G.L., *et al*. (2015). Microbial community of the deep-sea brine Lake Kryos seawater–brine interface is active below the chaotropicity limit of life as revealed by recovery of mRNA. *Environmental Microbiology*, 17 (3): 364–382.

Yücel, M. (2013). Down the thermodynamic ladder: A comparative study of marine redox gradients across diverse sedimentary environments. *Estuarine, Coastal and Shelf Science*, 131: 83–92.

Yücel, M., Gartman, A., Chan, C.S., *et al.* (2011). Hydrothermal vents as a kinetically stable source of iron-sulphide-bearing nanoparticles to the ocean. *Nature Geoscience*, 4 (5): 367–371.

Metagenomic Analysis for Obtaining Taxonomic and Functional Information on Marine Microbiomes

Javier Tamames

Microbiome Analysis Laboratory. Centro Nacional de Biotecnología, Madrid, Spain.

Abstract

Historically, microbial studies relied on culturing, but most environmental microbes resist isolation, significantly underestimating diversity. Early molecular techniques like rDNA sequencing opened insights but had limitations: they didn't reveal metabolism, suffered from primer biases, and couldn't detect truly rare microbes. While marker gene sequencing (*e.g.,* 16S/18S rRNA) advanced our understanding of community composition, it lacked functional information and was prone to biases. Metagenomics overcomes these issues by directly sequencing environmental DNA, providing comprehensive taxonomic and functional insights without PCR biases. Our SqueezeMeta pipeline makes complex metagenomic analysis accessible and efficient, generating Metagenome-Associated Genomes (MAGs). This portable platform, coupled with a MinION sequencer, enables rapid, autonomous analyses anywhere, as demonstrated by our studies revealing the distribution of plastic degradation and antibiotic resistance genes in oceans.

Keywords: metagenomics, microbial dark matter, amplicon sequencing, Metagenome-Associated Genomes (MAGs), SqueezeMeta pipeline, MinION sequencer

Main Text

Classically, microbial members of the biosphere could only be studied by isolating microbes into pure culture. This had the advantage that cultures could be studied physiologically and genetically. However, a large fraction (or even the major fraction) of microbes in the environment are not easily culturable using current techniques. For example, Crespo *et al.* (2016) compared results from very deep sequencing with a culturing approach and found that most of the cultured taxa were not obtained by HTS, despite the high sequencing effort. Sanz-Sáez *et al.* (2023) found that the proportion of culturable bacteria in deep ocean samples was higher than at the surface. Nevertheless, the major fraction of them were still not retrieved in culture.

When molecular techniques to retrieve rDNA from natural samples were developed in the 1990s, a window was opened into the diversity of natural microbial communities (Giovannoni *et al.* 1990). In the following years, very productive studies concentrated on analyzing this diversity. There were at least three caveats, however. First, while rDNA provides identity, the metabolism of the microbe remains inaccessible. Second, the primers used in PCR amplification are not as universal as desired, as shown, for example, by the discovery of the candidate phyla radiation (Hug *et al.* 2016). Thus, there is a microbial "dark matter" that remains to be discovered. And third, despite the sequencing power of recent high-throughput techniques, they are still not powerful enough to reach the truly rare microbes.

For studying the rare biosphere, marker gene sequencing-based methods are the most common approach. These marker genes are usually the V4 hypervariable region of the 16S rRNA gene in prokaryotes, and the V9 region of the 18S rRNA gene for eukaryotes (the ITS rDNA region

can also be used). Traditionally, an amplicon library of these genes is prepared using standard primers, sequenced, and the results are clustered into OTUs, defined as sequences with 97% identity. Recently, the use of ASVs (Amplicon Sequence Variants) has been widely adopted, where sequences are not clustered based on percent identity. Instead, single-nucleotide variants are evaluated to distinguish real variants from those most likely explained by sequencing errors (García-García *et al.* 2019).

Approaches based on amplicon sequencing are prone to biases and artifacts, including the lack of universality of the primers used in amplification (because PCR primers are designed based on well-known taxa), and the unequal coverage of sequence databases that are still not comprehensive for many rare taxa. Also, the different copy numbers of marker genes can significantly alter the measured abundances of each taxon. Contamination during sampling and/or DNA extraction must also be accounted for (Poretsky *et al.* 2014).

Nevertheless, these marker gene approaches only provide information on the composition and abundance of the microbial community. That is, they can identify the members of the microbiome but cannot provide information about the functional content encoded in their genomes. Metagenomics emerges as a very powerful tool to address both the taxonomy and composition of the rare biosphere, with the possibility of recovering genomic information about the functional potential of these organisms (Simon and Daniel 2011). Metagenomics always involves sequencing genomic DNA, instead of individual marker genes. Since metagenomics deals with non-amplified microbial DNA extracted directly from environmental samples, it is not affected by biases impacting amplicon sequencing. Additionally, the progress of sequencing techniques and instruments has made it an affordable experimental technique.

In metagenomics, DNA is isolated from the sample irrespective of its origin. Therefore, it is a mixture of DNA from the different organisms present in the sample (bacteria, archaea, eukaryotes, viruses). That mixed DNA is then subjected to sequencing using current high-throughput sequencing techniques. The resulting sequences are then analyzed by bioinformatics techniques, usually to obtain: 1) A taxonomic profile detailing the taxonomic composition of the samples. 2) A functional profile indicating the possible functionalities encoded in these DNAs. And 3) Metagenome-associated genomes (MAGs), which are draft genomes of the organisms in the sample. MAGs are important because they can provide a glimpse of the genome, allowing researchers to infer the possible lifestyle of the corresponding organism (Setubal 2021). For instance, this includes its metabolism, resistance genes, functional capabilities, and possible relationships with other organisms and the environment.

MAGs have substantially improved our knowledge of the microbial biosphere. For instance, Nishimura *et al.* were able to obtain more than 50,000 MAGs using 2,000 marine metagenomes (Nishimura and Yoshizawa 2022). These metagenomic analyses usually rely on assembling the raw sequences from the samples into contigs, which are pieces of the genomes present in the underlying microbiomes. However, as this assembly step is complex and computationally intensive, it can be avoided by analyzing the raw reads instead. This has the disadvantage of not providing MAGs, which are often one of the most desired results.

A standard metagenomic analysis can involve the following steps: assembly of raw reads to produce contigs, gene prediction to detect ORFs in these contigs, homology searching to assign taxa and functions to the ORFs, abundance estimation via mapping raw reads to the contigs, and MAG production by binning algorithms recognizing contigs belonging to the same species. The associated bioinformatics is thus complex and often presents a bottleneck. Many diverse software tools are involved, which must be installed and tailored, and typically demand ample computational resources.

To relieve this burden, we developed the SqueezeMeta pipeline (Tamames and Puente-Sánchez 2019) with several objectives in mind: 1) offering a fast and easy-to-use platform for performing the complete analysis of metagenomes. Our goal was to include all the usual steps in metagenomic analysis with state-of-the-art tools, but making them attainable to all users, regardless of their bioinformatic skills. 2) Breaking the dependence on large computers, enabling it to run with scarce resources, and even on laptops. 3) Providing additional tools for performing statistical analysis and sharing the results.

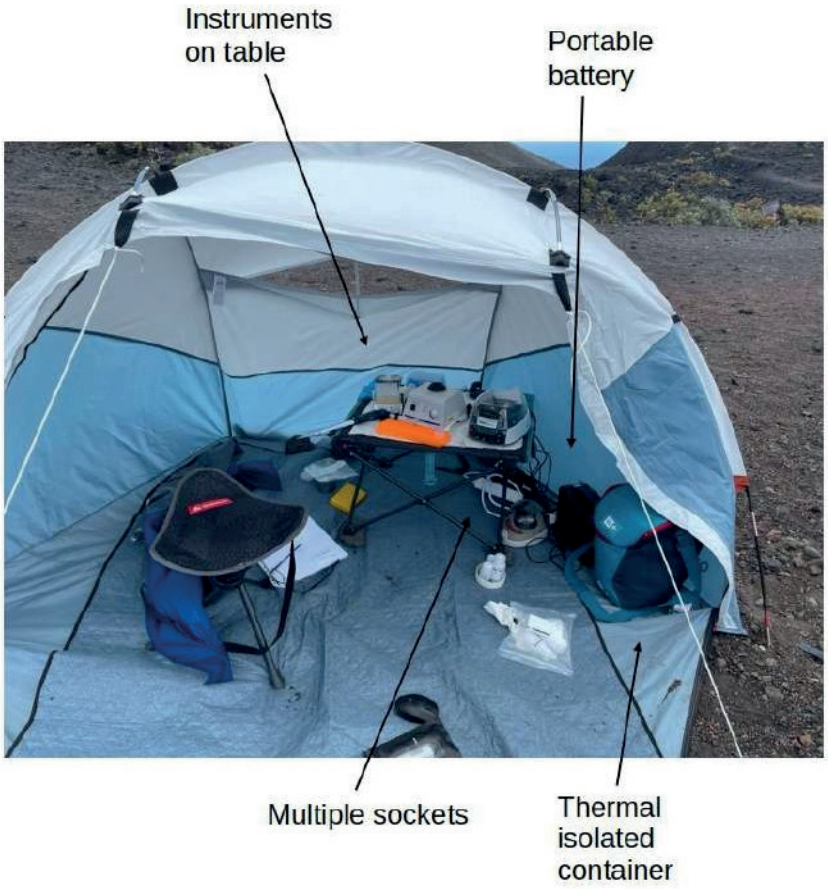


Figure 1. Setting of the portable sequencing laboratory for analysing lava rocks in La Palma island.

SqueezeMeta can easily obtain taxonomic and functional profiles from metagenomes. It is also capable of using custom databases provided by the user, making it possible to extend the functional annotation to any classification of interest. The pipeline also obtains MAGs using several different tools that can be combined into a single MAG dataset, which are then analyzed for completeness and contamination.

The development of this pipeline, together with the availability of portable sequencers, has allowed us to produce a full metagenomic analysis in less than 24 hours, in a fully autonomous way, with no reliance on electric power or internet connectivity (Tamames *et al.* 2019). The platform comprises the following modules (Figure 1):

1. A portable DNA extraction laboratory, small enough to be carried by a single person, to produce the required amounts of environmental DNA.
2. A MinION DNA sequencer, to sequence that DNA and produce metagenomic DNA sequences.
3. The bioinformatic pipeline SqueezeMeta, running on a small laptop (the same one that manages the sequencing), to analyze the metagenomic DNA sequences; and
4. The stand-alone statistical package SQMTools (Puente-Sánchez *et al.* 2019) is used for statistical analysis of the data. It is coupled with our new ShinySQM library (https://github.com/redondrio/ SQMxplore), which allows for the creation of interactive web pages and interfaces for openly sharing the results. We tested the platform in diverse circumstances and environments, including sampling soils and rocks in La Palma Island (Canary Islands, Spain), in Antarctica, and sampling marine waters onboard in Ria de Vigo (Spain).

To demonstrate the utility of the SqueezeMeta pipeline for studying metagenomic datasets, we analyzed several samples from the Malaspina oceanographic campaign, collected in 2011 from various ocean locations (Duarte 2015), for which metagenomic DNA sequences were available. We examined the taxonomic composition of the microbiomes in diverse oceanic locations. To illustrate our platform's capabilities in addressing functional composition using custom databases, we included the analysis of genes related to plastic biodegradation, as listed in the PlasticDB database (Gambarini *et al.* 2022). We also included the abundances of genes related to antibiotic resistance (ARGs) from the CARD database (Jia *et al.* 2017). The results can be seen in Figure 2 (upper panel), which shows only surface samples (3 m depth). Abundances are measured in RPKM (Reads Per Kilobase and per Million reads), a standard normalization method for sequencing depth and gene length, allowing for the comparison of values across different samples.

The abundance of plastic degradative genes is rather uniform across diverse areas. Interestingly, however, the highest abundance of these degradative genes occurs in the center of the Pacific Ocean, close to the so-called Great Pacific Garbage Patch (Wong *et al.* 1974). It was also possible to analyze vertical profiles for the same samples because, at each station (data point), sampling was done at different depths (Figure 2, lower panel). The distributions are heterogeneous, but the highest abundances can be found in Mesopelagic-Bathypelagic zones, in contrast to surface waters, where abundances tend to be lower. Some reports indicate that the concentration of microplastics can be higher in mid-water depths (Choy *et al.* 2019; Gunaalan *et al.* 2024), which is consistent with these results.

Putative ARGs are present in higher abundances, corresponding to their widespread distribution in all kinds of organisms and environments (Zhuang *et al.* 2021). As in the previous case, the abundances are fairly uniform across all samples, with the maximum abundance found in mid-Atlantic waters. Such similarity could be related to constant and homogeneous selective pressures throughout the ocean. As in the case of plastic degradation genes, higher abundances are found at greater depths. This is consistent with previous results obtained in selected locations, such as the Black Sea (Sabatino *et al.* 2022). Of course, it is possible to dissect these global abundances into individual ARGs and to inspect the taxonomic origin of these genes.

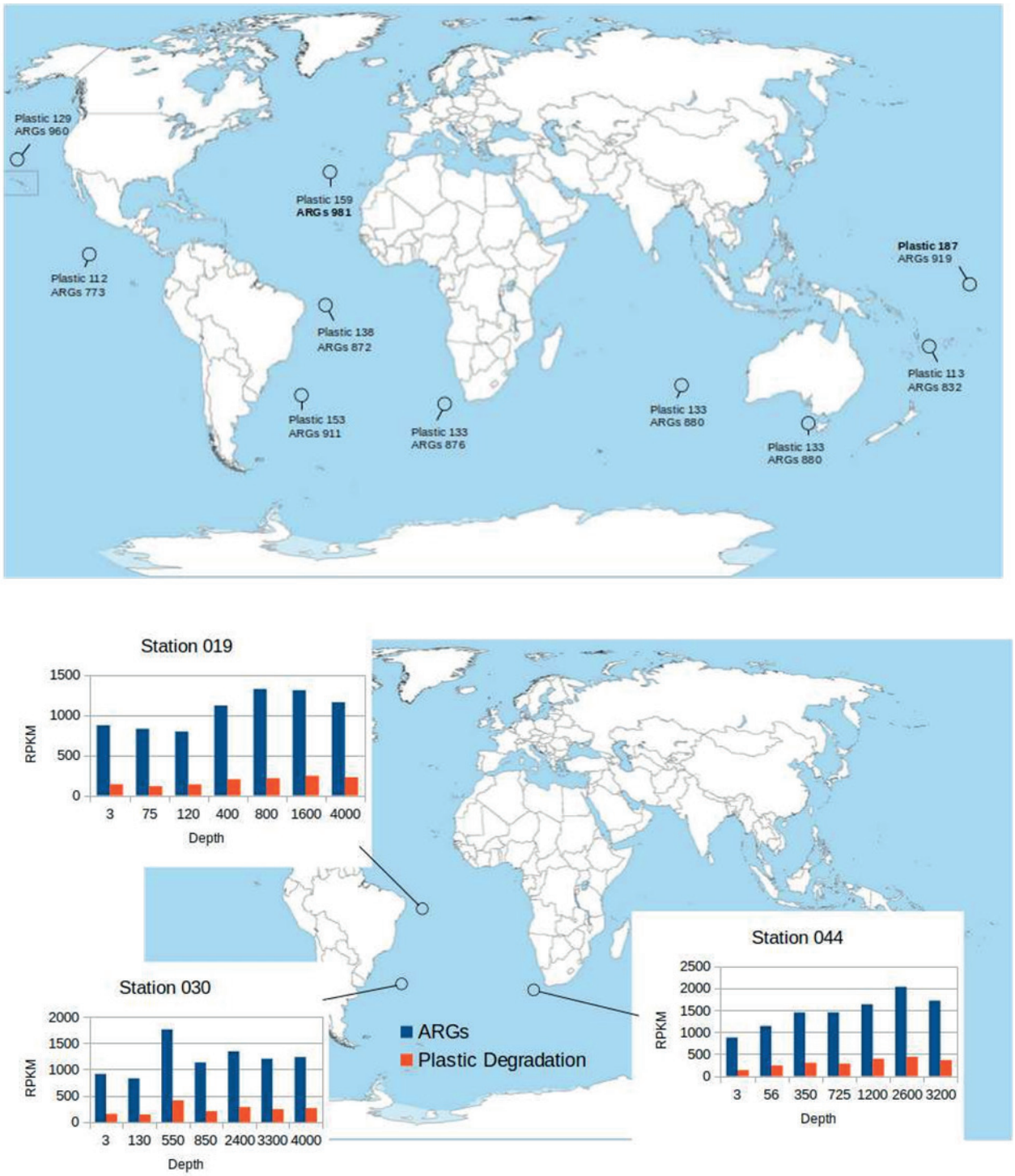


Figure 2. Upper: Abundance of Plastic degradative genes and Antibiotic Resistance genes (ARGs) in diverse oceanic locations. The samples were taken by the Malaspina Oceanographic Cruise (2011) and correspond to surface samples (3 meters depth). Lower: Vertical profile of abundances of the same genes in three selected locations, based on samples taken at different depths.

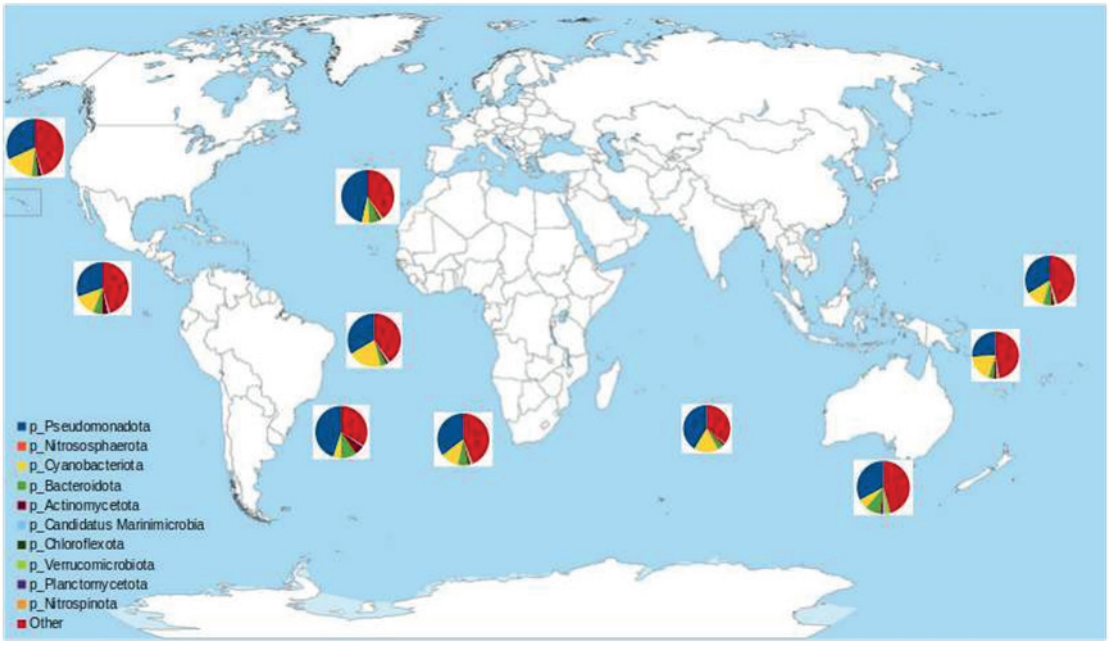


Figure 3. Taxonomic distribution of surface samples (3 meters depth) in Malaspina cruise (2011). The pie charts show the percentage distribution of the dominant groups.

The taxonomic classifications of the metagenomic samples can be seen in Figure 3, showing the most abundant phyla in the surface samples (3 m). Clear differences can be seen between the different samples, although Pseudomonadota and Cyanobacteriota are dominant groups in all samples. The choice of phylum taxonomic rank was made for simplicity, but the same analysis can be performed for other taxonomic ranks.

In summary, our methods are capable of analyzing metagenomic samples using a wide variety of tools, providing comprehensive functional and taxonomic information on the microbiomes present in a given environmental location. This can complement the abiotic information obtained by other means with a clear depiction of how the microbial communities function under these circumstances.

References

Choy, C.A., Robison, B.H., Gagne, T.O., *et al*. (2019). The vertical distribution and biological transport of marine microplastics across the epipelagic and mesopelagic water column. *Scientific Reports*, 9 (1): 7843.

Crespo, B.G., Wallhead, P.J., Logares, R., *et al.* (2016). Probing the Rare Biosphere of the North-West Mediterranean Sea: An Experiment with High Sequencing Effort. *PLoS ONE*, 11 (7): e0159195.

Duarte, C.M. (2015). Seafaring in the 21st century: The Malaspina 2010 circumnavigation expedition. *Limnology and Oceanography Bulletin*, 24 (1): 1–2.

Gambarini, V., Pantos, O., Kingsbury, J.M., *et al*. (2022). PlasticDB: a database of microorganisms and proteins linked to plastic biodegradation. *Database*, 2022: baac008.

García-García, N., Tamames, J., Linz, A.M., *et al*. (2019). Microdiversity ensures the maintenance of functional microbial communities under changing environmental conditions. *The ISME Journal*, 13 (11): 2969–2983.

Giovannoni, S.J., Britschgi, T.B., Moyer, C.L., *et al.* (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature*, 345 (6270): 60–63.

Gunaalan, K., Almeda, R., Vianello, A., *et al*. (2024). Does water column stratification influence the vertical distribution of microplastics? *Environnemental Pollution*, 340: 122865.

Hug, L.A., Baker, B.J., Anantharaman, K., *et al*. (2016). A new view of the tree of life. *Nature Microbiology*, 1 (5): 16048.

Jia, B., Raphenya, A.R., Alcock, B., *et al*. (2017). CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 45 (D1): D577-D583.

Nishimura, Y. and Yoshizawa, S. (2022). The OceanDNA MAG catalog contains over 50,000 prokaryotic genomes originating from various marine environments. *Scientific Data*, 9 (1): 1–11.

Poretsky, R., Rodriguez-R, L.M., Luo, C., *et al*. (2014). Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLoS ONE*, 9 (4): e93827.

Sabatino, R., Cabello-Yeves, P.J., Eckert, E.M., *et al*. (2022). Antibiotic resistance genes correlate with metal resistances and accumulate in the deep-water layers of the Black Sea. *Environmental Pollution*, 312: 120033.

Sanz-Sáez, I., Sánchez, P., Salazar, G., *et al*. (2023). Top abundant deep ocean heterotrophic bacteria can be retrieved by cultivation. *ISME Communications*, 3 (1): 92.

Setubal, J.C. (2021). Metagenome-assembled genomes: concepts, analogies, and challenges. *Biophysical Reviews*, 13 (5): 905–909.

Simon, C. and Daniel, R. (2011). Metagenomic Analyses: Past and Future Trends. *Applied and Environmental Microbiology*, 77 (4): 1153–1161.

Tamames, J. and Puente-Sánchez, F. (2019). SqueezeMeta, A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline. *Frontiers in Microbiology*, 9, 3349.

Tamames, J., Jiménez, D., Redondo, Á. et al. (2023). In situ metagenomics: A platform for rapid sequencing and analysis of metagenomes in less than one day. *Molecular Ecology Resources*, 1755–0998.13909.

Seafloor Dissolved Sulfide Microgradients and Sulfur Cycle Intermediates Probed by Voltammetry: Potential and Future Directions

Mustafa Yücel

Middle East Technical University (METU), Institute of Marine Sciences, Erdemli, Mersin, Türkiye

Abstract

Seafloor dissolved sulfide gradients are prevalent features in productive sea basins as well as deep-sea chemosynthetic habitats. These gradients can occur over hundreds of meters in the water column in deep anoxic basins and they reduce to centimeters to millimeters in surface sediments under productive waters as well as across hydrothermal vents and cold seeps. Microelectrode cyclic voltammetry allows for the collection of large quantities of biogeochemical data on the redox state of habitats, which can be used for big data integration and co-visualization with high resolution physical measurements and genomics analyses.

Keywords: redox, sulfide, voltammetry, anoxic sediments, hydrothermal vents

Introduction

Ocean chemical gradients are shaped by biological, geochemical and physical processes. Among the many biologically important elements involved in biogeochemical cycles, sulfur cycle becomes significant in low-oxygen to anoxic seas and sulfur-based chemosynthetic ecosystems such as hydrothermal vents. Seafloor dissolved sulfide gradients are prevalent features in such habitats. The sizes of these gradients reduce to millimeters in surface sediments under productive waters and throughout microbial mats, but they can occur over hundreds of meters in the water column in deep anoxic basins (Jorgensen *et al.* 2019). Sulfide fluxes can also happen at much smaller spatial scales across gas seeps, mud volcanoes, and seafloor hydrothermal vents, with the added characteristics of rapid transport, high mixing rates, and intense precipitation-redox cycles (Le Bris *et al*. 2019).

Kinetics of sulfide oxidation reactions can lead to the accumulation of redox reaction intermediates, particularly in fast sulfur cycle systems like surface marine sediments or vent mixing zones. The formation and persistence of these intermediates put additional constraints (or opportunities) on the benthic ecosystem functioning, with further implications to oxygen dynamics, nutrient cycling and biological carbon fixation in the marine environment. Biotic and abiotic processes can eliminate hazardous sulfide from the environment via sulfide oxidation; however, this occurs at the cost of oxidants like oxygen and nitrate. However, some chemoautotrophic microorganisms use the process as a source of energy (Sievert *et al.* 2008), which allows for the development of highly productive communities like those that surround seafloor hydrothermal vents. While the oxidation of sulfide in pelagic marine sediments primarily occurs through the reduction of Fe(III) and Mn(IV) (Aller and Rude 1988; Zopfi *et al*. 2008), or nitrate reduction in organic-rich environments (Preisler *et al*. 2007; Grunke *et al.* 2011), direct reactions between O_2 and H_2S are more frequent for hydrothermal vent diffuse flows. This is because these chemical species can coexist in these waters, which creates a redox disequilibrium that is utilized by chemoautotrophic microbes (Johnson *et al*. 1988; Gartman *et al*. 2010; Luther *et al*. 2011).

As these factors regulate the short-term availability of sulfide in sulfide-based chemosynthetic habitats, rapid probing of chemical disequilibrium and the detection of co-existence of H_2S, O_2, and catalytic metals like Fe, Mn, Cu, and Zn have become important targets of deep-sea detection (Le Bris *et al*. 2019). Furthermore, because of our growing knowledge of the role that metals have played in Earth's evolution (Mele *et al*. 2023) and the potential implications for remote detection of redox-disequilibrium in other ocean habitats of the Solar System (Yücel *et al.* 2025), co-availability of redox couples is also becoming a more urgently needed research frontier (Giovannelli, this volume).

In this communication, using newly analyzed data from various seafloor habitats we demonstrate utility of the voltametric electrochemical approach from i) seafloor sulfide-rich sediments of the Baltic Sea and the Black Sea ii) deep-sea hydrothermal vents. Along with providing spatially resolved dissolved sulfide datasets, we present a comparative analysis of a large set of voltametric scans from sulfide gradients from these habitats into their potential to generate sulfur intermediate products such as electrochemically (gold-amalgam voltammetry) detectable polysulfide and iron monosulfide species.

Application of Microelectrode Voltammetry

The application of voltametric microelectrodes has been exhaustively described in many studies (*e.g.,* Brendel and Luther 1995; Luther *et al.* 2008; Yücel 2013). These microelectrodes, named also as voltametric sensors, are solid-state gold-amalgam micro-discs that are prepared through the plating of mercury on a 100µm-diameter gold disc electrode, fabricated via the sealing of a gold microwire in polyethyletherketone (PEEK) tubing or glass encasings (Luther *et al*. 1999). They are capable of simultaneously measuring O_2, H_2S, Fe(II), Mn(II) as well as qualitatively detecting soluble FeS and Fe(III) species. Therefore, a high-resolution vertical profile of these chemical species can be obtained and their fluxes (if any) across the sediment-water interface can be quantified.

The voltametric measurements reported in this communication were performed by cyclic voltammetry with scan parameters tailored for sulfide-rich conditions (Luther *et al*. 2001a, 2008; Rozan *et al*. 2000; Yücel *et al*. 2013). The preparation of electrodes includes solid state gold-amalgam, constructed using 100µm gold wire. The reference electrode is a solid state, silver wire that is coated by AgCl using a 3M KCl or NaCl solution. Scan rates ranged from 400-2000 mV/s for the work presented here. In sulfide-rich conditions an electrochemical conditioning step is required to condition (or clean) the electrode surface in between each measurement. The conditioning was performed at -0.1 C and -0.9 V a time range between 2-10 s, depending on the matrix, for removing of the electroactive species on the electrode surface. Cyclic voltammetry is the scan from positive to negative and final positive values. The scan rate, conditioning time, cleaning time, scan type are the prominent parameters for one measurement. To illustrate, cyclic voltammetry with conditioning at -0.1 V for 2s and then at -0.9 V for 5s, takes about 9s to run at a scan rate of 1000 mV/s. This approach provides a detection limit of 0.1 µM and 10 µM for H_2S and O_2 calculated by linear regression of standards.

A typical sediment profile with voltametric microelectrodes was taken immediately after the retrieval of the multi-cores. Micro-profiling on intact sediment cores started within 10 minutes of the core retrieval on board and the measurements were completed typically within 45 minutes. Repeated profiling on cores left undisturbed for several hours in the laboratory resulted in similar profiles. A micro-manipulator was used to attach the working electrode, with the counter and reference electrodes remaining fixed in the overlying core-top water.

The hydrothermal system voltametric results shown here were obtained *in situ*, with a submersible electrochemical analyzer (AIS, USA) manipulated from the Alvin submarine, on board R/V Atlantis. A four-electrode multiplexed channel consisting of four gold-amalgam working electrodes were mounted inside a wand where only the tips were exposed to outside. The wand was connected to the main AIS *in situ* unit via a 4-m cell cable. Reference and counter electrodes were fixed to a secure spot in the basket. Submersion in a conductive seawater environment ensured electron flow between electrodes also excellent insulation from any electrical noise.

Sulfide Calibrations and Utility of Comparison of Forward vs Reverse Scans

Regular calibrations during a field trip or an expedition are necessary to ensure the quality of the in-situ data. We have performed laboratory sulfide calibrations using standard additions of a sulfide stock solution to an argon-purged, anoxic seawater ideally taken from the study environment. The sulfide stock solutions were prepared from Na_2S-$9H_2O$ crystals and their final concentration were determined by iodometric titration. A window of data between -0.4 to -1.2 V extracted from a typical cyclic voltammogram resulting from the scan from -0.1 V to -1.8 V (so called 'forward scan', the upper part of the curve) and the back-scan from the -1.8 V end point to the starting potential (so called 'reverse scan) is shown in Figure 1. The large peak at around -0.9 V in the forward scan is due to the reaction of unbound S(-2) species in H_2S, HS^- and S_x^{2-} with the amalgam electrode surface. The wave-like signal during the reverse scan also belongs to these species but also to the S(-2) in FeS or organic polysulfides. The latter two do not generate a forward wave, hence the comparison of the two signals with laboratory calibrations can be useful for a first order estimate of aqueous sulfide speciation.

In Figure 1, laboratory calibrations of S(-2) (H_2S/HS^-) are also shown for comparison. Each electrode will have a different calibration hence a specific ratio of forward to reverse scans in sulfide calibrations. In the case shown, the forward and reverse wave signal ratios ranged from 4 to 6 for the set of electrodes used in the study. In the voltammogram taken from the 3mm depth below a *Beggiatoa* mat, sulfur speciation was dominated was sulfide oxidation products as evidenced by the splitting of the forward peak. As such, if any polysulfides, or FeS or organic S forms which do not yield a 'free' S(-2) signal would result in the suppression of the forward wave, but the total S(-2) wave on the return leg of the voltammogram would remain constant.

The comparisons in deviations from the forward/reverse S(-2) signal ratios during the lab calibrations with the field data may be useful in post-processing and data analysis can yield the estimation of a 'bound sulfide' pool. Here we show that such an approach may already point to previously unknown dynamics in different seafloor systems, demonstrating three cases in the Baltic Sea, the Black Sea sediments and seafloor hydrothermal vents at the East Pacific Rise.

Case 1: *Beggiatoa* Mats at the Floor of the Baltic Sea

In Yücel *et al.* 2017, we have described in detail the sulfur cycling in the extensive bacterial mats covering the seafloor under a dynamic redoxcline in the Gotland Basin. The study was conducted in 2013 during an expedition by R/V Alkor. High dissolved sulfide fluxes from sediments (Figure 2) are consumed by mats composed of filamentous *Beggiatoa.* In the Baltic Sea, a hypoxic basin, the presence of sulfur oxidation intermediates occurs in spatially distinct zones along a sedimentary redox gradient, pointing to a polysulfide-yielding oxidation pathway within benthic microbial mats but a FeS-yielding - probably abiotic process in deeper sediments.

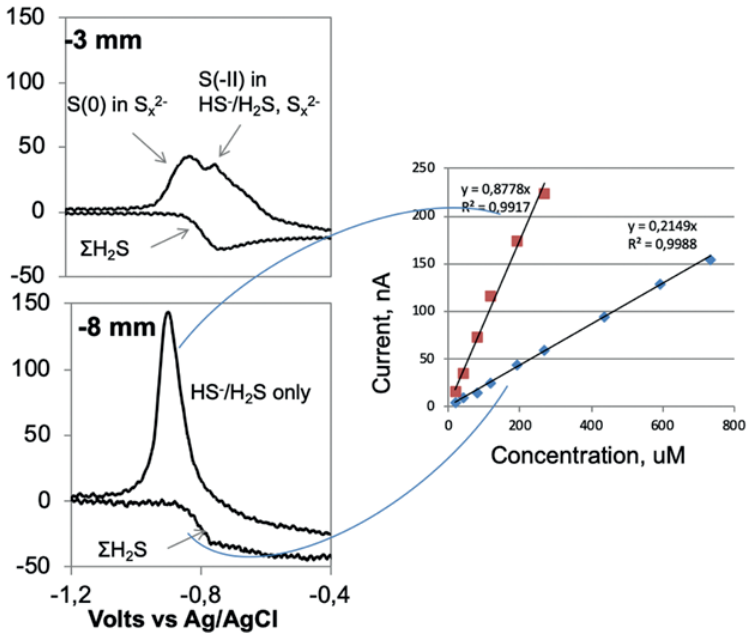


Figure 1. Left: Microelectrode voltammograms over a bacterial mat (3 mm) and below a *Beggiatoa* mat (8 mm) in the Baltic Sea. Right: Laboratory calibrations of sulfide and the calibration of the forward peak and reverse 'wave' type signal.

Here, we show that this explicit detection of FeS is well corroborated by the shifting reverse to forward cyclic voltametric signal ratio (as defined above) related to sulfide deeper in the core. In Figure 3 the ratio of the forward scan sulfide peak height to the magnitude of the wave-like signal in the reverse scan is plotted.

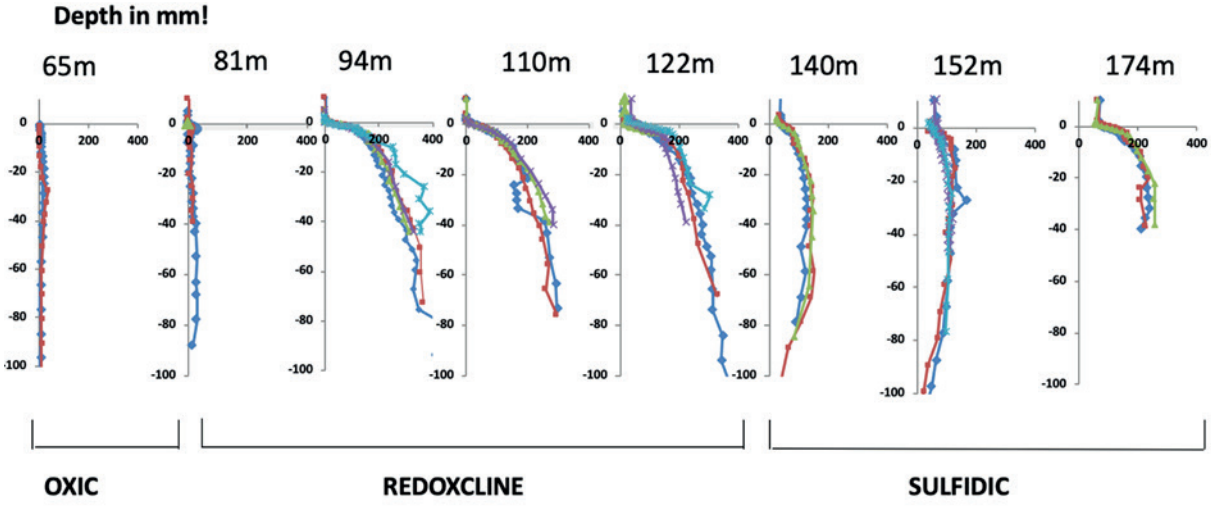


Figure 2. Total dissolved sulfide in the Eastern Gotland Basin seafloor (from Yücel *et al.* 2017).

While in most cores the ratio progressively shifted to higher values, indicating the dominance of free sulfide, in the uppermost few mm depths a clear shift to smaller values was detected in all the cores. This shift is more notable in the cores with bacterial mats on their surface.

Considering the laboratory calibration ratios of the forward/reverse signals as typical signature of a free sulfide pool, we estimate a percentage fraction of the 'bound' sulfide pool in the zones where the ratios become much smaller. Shown in Figure 4, this exercise points to that up to 60% of the total sulfide pool may be in 'bound' forms such as organic sulfides or inorganic, acid-volatile metal sulfides such as FeS.

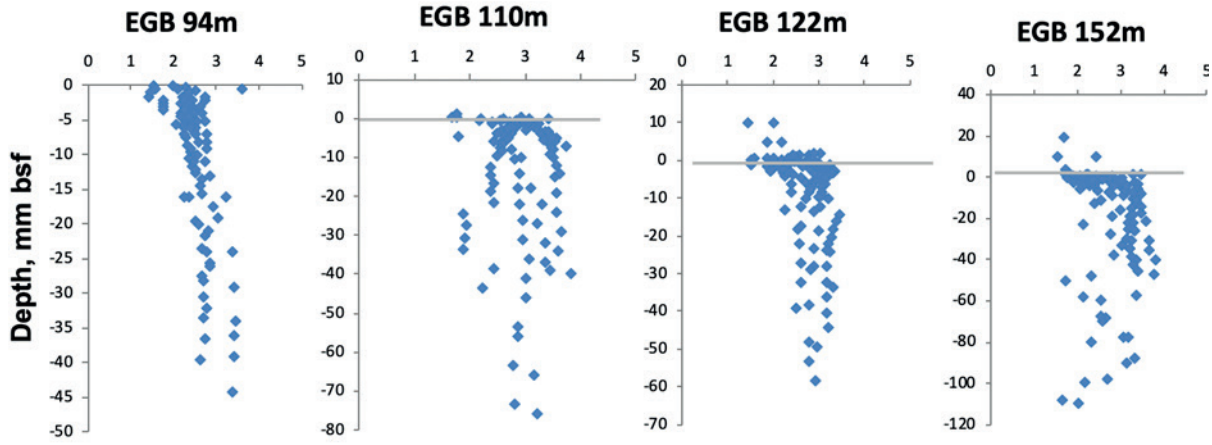


Figure 3. Free vs total sulfide signal ratio (nA/nA) significantly decreases near the interface (gray line). 'EGB' stands for Eastern Gotland Basin, where the sediment cores were taken in the Baltic Sea. The water depth of the coring locations is also indicated in the title.

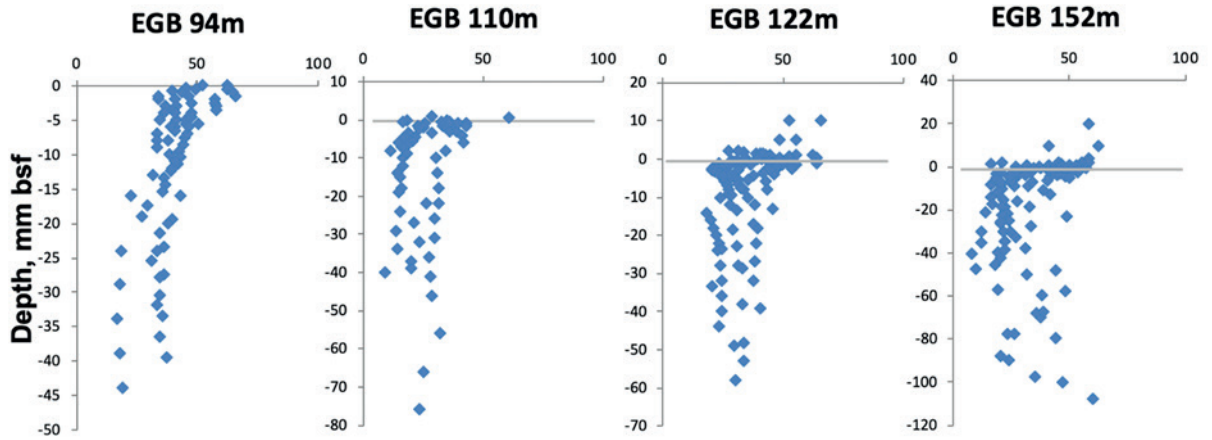


Figure 4. Concentration-based % bound sulfide sediment-water interface (gray line).

Case 2: Black Sea Suboxic Zone and Deep Anoxic Zone Sediments

During the R/V Pelagia expedition in 2013, we collected extensive microelectrode voltammetry data from multicores collected from deep Black Sea seafloor. Results related to metal cycling and turbidites have been discussed in Kraal *et al.* (2019). Here for the first time, we report the sulfide micro-profiles across the basin. The cores taken from the anoxic basin had generally increasing sulfide concentrations downcore. Profiles from the deep basin are shown in Figure 5. Highest measured sulfide concentrations in the porewaters were in Station 12 in the Western Central basin of the Black Sea, reaching about 800 µM at 10 cm depth. Diffuse flux calculations based on these

data indicate sulfide fluxes from the sediments to the water column increase with water depth.

The analysis of the data obtained from sediments underlying suboxic waters of the Northwestern shelf did not indicate that these areas are an important source of manganese or iron to the overlying water column. However, the core obtained from the suboxic station 13, which is under the influence of the Mediterranean inflow and a dynamic redoxcline, had at least several hundred micromolar of dissolved manganese in the top 10 cm of the sediments with increasing concentrations downcore, indicating that these sediments are a strong source of dissolved Mn to the overlying suboxic zone of the Black Sea (Figure 6).

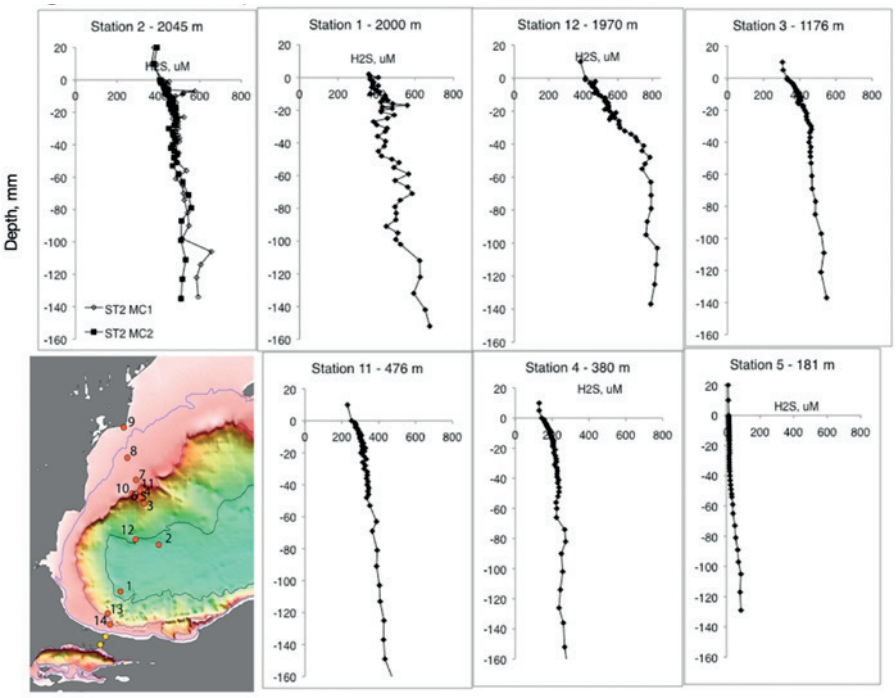


Figure 5. Stations occupied in the 2013 R/V Pelagia expedition and sediment microelectrode profiles of total S(-2) in the porewaters.

Similar to observations in the Baltic Sea, an ongoing sulfidation-related pathway in the deeper part of the upper sediment column of the Black Sea's benthic redox gradient is only implicitly indicated by the changing ratios of sulfide-dependent signal magnitudes during forward and reverse scans. Station 13 results, shown in Figure 6, are quite demonstrative here as the forward/reverse scan ratio even goes down below 1, with the co-occurring high Mn(II) signal. The lack of a distinct FeS signal here either points to larger than 100 micrometer (hence not electroactive) dissolved particles and/or organic sulfur species.

A similar computation was made for the Black Sea sediments. Cumulative percentages of estimated 'bound sulfide' levels for all data less than 250 µM are plotted for both Baltic Sea and Black Sea cores in Figure 7. We find that overall redoxcline (Baltic Sea) and suboxic core (Black Sea) scans tend to have a higher fraction of bound sulfide, exceeding 60%. In cores collected from deep locations, the ratios computed for the porewaters < 250 µM had percentages stabilized around 20%, indicating that a fraction of the total sulfide still did not behave as truly 'free' sulfide, compared to laboratory calibrations.

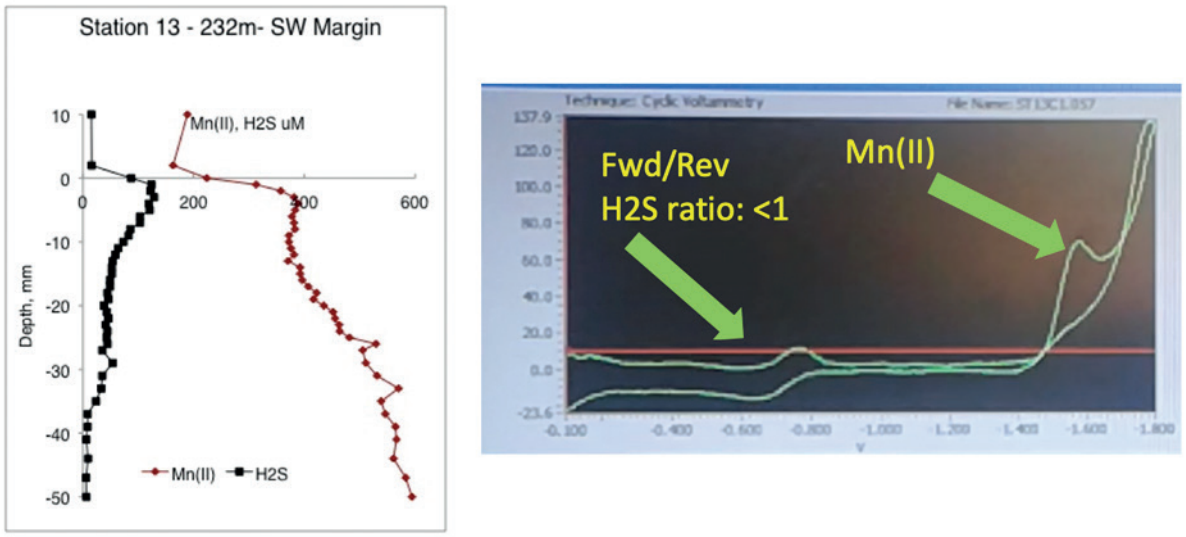


Figure 6. Black Sea, Station 13 sediment core micro-profiles, indicating high levels of Mn(II) and strong evidence for a 'bound' nature of the total sulfide pool

In conclusion, a comprehensive examination of the microelectrode voltammetry data in the Baltic Sea chemocline revealed that deep Fe flux leads to FeS and that NO_3-driven sulfide oxidation produces accumulation of S_x^{2-}, all of which are represented in the patterns of variations in the forward/reverse scan sulfide peak/wave heights. Deep sediment voltametric scans in the Black Sea show a large fraction of bound sulfide without any clear peaks for sulfur intermediates. Some unidentified role exists for organic sulfide species and/or FeS colloids/nanoparticles (which are too large to be electroactive), particularly at the seafloor of the Black Sea suboxic zone. As a result, more research is currently underway, such as the ERC Consolidator Grant DeepTrace (PI: M Yücel, 20232028) to uncover this "black box" of nanoparticulate species.

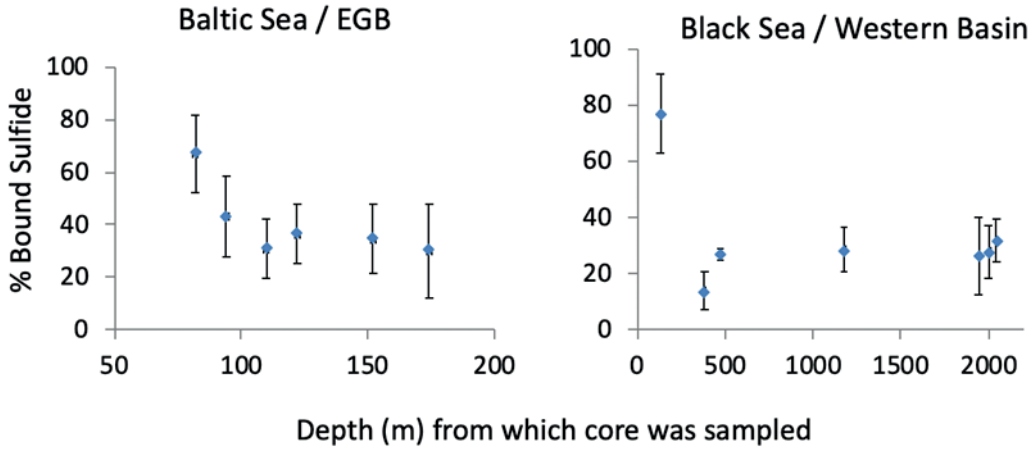


Figure 7. Comparison of 'bound sulfide' estimations in the Baltic and Black Sea sediments, based on microelectrode voltammetry.

Case 3: Sulfide Micro-Gradients in the East Pacific Rise Hydrothermal Vents

In the hydrothermal ecosystem, the availability of free sulfide for biological use is determined both by the rate of abiotic/biotic oxidation of sulfide in the environment and its chemical form inside the host organism (Luther *et al*. 2001b) or surrounding free-living communities. Since the discovery of sulfide-based hydrothermal habitats in 1979 (Corliss *et al*. 1979), research on sulfur speciation has been at the forefront due to the possibility that a rapid abiotic oxidation of sulfide by oxygen will hinder its biological usage.

Luther *et al.* (2001) initially demonstrated the substantial shift in sulfur speciation in various vent habitats using in-situ voltammetry obtained from a submersible or remotely operated vehicle. Here, the tubeworm *Riftia* fields are abundant in "free" sulfide, frequently in conjunction with dissolved oxygen, and have high forward to reverse sulfide signal ratios. It has been demonstrated that the fluid chemistry of the *Riftia* or *Tevnia* type tubeworm habitats is characteristic for the chemosymbionts of these organisms, which feed on both oxygen and sulfide (also see Nees *et al.* 2008; Luther *et al*. 2012; Le Bris *et al*. 2019). By pumping in surrounding seawater and precipitating sulfide as FeS and other metal sulfides, the filter-feeder worm *Alvinella*, which is most likely heterotrophic, colonizes the highest temperature sites and requires detoxification of sulfide. The ratios of the forward to reverse sulfide signals in this instance become substantially less than 1, which is comparable to the previously mentioned suboxic zone scans of the Black Sea. A minor "free" sulfide signal was identified in the forward scan for both FeS and Fe(II), but a sizable acid-volatile sulfide signal was detected in the return wave.

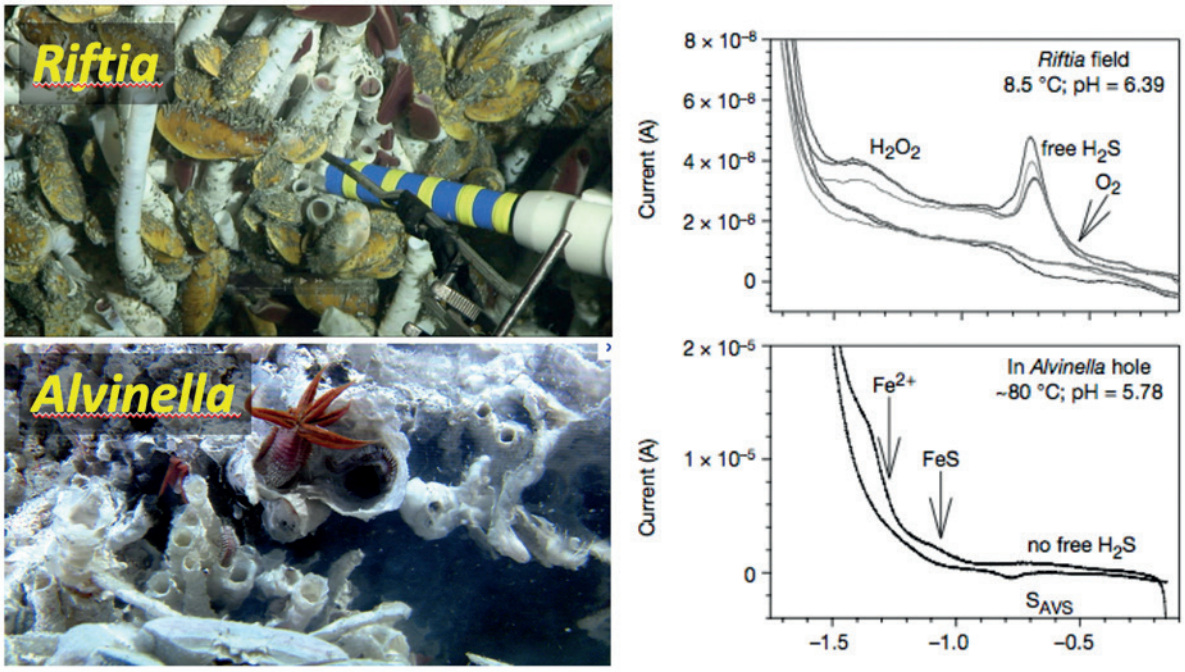


Figure 8. Voltammograms from *Riftia* and *Alvinella* habitats from EPR 9N hydrothermal fields. The photos are from the 2017 EPR expedition (Chief Scientist GW Luther, UDEL). The scans on the right side are adapted from Luther *et al.* 2001.

The co-existence of sulfide in vent mixing fluids with metals and oxygen is essential for an effective kinetic setup that benefits chemosynthetic microorganisms. With a half-life of 55 days

at pH 12, the abiotic oxidation of sulfide in trace metal clean solutions is slow (Luther *et al*. 2011. However, the kinetics of the process are accelerated by transition metals. Fe(II), Pb(II), Cu(II), Fe(III), Cd(II), Ni(II), Co(II), and Mn(II) are among the metals whose effects have been studied at different concentrations at pH 8 for seawater at 25 ºC (Vazquez *et al*., 1989). Their findings indicate that Fe(II) is 20 times more effective than the other metals at catalyzing sulfide oxidation. Because iron concentrations can surpass 100 μM and pH can drop below neutral values in some diffuse flow microenvironments, iron catalysis of the sulfide oxidation process is crucial for the hydrothermal environment (Le Bris *et al*. 2003, 2006). Future integrated studies of both shallow and deep hydrothermal systems could greatly benefit from the use of *in situ* voltammetry due to its direct ability to document sulfide, iron, and oxygen in a single scan as well as its indirect potential to estimate sulfide speciation *in situ* (using the forward and reverse scans in the cyclic voltametric mode). This could reveal the intricate nature of the micro-gradients that support extreme life in those habitats.

Conclusions

Microelectrode cyclic voltammetry allows for the collection of large quantities of data, which can be used to direct research on the cycling of cryptic sulfur across natural redox gradients in seafloor habitats. We have demonstrated that this approach is not only beneficial for detecting selected analytes, but it can also offer more information on processes involving sulfur intermediates if a comparative analysis of reverse and forward scans of cyclic voltammograms is performed. Although the analytical precision and accuracy of cyclic voltammetry may be limited relative to many alternative methodologies, significant gains can be made in response times, frequency, and resolution. This, in turn, supports autonomous deployment and integration with other platforms. The data generated is also suitable for a novel, emergent analysis strategy based on big data and visualization, which are showcased in different chapters of this monograph.

Acknowledgements

The author is grateful to Prof. Laura Giuliano and the CIESM team for an inspiring workshop on ‘Ocean of Gradients’. The dataset and samples that supported the synthesis of this contribution has been collected in research expeditions with RV Pelagia (2013, Chief Scientist: Caroline Slomp), RV Alkor (2013 and 2014, Chief Scientists: Olaf Pfankuche and Stefan Sommer), RV Atlantis (2007, 2008, 2017 Chief Scientist: G Luther). The author is thankful to the science parties and crew of all these expeditions. The drafting of this chapter was supported by the EU Horizon Europe European Research Council Consolidator Grant (ERC-CoG) DeepTrace Project (Grant agreement No.101043381) Turkish Academy of Sciences (TÜBA).

References

Aller, R.C. and Dude, P.D. (1988). Complete oxidation of solid phase sulfides by manganese and bacteria in anoxic marine sediments. *Geochimica et Cosmochimica Acta*, 52: 751–765.

Brendel, P. and Luther, G.W. (1995). Development of a gold-amalgam voltametric microelectrode for the determination of dissolved Fe, Mn, O2 and S(-2) in porewaters of marine and freshwater sediments. *Environmental Science & Technology*, 29 (3): 751–761.

Corliss, J.B., Dymond, J., Gordon, L.I., *et al*. (1979). Submarine Thermal Springs on the Galapagos Rift. *Science*, 203 (4385): 1073–1083.

Gartman, A., Yucel, M., Madison, A.S., *et al*. (2010). Sulfide oxidation across diffuse flow zones of hydrothermal vents. *Aquatic Geochemistry*, 17 (4): 583–601.

Johnson, K.S., Childress, J.J., Hessler, R.R., *et al*. (1988). Chemical and biological interactions in the Rose Garden hydrothermal vent field, Galapagos spreading center. *Deep-Sea Research Part I: Oceanographic Research Papers*, 35 (10-11): 1723–1744.

Jorgensen, B.B., Findlay, A.J., Pellerin, A. (2019). The Biogeochemical Sulfur Cycle of Marine Sediments. *Frontiers in Microbiology*, 10: 849.

Kraal, P., Yücel, M., Slomp, C. (2019). Geochemical alteration of turbidites in the southwestern Black Sea: implications for biogeochemical cycling in an anoxic basin. *Marine Chemistry*, 209: 48–61.

Le Bris, N., Yücel, M., Das, A., *et al*. (2019). Hydrothermal energy transfer and organic carbon production on the deep-seafloor. *Frontiers in Marine Science*, 5: 531.

Luther, G.W., Glazer, B.T., Hohmann, L., *et al*. (2001a). Sulfur speciation monitored *in situ* with solid state gold amalgam voltametric microelectrodes: polysulfides as a special case in sediments, microbial mats and hydrothermal vent waters. *Journal of Environmental Monitoring*, 3 (1): 61–66.

Luther, G.W., Reimers, C.E., Nuzzio, D.B., *et al*. (1999). In situ deployment of voltametric, potentiometric and amperometric microelectrodes from a ROV to determine O2, Mn, Fe, S(2-) and pH in porewaters. *Environmental Science & Technology*, 33 (22): 4352–4356.

Luther, G.W., Rozan, T.F., Taillefert, M., *et al*. (2001b). Chemical speciation drives hydrothermal vent ecology. *Nature*, 410 (6829): 813–816.

Luther, G.W., Findlay, A.J., MacDonald, D.J., *et al*. (2011). Thermodynamics and kinetics of sulfide oxidation by oxygen: a look at inorganically controlled reactions and biologically mediated processes in the environment. *Frontiers in Microbiology*, 2: 62.

Luther, G.W., Glazer, B.T., Ma, S., *et al*. (2008). Use of voltametric solid-state (micro)electrodes for studying biogeochemical processes: laboratory measurements to real time measurements with an *in situ* electrochemical analyzer (ISEA). *Marine Chemistry*, 108 (3-4): 221–235.

Luther, G.W., Gartman, A., Yücel, M., *et al*. (2012). Chemistry, temperature and faunal distributions at diffuse flow hydrothermal vents: comparisons of two geologically distinct ridge systems. *Oceanography*, 25 (1): 234–245.

Mele, B.H., Monticelli, M., Leone, S., *et al*. (2023). Oxidoreductases and metal functioning in the functioning of Earth. *Essays in Biochemistry*, 67 (4): 653–670.

Nees, H.A., Moore, T.S., Mullaugh, K.M., *et al*. (2008). Hydrothermal vent mussel habitat chemistry, pre-and post-eruption at 9°50'North on the East Pacific Rise. *Journal of Shellfish*

Research, 27 (1): 169–175.

Preisler, A., de Beer, D., Lichtschlag, A., *et al*. (2007). Biological and chemical sulfide oxidation in a *Beggiatoa* inhabited marine sediment. *The ISME Journal*, 1 (4): 341–353.

Sievert, S.M., Hügler, M., Taylor, C.D., *et al*. (2008). Sulfur oxidation at deep-sea hydrothermal vents. In: Dahl, C. and Friedrichs, C.G. (Eds.), *Microbial Sulfur Metabolism*. Springer: Berlin, pp. 247–273.

Vazquez, F.G., Rey, P.G., Bermejo, M.P. (1989). Chemical speciation of trace metals in seawater. *Marine Chemistry*, 28(4), 315–325.

Yücel, M. (2013). Down the thermodynamic ladder: A comparative study of marine redox gradients across diverse sedimentary environments. *Estuarine, Coastal and Shelf Science*, 131: 83–92.

Yücel, M., O'Connell, M., Luther, G.W. (2013). Iron sulfide formation in the presence H2S: An electrochemical approach to measure the rate of sulfide oxidation. *Environmental Science & Technology*, 47(1), 329-335.

Yücel, M., Sommer, S., Dale, A.W., *et al*. (2017). Microbial sulfide filter along a benthic redox gradient in the Eastern Gotland Basin, Baltic Sea. *Frontiers in Microbiology*, 8: 169.

Zopfi, J., Bottcher, M.E., Jorgensen, B.B. (2008). Biogeochemistry of sulfur and iron in Thioplaca-colonized surface sediments in the upwelling area off central Chile. *Geochimica et Cosmochimica Acta*, 72 (3): 827–843.

WORKSHOP COMMUNICATIONS

C) EXTREME MARINE AND HUMAN MICROBIOTA

Microbial Metabolic Diversity and Pressure Adaptation at Deep-Sea Hydrothermal Vents

Mohamed Jebbar

Univ Brest, CNRS, Ifremer, EMR 6002 BIOMEX, Unité Biologie et Écologie des Écosystèmes marins Profonds BEEP, Plouzane, France

Abstract

Hydrothermal vents form when seawater penetrates the oceanic crust, becomes superheated by magma, and is expelled back into the ocean. The resulting fluids give rise to vent structures like black smokers (high temperature) and white smokers (lower temperature). Discovered in 1977, these deep-sea systems support ecosystems driven by chemosynthesis rather than photosynthesis, relying on microbes that use inorganic compounds for energy. Microbial communities in hydrothermal vents are metabolically diverse, including autotrophs, heterotrophs, methanogens, sulfur oxidizers, and iron reducers. Hydrostatic pressure, which increases with depth, plays a crucial role in shaping microbial physiology, influencing metabolic pathways and cellular processes. Pressure-adapted microbes, known as piezophiles, thrive under conditions that mimic their deep-sea environments. While both culture-based and molecular methods have advanced our understanding of these extremophiles, high-resolution biogeochemical mapping linked to active microbial processes is still lacking. Visual databases with images and videos offer valuable insight into the complex terrain and biodiversity of these unique ecosystems.

Keywords: hydrothermal vents, chemosynthesis, extremophiles, hydrostatic pressure, microbial diversity, seafloor geology

Introduction

From a geological perspective, hydrothermal vents are associated with seafloor spreading zones, located along mid-ocean ridges—regions that lie at the boundaries of tectonic plates. These ridges, which extend continuously across the global ocean floor, form the longest mountain range on Earth, with an estimated total length of 60,000 to 80,000 kilometers.

Driven by plate tectonics and active volcanism, seawater penetrates fractures in the oceanic crust, becomes superheated as it approaches underlying magma, leaches minerals from the surrounding rocks, and is eventually expelled back into the ocean under the combined influence of high temperatures and hydrostatic pressure. These underwater hot springs are known as hydrothermal vents.

When the superheated hydrothermal fluid (typically 350–450 °C) rapidly mixes with the near-freezing ambient seawater (2–4 °C), the dissolved minerals in the fluid precipitate out. These minerals accumulate around the vent orifice, giving rise to hydrothermal chimneys—partially porous mineral structures whose composition and morphology evolve over time.

The temperature of the hydrothermal fluid, together with its chemical composition, determines the sequence in which various minerals precipitate to form the vent structures. Chimneys are classified as “black smokers” when the fluid is emitted at high temperatures (above 300 °C).

These emit dark, mineral-rich plumes due to high concentrations of metal sulfides. In contrast, "white smokers" form when hydrothermal fluids are partially diluted with cold seawater before being expelled at lower temperatures (typically between 150 and 290 °C). These vents release lighter-colored fluids with lower mineral content and reduced flow rates. When the hydrothermal fluid is extensively diluted by seawater before emerging from the seafloor, the temperature may drop below the threshold required for mineral precipitation. In such cases, no chimney structures form, and the fluid is discharged diffusely through the seabed—this is referred to as a zone of diffuse fluid emission.

Deep-sea hydrothermal vents were discovered in 1977 (Corliss *et al.* 1979) and revealed ecosystems whose existence depends on bacterial chemosynthesis rather than photosynthesis. These ecosystems support thriving communities that rely on the metabolic activity of microorganisms surrounding the vents. The discovery of hydrothermal vents in various deep-sea regions significantly expanded our understanding of the diversity and extent of life on Earth.

Hydrothermal vents are found in oceans and seas around the world, from relatively shallow depths to some of the deepest known, such as those in the Cayman Trough at 4,960 meters (Connely *et al*. 2012).

There are several compelling reasons why scientists study hydrothermal vents. First, the deep sea remains one of the least explored regions of our planet—there are more detailed maps of the Moon and Mars than of Earth's ocean floor, largely because satellites cannot penetrate water. Researchers are eager to understand what organisms live in these environments and how they survive under such extreme conditions.

Another motivation lies in the potential these ecosystems hold for shedding light on the origins of life. Some scientists propose that hydrothermal vents may have been the cradle of life on Earth, as some of the oldest known organisms are hyperthermophiles—microbes that thrive at temperatures above 80°C. This hypothesis, while intriguing, remains a subject of ongoing debate and scientific scrutiny.

Moreover, the organisms inhabiting hydrothermal vents—particularly those in the deepest regions of the ocean—are remarkable for their ability to thrive in chemical environments that would be toxic to most other life forms on Earth. These extremophiles are uniquely adapted to withstand extreme hydrostatic pressure, high temperatures, and chemically harsh conditions. As a result, they are of significant interest not only for fundamental scientific research but also for applied sciences. The biomolecules and enzymes they produce, fine-tuned for survival under such extremes, hold great promise for a range of biotechnological applications.

Phylogenetic and Metabolic Microbial Diversity

Numerous microbiological studies employing culture-independent approaches—such as metabarcoding and metagenomics—have been conducted on various deep-sea hydrothermal vent sites, revealing remarkably diverse and complex microbial communities. Metagenome-assembled genome (MAG) analyses have shown that bacterial diversity spans 29 phyla, while archaeal diversity encompasses 19 phyla (Zhou *et al.* 2022). In contrast, culture-based methods have so far enabled the enrichment and isolation of only about a hundred species, belonging to just 2 archaeal phyla and 11 bacterial phyla (Zeng *et al.* 2021) (Figure 1). Most of these isolates are thermophilic or hyperthermophilic, primarily anaerobic, although some mesophilic and mostly aerobic species have also been recovered. These microbial communities exhibit a wide range of metabolic strategies, including autotrophy and heterotrophy, and encompass functional

groups such as methanogens, methanotrophs, sulfur oxidizers, sulfur reducers, sulfate reducers, denitrifiers, iron oxidizers, iron reducers, and hydrogen oxidizers (*ibid.*).
Functional and quantitative studies are predominantly conducted in laboratory settings, often using microbial isolates. However, some research efforts have also focused on measuring specific metabolic activities *in situ* in order to better understand ecosystem functioning and identify key microbial players. Establishing a direct and accurate link between biogeochemical cycles in the deep sea and the actual contribution of the microbial communities at hydrothermal vents remains a significant challenge. One of the major and pervasive environmental factors in deep-sea hydrothermal systems is hydrostatic pressure, which exerts a strong influence on physico-chemical equilibria and, consequently, on microbial physiology and metabolism. Despite this, most laboratory studies are performed at atmospheric pressure, even when working with microorganisms isolated from high-pressure environments. As a result, the insights gained from these experiments may be imprecise or even misleading when applied to models of deep-sea hydrothermal ecosystem functioning.

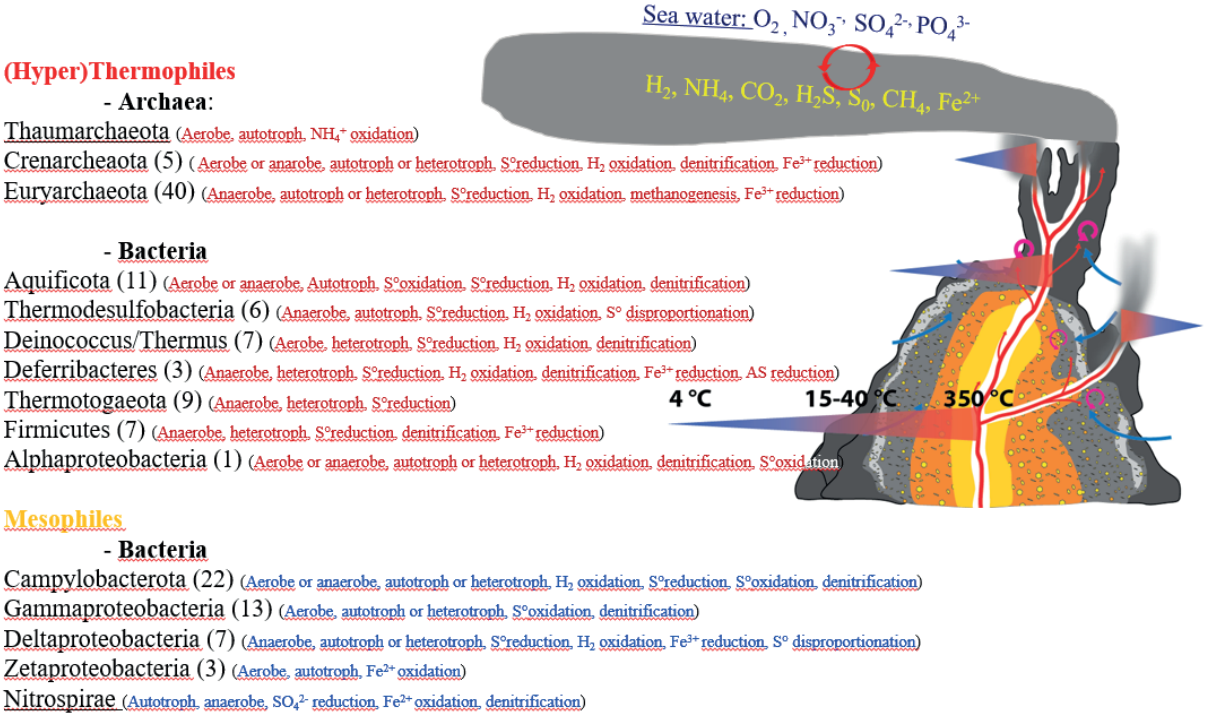


Figure 1. Overview of the microbial metabolisms of prokaryotes isolated from deep sea hydrothermal vent field samples adapted from Zeng *et al.* 2021.

Hydrostatic Pressure Adaptation

Hydrostatic pressure is the force exerted by a column of water on a given surface. In the ocean, it increases by approximately 0.1 MPa for every 10 meters of depth, meaning that at a depth of 2,500 meters, the pressure reaches about 25 MPa.

Hydrostatic pressure is commonly expressed in three units: atmospheres (atm), bars (bar), and pascals (Pa). The atmosphere and the bar are nearly equivalent, with 1 atm equal to 1.01325 bar. The pascal is the SI (International System of Units) reference unit, where 1 Pa equals 10^{-5} bar.

For practical purposes in deep-sea studies, pressure is often expressed in megapascals (MPa), with 1 MPa equal to one million pascals.

The physical effects of high hydrostatic pressure (HHP) were first described by Le Chatelier in 1884. He formulated the principle that a system subjected to a change in an external condition—such as temperature, pressure, or volume—will adjust in a way that counteracts the imposed change. Specifically, when hydrostatic pressure increases, the system responds by minimizing its volume. This pressure-induced volume change, denoted as ΔV (the difference between the final and initial volumes of the system), is governed by the thermodynamic relationship outlined in Silva *et al.* (2013), as shown in the equation below.

$$\Delta V = \Delta V^0 + \Delta\alpha(T - T_0) - \Delta\beta(P - P_0)$$

In this equation, ΔV^0, $\Delta\alpha$, $\Delta\beta$, P, and T are, respectively, the variation in volume under standard conditions, the thermal expansion factor, the compressibility factor, the pressure, and the temperature. When the temperature is constant, only pressure and volume are included in the reaction. The compressibility factor therefore plays an important role, particularly as the volume of biomolecules (*e.g.,* proteins, lipids) can vary significantly. Similarly, various experiments have shown that if a reaction generating a decrease in volume were subjected to HHP, its activity would be greatly increased (Smeller 2002).

Microorganisms inhabiting the deep sea—including hydrothermal vent environments located below 2,000 meters—often exhibit optimal growth at hydrostatic pressures that approximate *in situ* conditions and far exceed atmospheric pressure. These pressure-adapted organisms are referred to as piezophiles (from the Greek *piezein*, meaning "to press," and *philos*, meaning "loving").

Although hydrostatic pressure is a pervasive environmental factor in the deep ocean, it is frequently overlooked or underestimated in microbiological studies. Nevertheless, research on a limited number of piezophilic microorganisms—primarily Bacteria (notably Gammaproteobacteria) and Archaea (mainly Euryarchaeota)—has provided valuable insights. These organisms have been isolated and studied across multiple levels of analysis, including omics (genomics, transcriptomics, proteomics), physiology, metabolism, phylogeny, biochemistry, and biophysics, in efforts to unravel the mechanisms underlying adaptation to high hydrostatic pressure (Jebbar *et al*. 2015).

Most studies on piezophilic microorganisms have demonstrated that hydrostatic pressure influences the activity of various metabolic pathways and key cellular processes. These findings highlight the importance of incorporating hydrostatic pressure as a parameter in statistical models that aim to simulate the functioning of deep-sea ecosystems. This consideration is increasingly being integrated into underwater observation initiatives, which enable the direct monitoring and characterization of functional microbial diversity, among other ecological and environmental aspects.

Conclusions and Prospects

This is a concise overview of the metabolic and taxonomic diversity of prokaryotes inhabiting deep-sea hydrothermal vents. Both culture-dependent and culture-independent approaches have been employed to investigate the diversity of bacteria and archaea that have adapted to and thrive in these extreme environments, characterized by their instability, steep physicochemical gradients, and pronounced environmental contrasts. High-resolution images and videos of the

explored ecosystems are available through dedicated websites and databases, offering visual insight into the topography and associated fauna. These are complemented by geochemical measurements and *ex situ* analyses, which together contribute to the characterization of microbial diversity. In my view, substantial work remains before we can achieve high-resolution 3D mapping and visualization of marine biogeochemical data from hydrothermal vent systems, particularly data that can be correlated with the metabolically active fraction of the prokaryotic community.

References

Connelly, D.P., Copley, J.T., Murton, B.J., *et al*. (2012). Hydrothermal Vent Fields and Chemosynthetic Biota on the World's Deepest Seafloor Spreading Centre. *Nature Communications*, 3 (1): 620.

Corliss, J.B., Dymond, J., Gordon, L.I., *et al*. (1979). Submarine Thermal Springs on the Galápagos Rift. *Science*, 203 (4385): 1073–1083.

Jebbar, M., Franzetti, B., Girard, E., *et al*. (2015). Microbial Diversity and Adaptation to High Hydrostatic Pressure in Deep-Sea Hydrothermal Vents Prokaryotes. *Extremophiles*, 19 (4): 721–740.

Silva, J.L., Oliveira, A.C., Vieira, T.C., *et al*. (2014). High-Pressure Chemical Biology and Biotechnology. *Chemical Reviews*, 114 (14): 7239–7267.

Smeller, L. (2002). Pressure–Temperature Phase Diagrams of Biomolecules. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, 1595 (1–2): 11–29.

Zeng, X., Alain, K., Shao, Z. (2021). Microorganisms from Deep-Sea Hydrothermal Vents. *Marine Life Science & Technology*, 3 (2): 204–230.

Zhou, Z., St. John, E., Anantharaman, K., *et al*. (2022). Global Patterns of Diversity and Metabolism of Microbial Communities in Deep-Sea Hydrothermal Vent Deposits. *Microbiome*, 10 (1): 241.

Deep Lakes Saturated with $MgCl_2$ Located at the Bottom of the Mediterranean Ridge, as Unique Abiotic Hydrological Formations

Violetta La Cono[1], Laura Giuliano[2] and Michail M. Yakimov[1]

[1]Centro Nazionale di Ricerca (CNR), Institute for Polar Sciences, Messina, Italy
[2] Mediterranean Science Commission (CIESM), Monaco

Abstract

The recent discovery of magnesium-rich evaporite minerals and subglacial brines on Mars is stimulating interest in exploring the potential habitability of terrestrial Mg^{2+}-rich environments. The deep-sea, magnesium-rich Lakes *Discovery*, *Hephaestus* and *Kryos*, located at the bottom of the Mediterranean Ridge, are among the most challenging, extreme and unusual environments on planet Earth. Moreover, they seem to be only known terrestrial areas, the ecological conditions of which (high pressure, anoxia, athalassohaline hydrochemistry, rich in magnesium, excessive chaotropicity and extremely low water activity) can be considered as the most closed analogues of Martian subglacial lakes. From the point of view of 3D images of ocean gradients, Lakes *Discovery*, *Hephaestus* and *Kryos* are of extraordinary attraction and interest. Due to their elevated density (> 1.30), they form stable and long-lived (at least 2-3 thousand years) hydrological formations, separated from the overlying 3- kilometer layer of seawater in the form of lakes, easily identified by swath-bathymetry methods of observation. Despite *Discovery*, *Hephaestus* and *Kryos* are anathemas for living beings, their brine- seawater interfaces are inhabited by highly stratified, diverse and enigmatic members of prokaryotes, adapted to withstand above mentioned polyextremophily. Being completely different from seawater microbes, these obligate anaerobic and extremely halophilic communities are strikingly similar to each other in all these three lakes. We believe, therefore, that they originated, in part at least, from an ancient late-stage evaporitic suite and have existed in either dormant or metabolically active state in brine lenses that were trapped in subsurface sediments for an indefinitely long time.

Keywords: Deep-sea Hypersaline Anoxic Lakes (DHALs), Messinian Salinity Crisis, Mediterranean Ridge, hypersaline brines, microbial communities, extremophiles, water activity

Introduction

During the late Miocene epoch (5.33–5.96 million years ago), several repeated desiccations and re-fillings of the Mediterranean basin, named Messinian salinity crisis, resulted in the formation of enormous deposits of layered evaporites that attain a thickness of up to 3.5 km in some places of eastern Mediterranean (Cita 2006), where continuous subduction of the African plate under the Eurasian and Anatolian plate formed the Mediterranean Ridge accretionary complex. The deformational tectonic activity of the Mediterranean Ridge, accompanied by the presence of huge subsurface salt deposits, appears to control the creation in this area of peculiar large-scale submarine hydrological formations within confined depressions, named deep-sea hypersaline anoxic lakes (DHALs). Currently, nine such lakes, *L'Atalante*, *Bannock*, *Discovery*, *Hephaestus*, *Kryos*, *Medee*, *Thetis*, *Tyro* and *Urania* have been discovered and studied in the deep eastern Mediterranean in the last 40 years (De Lange and Ten Haven

1983; MEDRIFF Consortium 1995; Wallmann *et al*. 1997, 2002; Chamot-Rooke *et al*. 2005; La Cono *et al*. 2011, 2019; Yakimov *et al*. 2013, 2015) (Figure 1). The surfaces of these brine lakes lie between 3.0 and 3.5 km below sea level, and the salinity of their brines ranges from five to 13 times higher than that of seawater. As it generally accepted, the deliquescence of subterranean evaporitic deposits, tectonically exposed to seabed surface is the main reason of their formation (Vengosh *et al*. 1994). Although these DHALs lie geographically close to each other (Figure 1), their hydrochemical diversity suggests that the processes leading to their formation were qualitatively different and can be unambiguously explained by the salt deposition dynamics during desiccation/re-flooding cycles occurred in the past.

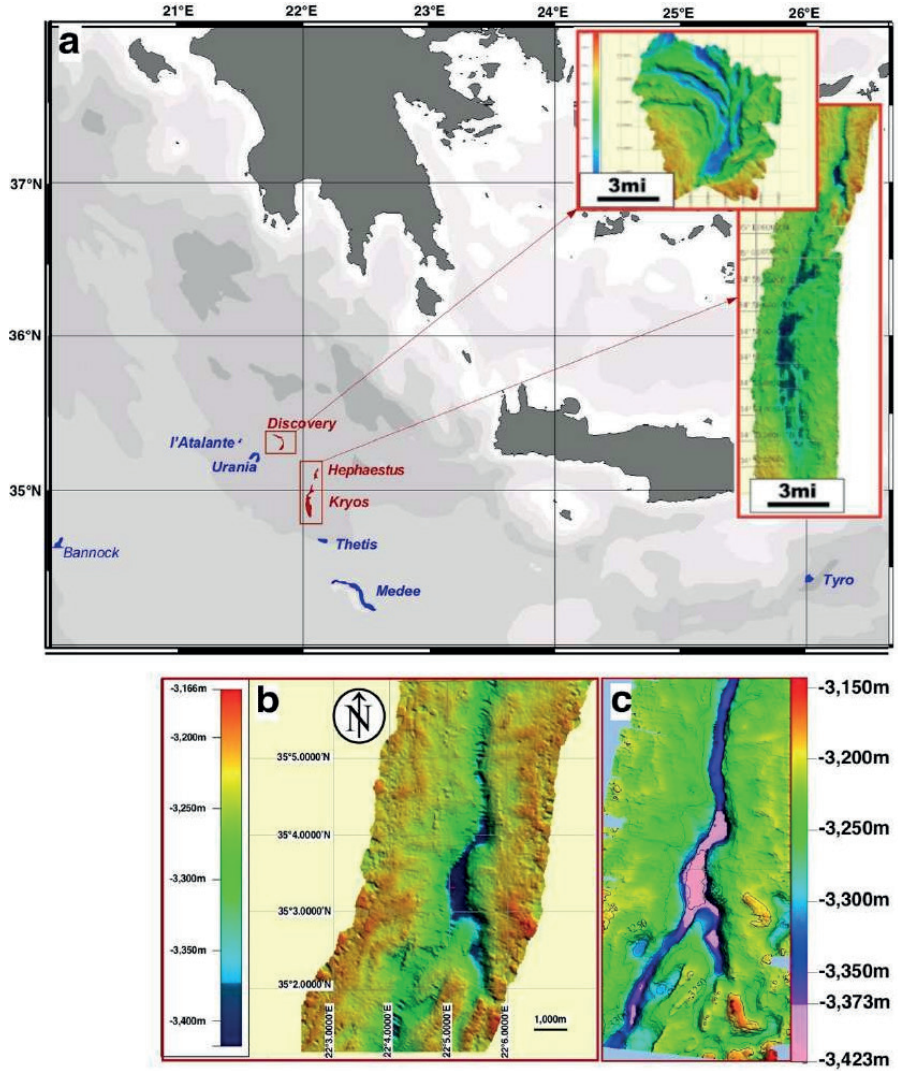


Figure 1. Locations of all nine large-scale (up to 40 km long and 3.5 km wide) deep-sea brine lakes currently known to exist on the floor of the Mediterranean Ridge. Lakes filled with $MgCl_2$ are shown in red and in the insets at the top right while thalassohaline (NaCl-dominated) lakes are shown in blue (a). High 10-m resolution swath-bathymetry map of the lastly recently discovered Lake *Hephaestus* demonstrates its location in a very narrow steep NE depression (b). Exact positioning of the surface of the lake (seawater : brine interface) was confirmed by multiparameter conductivity-temperature-dissolved oxygen (CTD) cast at the depth of 3,373 m and exact shape of Lake *Hephaestus* was obtained by swath bathymetry with 5-m resolution, shown in pink (c). Multibeam swath bathymetry was obtained by echo-sounding and processed with NEPTUNE, CARIS and GMT packages (Wessel *et al.* 2013).

Seawater can be evaporated 10-fold without salt precipitation, resulting in formation of brine with salinity ≤ 330 practical salinity units (PSU). This early-stage primary brine remains thalassohaline, *i.e.,* it has proportions of all major ions, characteristic to that of seawater. When the evaporation of primary brine continues (Figure 2a-c), salinity increases and the salts begin to precipitate, starting with calcium minerals, followed by precipitation of halite (NaCl), kieserite ($MgSO_4 \cdot KCl \cdot 3H_2O$), carnallite ($KMgCl_3 \cdot H_2O$), kainite ($MgSO_4 \cdot KCl \cdot 3H_2O$) and ending with formation of bischofite mineral ($MgCl_2 \cdot 6H_2O$), which is the most soluble of all marine evaporitic salts (De Lange *et al*. 1990; Wallmann *et al*. 1997; Cita 2006). Interruption of the evaporitic path at the stage of halite precipitation by flooding, followed by countless repetition of desiccation/re-flooding cycles, would cause massive formation of halite-rich evaporitic deposits. Dissolution of such evaporites would originate the creation of NaCl-pure brine, which is likely the case of Lake *Tyro* (Cita 2006).

Results

Origin of the Mediterranean Ridge DHALs

Among all Mediterranean DHALs explored so far, only the *Discovery, Hephaestus* and *Kryos* lakes are filled with near-saturated $MgCl_2$-brines (4.38-5.05 M), suggesting that they derived via dissolution of nearly pure bischofite, which should be located in the uppermost layer of the evaporitic suite (Table 1). The brine of these lakes contains approx. 8 g kg^{-1} Br (Wallmann *et al.* 1997, 2002; Yakimov *et al.* 2015; La Cono *et al.* 2019), the concentration, typically found in bischofite which precipitates from the latest stage brine containing higher Br concentration (12 g kg^{-1}) that occurs when seawater is evaporated to less than 0.5-1% of its initial volume.

	MedSea water	Lake Discovery	Lake Hephaestus	Lake Kryos
Major ions, mmol kg $^{-1}$*				
Cl^-	630	**10,150**	**9,120**	**9,043**
$SO4^{2-}$	33	110	203	320
Br^-	1	110	78	70
Na^+	540	84	93	125
K^+	12	20	28	80
Mg^{2+}	61	**5,150**	**4,720**	**4,380**
Ca^{2+}	12	1	2	1
Main physicochemical parameters				
Surface location, mbsl	NA	3,580	3,373	3,337
Maximum Depth, m	NA	50	50	160
Surface area, km^2	NA	7.5	1.7	25
Water activity, a_w	0.980	**0.382**	**0.395**	**0.399**
Stable isotopes in brine				
$\delta^{18}O$, ‰	0.97 ± 0.68	−2.39 ± 0.79	−3.06 ± 0.22	−2.54 ± 0.64
δD, ‰	5.1 ± 5.7	−17.9 ± 1.5	−16.5 ± 0.4	−15.8 ± 0.8

Table 1. Major ion composition and other parameters of $MgCl_2$-saturated, deep-sea brines of the lakes *Discovery*, *Hephaestus* and *Kryos* located at the floor of the Mediterranean Ridge.* due to elevated density (> 1.30) the mmol values are given in kg^{-1}, not in $liter^{-1}$. NA, not applicable.

Thus, the existence of the *Discovery, Hephaestus* and *Kryos* brines reinforces the evidence that the Eastern Mediterranean basin evaporated close to dryness during the Messinian salinity crisis (Wallmann *et al*. 1997; Cita 2006). Plotting the average of DHALs brine composition over the evaporation path of seawater can easily explain their origin, namely the deliquescence of which strata of the evaporitic suite, resembling a layer cake, caused the formation of each of these milieux (Figure 2 a-c).

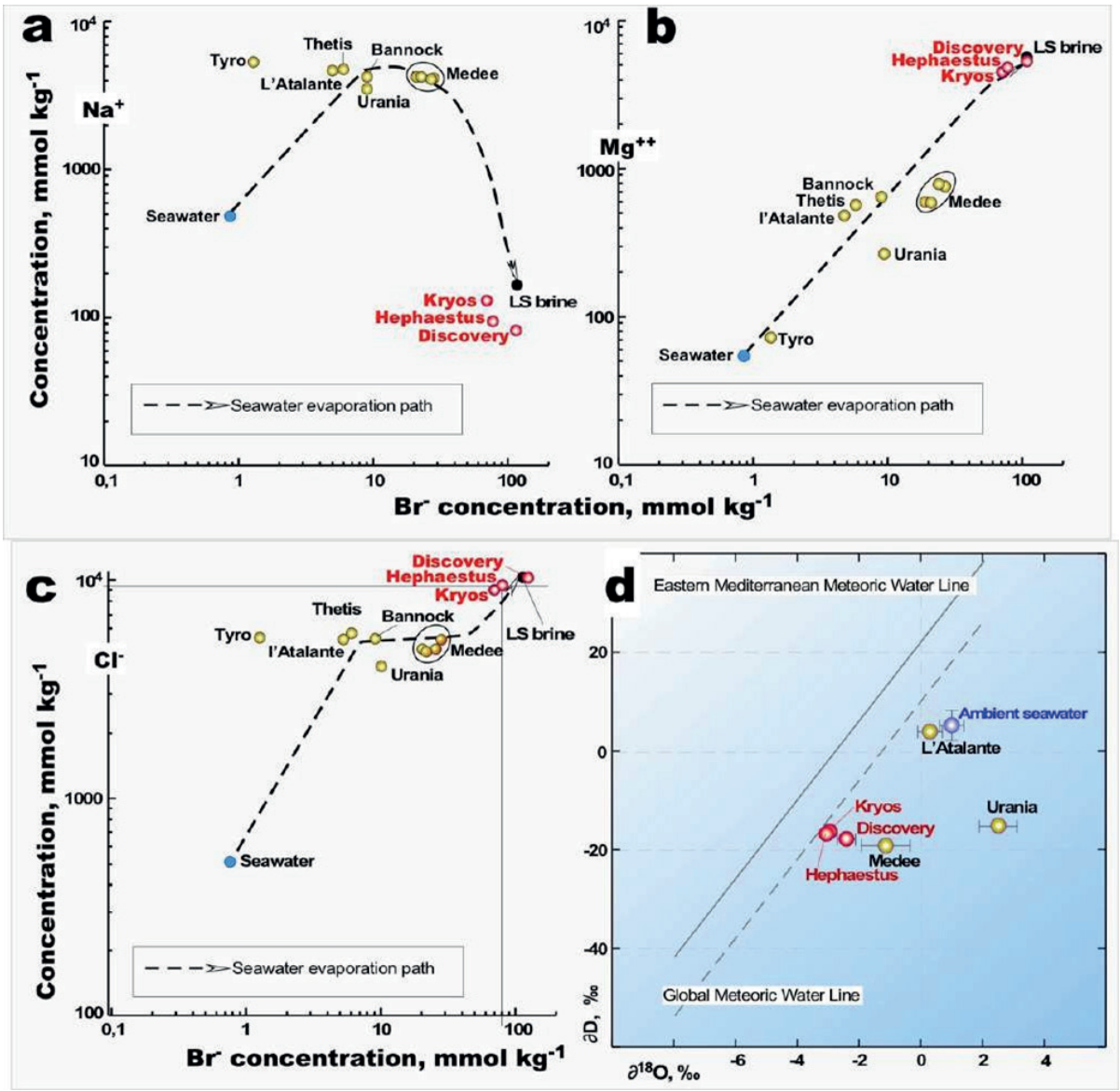


Figure 2. Evaporation path of seawater in the Na^+ vs. Br^- (a); Mg^{2+} vs. Br^- (b) and Cl^- vs. Br^- plots (c). Data are after De Lange et al. (1990), Yakimov et al. (2015) and La Cono et al. (2019). Average brine compositions have been indicated as yellow bubbles for thalassohaline DHALs and as red bubbles for $MgCl_2$-filled athalassohaline brines *Discovery* Hephaestus and *Kryos*. LS brine is the latest stage brine occurred when seawater is evaporated to less than 1% of its volume. (d) Plot of δD versus $\delta^{18}O$ for brine samples collected in DHALs *Discovery*, *Hephaestus*, *Kryos*, *l'Atalante*, *Medee* and *Urania*. Eastern Mediterranean seawater values, measured in the seawater overlaying the lakes, are shown as blue bubble. The solid and dashed lines represent Eastern Mediterranean Water Line and Global Meteoric Water Lines, respectively.

As mentioned in the Introduction Section, most deep-sea hypersaline lakes are proposed to have originated from the dissolution of subsurface evaporitic layer(s) after they were (tectonically) exposed to seawater (Cita 2006). Analysis of stable isotopes of $\delta^{18}O\text{-}H_2O$ and $\delta^{2}H\text{-}H_2O$ in DHAL brines helped us to understand the origin of these lakes. We used this approach to analyze six Mediterranean brine lakes (*L'Atalante*, *Discovery*, *Hephaestus*, *Kryos*, *Medee* and *Urania*). Apparently, such a scenario is applicable only to some but not to all Mediterranean DHALs. As shown in Figure 2d, analyzed DHALs can be separated in two principal groups: 1) brines with the $\delta^{18}O\text{-}H_2O$ / $\delta^{2}H\text{-}H_2O$ values similar to that of modern seawater (*L'Atalante*), 2) brines with $\delta^{18}O\text{-}H_2O$ / $\delta^{2}H\text{-}H_2O$ values other than seawater (all other analyzed brines). Thus, it seems only DHAL *L'Atalante* exhibits 'seawater-dissolution' scenario, while brines of other DHALs originated from more ancient subterranean water masses. The $\delta^{18}O$ / $\delta^{2}H$ isotopic ratios in *Discovery*, *Hephaestus*, *Kryos* and *Medee* brines are quite similar to each other and both isotopes are depleted relative to those of Mediterranean seawater and *L'Atalante*. Moreover, the $\delta^{18}O\text{-}H_2O$ values are enriched versus $\delta^{2}H\text{-}H_2O$, indicating that this shift cannot originate from the evaporation of modern seawater or fresh rain (meteoric) water. This shift was much more pronounced in the *Urania* brine. While possessing similarly depleted $\delta^{2}H\text{-}H_2O$ values, this brine was evidently enriched with the $\delta^{18}O\text{-}H_2O$ isotope (2.0 – 5.2 ‰). For the latter DHAL, this apparent enrichment can be explained by prolonged rock-water interactions of the ancient water under high temperature, which is in concordance with the influence of an active hydrothermal mud vent located underneath Lake *Urania* (Yakimov *et al.* 2007). Hereby, it is likely that the brines of the majority of the studied Mediterranean DHALs were formed in the past upon contact of various layers of the evaporitic suite with interstitial water, thus forming subsurface brine lenses with different hydrochemistry, which were later extruded as already established hypersaline water masses onto the seabed surface through vents or fractures caused by frequent tectonic activity in the Mediterranean Ridge area, and settled in nearby basins. Noteworthy, existence of magnesium-rich interstitial brine pools, entrapped in sediments at depths > 300 meters, was already evidenced more than 20 years ago in the Mediterranean Ridge area (Vengosh *et al*. 1994).

$MgCl_2$-Saturated Dhals Are Obstacles for Life, but their Seawater: Brine Interfaces Are Biological Hot Spots

DHALs of the Mediterranean Ridge represent some of the most hostile environments on our planet, as the saltiest, pressurized, highly reduced and H_2S-rich habitats. Notwithstanding these harsh environmental conditions, both brine and seawater-interface of thalassohaline *Bannock, l'Atalante, Medee, Thetis, Tyro,* and *Urania* are populated by diverse and enigmatic extreme halophilic organisms that belong to all three domains of life (van der Wielen *et al.* 2005; Alexander *et al.* 2009), while brines of *Discovery, Hephaestus* and *Kryos* are sterile and only their interfaces act as refuges for Mg^{2+}-adapted microbial communities (Hallsworth *et al.* 2007; Edgcomb *et al.* 2009; Yakimov *et al.* 2015; La Cono *et al.* 2019). Such a drastic difference in DHALs hospitality can be explained by the contrasting thermodynamic availabilities of water (water activity, a_w) and chaotropic activities (entropic disordering of biomolecules) of their brines.

During the evaporation path, water is, to a large degree, made chemically unavailable by ion hydration. Such a decline in the chemical availability of water strongly affects biological activity by imposing both qualitative and quantitative biophysical limitations on microbial systems (Grant 2004). The a_w of pure freshwater is 1.0, of seawater is 0.98, and of most NaCl-saturated brines, such as saltern crystallizer ponds, inland salt lakes, and thalassohaline deep-sea brines, is $\geq$ 0.741 (Hallsworth *et al.* 2007; Lee *et al.* 2018). This a_w boundary was for a long time considered to be the limit for the most extremely halophilic bacteria and archaea.

However, recent studies have demonstrated biotic activity—namely, differentiation and cell division—close to 0.600 for halophilic bacteria and archaea in saline milieu, and at 0.585 a_w for an ascomycete fungus *Aspergillus penicillioide*s in glycerol-supplemented media (Stevenson *et al.* 2017). Extrapolations suggest theoretical minima for extreme fungal xerophiles in the range 0.570-0.565 (Hallsworth 2019). The *Discovery*, *Hephaestus*, and *Kryos* lakes are polyextreme environments, where the $MgCl_2$ concentrations (5,150, 4,720, and 4,379 mmol kg^{-1}, respectively) are approaching the maximum solubility of this salt. The a_w values of saturated $MgCl_2$ is 0.340, and these values for the *Discovery*, *Hephaestus*, and *Kryos* brines, empirically determined at *in situ* temperature, are 0.382, 0.392, and 0.399 (Hallsworth *et al.* 2007; Yakimov *et al.* 2015; La Cono *et al.* 2019). Thus, there is a substantial thermodynamic distance across the water-activity scale between the points where life is known to cease.

Another harmful feature of $MgCl_2$-rich water solutions, incompatible with the existence of actively metabolizing organisms, is their exceptional chaotropicity (Hallsworth *et al.* 2007; Cray *et al.* 2013a). This limits the propagation of microbial biosphere into high-$MgCl_2$ milieux, as well as other highly chaotropic environments, like $CaCl_2$-saturated Don Juan Pond (Hallsworth *et al.* 1998, 2003, 2007; Williams and Hallsworth 2009; Cray *et al.* 2013b). Empirical determinations show that a 5 M $MgCl_2$ solution has a chaotropic activity of 212 kJ g^{-1}, which is more than twice that of a saturated solution of phenol (Hallsworth *et al.* 2007). In addition, $MgCl_2$ concentrations > 3.0 M have been shown to be beyond the limits of cellular tolerance, regardless of the domain of life (Hallsworth *et al.* 2007; Alves *et al.* 2015; Stevenson *et al.* 2015). Supporting this, previous study on microbial communities of the *Discovery* Lake and recovery of unstable biomarkers, such as mRNA, suggested that in almost pure solutions of $MgCl_2$ representing the *Discovery* brine, active life, as we currently know it, is not likely at $MgCl_2$ concentrations > 2.3 M (Hallsworth *et al.* 2007), which corresponds to < 0.790 a_w. However, various sources of evidence suggest that this limit can also be expanded for other habitats where chaotropic ions can to some extent be compensated by kosmotropic ones (Oren 1983; Hallsworth *et al.* 2003, 2007; Williams and Hallsworth 2009; Bhaganna *et al.* 2010; Bell *et al.* 2013). Thus, the presence of other anions like kosmotropic sodium and sulfate can reduce the net chaotropicity of a hypersaline environment and widen the chaotropicity windows of life. This is exactly the case for DHAL *Hephaestus* and *Kryos*, which contain slightly higher concentrations of Na^+ and SO_4^{2-} ions compared to the brine of *Discovery*, and accordingly, mRNA molecules—indicators for metabolically active cells—were recovered from the 2.27–3.03 M $MgCl_2$ layer (aw 0.747–0.631), thereby expanding the previously recognized chaotropicity window-for-life for *Discovery* Lake (Yakimov *et al.* 2015; La Cono *et al.* 2019).

Identification of the Boundary for Active Microbial Life Within the Seawater-Brine Interface of $MgCl_2$-Saturated Lakes

The 1.5 to 3.0 m thick interface between the seawater column and the underlying $MgCl_2$-dominated brines of Lakes *Discovery*, *Hephaestus*, and *Kryos* provides a unique biophysical environment for studying the limits of life at high magnesium levels. As we mentioned above, recent studies have shown that the brines of DHAL *Discovery*, *Hephaestus*, and *Kryos* are sterile, and their haloclines contain a huge number of cryptic microorganisms (including eukaryotes), each of which differs from "regular" deep seawater (SW) microbiota. Using reconstructed 16S rRNA, assessment of phylogenetic diversity in the last studied Lake *Hephaestus* revealed the existence of a stratified native prokaryotic community at the seawater-*Hephaestus* brine interface down to 2.97 mM $MgCl_2$, referred to as the low interface (LIF).

Despite the juvenility of Lake *Hephaestus* (La Cono *et al.* 2019), it is clear that LIF layer-specific microbial communities inhabit its upper interface. Furthermore, they are phylogenetically distinct from microbial populations of the overlying seawater. In relation to 16S rRNA phylogeny, and even more so for mRNA phylogeny, this community resembled those of the seawater-brine interfaces of *Discovery* and *Kryos* (Hallsworth *et al.* 2007; Yakimov *et al.* 2015). Collectively, these data allude to the existence of hitherto uncharacterized hyperhalophiles, adapted to resist the chaotropicity of $MgCl_2$ and capable of metabolic activity under harshly athalassohaline conditions.

As shown in Figure 3, this layer is dominated by methylotrophic methanogenic (*Methanohalophilus* group) phylum Methanobacteriota and acetogenic representatives of the KB1 group, recently assigned to Candidatus Acetithermum (phylum Candidatus Bipolaricaulota) (Nigro *et al.* 2016). It is worth noting that, like the methylotrophic methanogens and acetogens mentioned above, the LIF community has a significantly higher percentage of sequences belonging to known extremely halophilic archaeal groups (13% vs. 1%) such as the class Halobacteria and Candidate Division Nanohaloarchaeota, compared to MIF. These four groups of highly halophilic and chaotolerant prokaryotes deserve special attention because their trophic interactions may shed light on understanding the adaptation mechanisms that allow these organisms to withstand polyextreme conditions in a $MgCl_2$-saturated environment and thrive at the edge of life.

Syntrophic Relations of Extremely Halophilic Microbial Communities at the Chaotropic Edge of Life

All known DHALs are ecosystems open to and influenced by the overlying seawater column. "Ocean snow" sinks into the DHALs and serves as a source of many biological macromolecules and energy-rich metabolites. These support microbial life in thalassohaline anoxic brine lakes and at the interfaces of athalassohaline DHALs. It is likely that methylated compounds, such as methanol, methanethiol, dimethylsulfide mono-, di-, and trimethylamines, may be among the major catabolic substrates that fuel and sustain the trophic network occurring at the interface (King 1988; Oren 1990, 2011; McGenity 2010; Yakimov *et al.* 2013). Methylamines are often found in saline systems, where they are formed by the decomposition of glycine betaine (GB) or other organic osmoprotectants (compatible solutes) produced by a variety of moderate halophiles possessing "salt-out" adaptation (King 1988; Oren 1990, 2011; McGenity 2010). These organisms typically thrive in both seawater and in superficial, less-saline aerobic or microaerobic layers of the DHAL interface. Apparently, in oxygen-depleted ecosystems, even at moderate salinity (120–150 PSU), neither sulfate reducers nor methanogens can directly utilize GB (Oremland and King 1989; Zhilina and Zavarzin 1990; Andrei *et al.* 2012). Although GB can be reduced with either an external electron donor (for example, hydrogen) or, similarly to amino acids (GB is a methylated derivative of glycine), via Stickland fermentation. In both cases, it requires the activity of halophilic fermenting members of the community, reductively degrading GB to trimethylamine and acetate (King 1988; Oren 1990, 2011). Recently, we investigated a stable anaerobic microbial consortium obtained from the brine of the deep-sea NaCl-saturated Lake *Thetis*. The trophic network of this consortium, established at salinities up to 240 g l^{-1}, relies on the fermentative decomposition of GB to trimethylamine and acetate, carried out by *Halobacteroides lacunaris* TB21, a fermenting member of the *Thetis* enrichment (La Cono *et al.* 2015). In contrast to acetate, which cannot be easily oxidized in salt-saturated anoxic environments, trimethylamine represents an advantageous C_1-substrate for *Methanohalophilus* sp. TA21, a methylotrophic methanogenic member of the *Thetis* enrichment. This second member of the consortium likely produces hydrogen via methylotrophic modification of the

reductive acetyl-CoA pathway because the initial anaerobic GB cleavage reaction requires the consumption of reducing equivalents (La Cono *et al.* 2015).

The description of the Thetis GB-degrading consortium strongly supports the hypothesis that anaerobic degradation of osmoregulatory molecules may play an important role in organic carbon turnover also at the interface of $MgCl_2$-saturated lakes. As we noted above, the LIF of Hephaestus is dominated by acetogenic representatives of the KB1 group (assigned to Candidatus Acetithermum of the Candidate phylum Bipolaricaulota) and methylotrophic methanogens of the *Methanohalophilus* group (Figure 3). Direct cultivation of KB1 (Yakimov *et al.* 2013), supported by fragmentary reconstructions of its genome (Nigro *et al.* 2016), indicated that this acetogenic organism possesses the acetyl-CoA pathway and can use both demethylation and cleavage of GB, depending on environmental factors and potential syntrophic relationships with its partner, the methanogenic *Methanohalophilus*. Thus, it seems reasonable that the well-documented evolution of both acetate and methane in saturated brines of the Mediterranean DHALs (Borin *et al.* 2009; Daffonchio *et al.* 2006; Lazar *et al.* 2011) is at least partially based on a C_1-syntrophic network, where KB1 candidate division and *Methanohalophilus* play roles in the consortium as methylotrophic homoacetogenic bacteria and methylotrophic methanogenic archaea, respectively.

Obviously, in addition to simple low-molecular-weight organic compounds, the aforementioned "ocean snow" also provides DHALs with various high molecular biopolymers, such as polysaccharides of algal/phytoplanktonic (cellulose, hemi- and lignocellulose) and zooplankton (chitin) origin. Therefore, the presence of haloarchaea among the dominant forms of microorganisms in the lower interphase layer of Lake *Hephaestus* is not at all surprising and, moreover, is expected. The haloarchaea discovered there belong to the genera *Halorhabdus* and *Halomicrobium*, known throughout the kingdom Archaea for their unique metabolic capabilities to depolymerize various complex polysaccharides. Indeed, genome sequencing of *Halorhabdus tiamatea*, isolated from DHAB Thetis, demonstrated that this organism possesses one of the highest numbers of glycoside hydrolases, including numerous xylanases, the majority of which were expressed in proteome experiments (Werner *et al.* 2014). The same is true for *Halomicrobium*, which has proven to be a key organism responsible for the depolymerization of chitin in suboxic hypersaline environments worldwide (La Cono *et al.* 2020).

And finally, about the fourth group of extremophilic organisms found in the lower interface of *Hephaestus*, namely nanohaloarchaeota (Figure 3). These nanosized (200-300 nm in diameter) archaea, with their small genomes and limited metabolic capabilities, yet ubiquitous organisms, thrive in hypersaline habitats they share with haloarchaea. We recently revealed the genetic and physiological nature of a nanohaloarchaeon–haloarchaeon association, with both microbes reproducibly cultivated together *in vitro*. The nanohaloarchaeon *Candidatus Nanohalobium* constans is a sugar-fermenting anaerobe, lacking key anabolic machinery and any of respiratory complexes. The cells of the nanohaloarchaeon are found physically connected to the chitinolytic haloarchaeon *Halomicrobium* sp. (Figure 4).

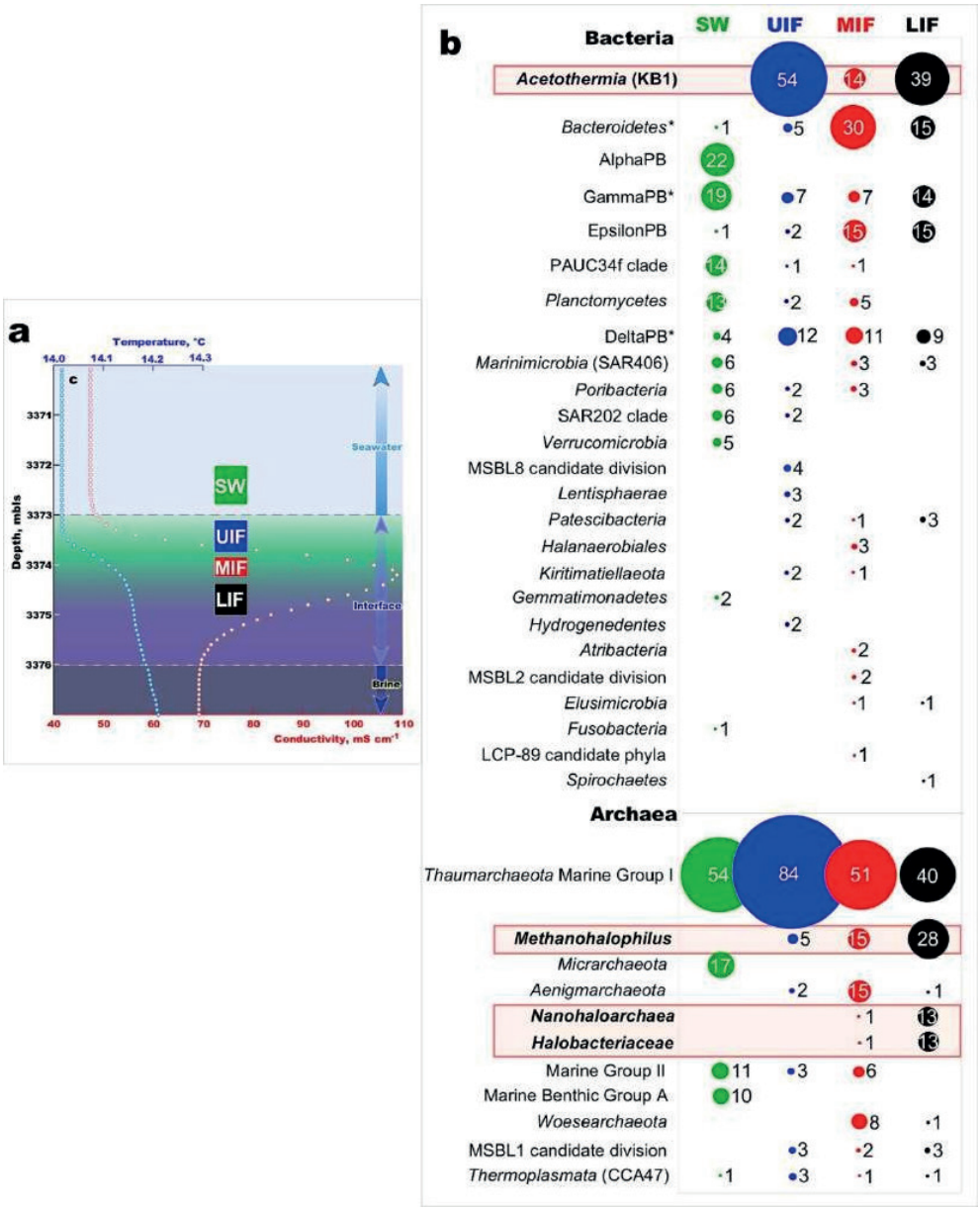


Figure 3. The detailed CTD profile of the of conductivity and temperature data defining the exact depth location of the interface and brine of the lake *Hephaestus* (a). As far as all conventional online CTD sensors are calibrated for seawater and other thalassohaline hydrological formations, they are not fully functional in athalassohaline environments, dominated by bivalent cations. Therefore, the CTD profiling was used only to qualitatively localize the depth and thickness of the *Hephaestus* interface, while chemical analyses were done further in on-land laboratory. Stratification of microbial community, inhabiting various layers of the *Hephaestus* interface given as a percentage of all analyzed clones obtained from corresponding fractions of the interface (b). Groups of organisms dominating LIF microbial community are highlighted by red boxes. Abbreviation used: LIF, the subsample of low interface; MIF, middle interface; UIF, upper interface; SW, seawater sample, collected at 3,372 mbsl (one meter above the *Hephaestus* interface).

However, *Halomicrobium* could not metabolize either glycogen or starch. Surprisingly, the nanohaloarchaeon's ability to hydrolyze these polysaccharides to glucose enabled the growth of *Halomicrobium* sp. in the absence of chitin. These results indicated that the nanohaloarchaeon–haloarchaeon association is both mutualistic and symbiotic; in this case, each microbe relies on its partner's ability to degrade different polysaccharides, and the susceptibility of *Halomicrobium* to colonization by *Ca.* Nanohalobium can be interpreted as a strategy to maximize the long-term fitness of the host.

Figure 4. Scanning electron micrographs of *Halomicrobium* sp. LC1Hm and *Ca.* Nanohalobium constans LC1Nh coculture growing in the high salted medium supplemented with chitin. Images depict tiny coccoidal nanohaloarchaeal cells (285±50 nm in diameter) either detached or adhered to the hosthaloarchaeal cells. Up to 20 nanohaloarchaeota cells can closely interact with the single host cell; nanohaloarchaeal cells express the pilus-like structures (thick and long protein stalks of the archaella, right).

These two syntrophic/symbiotic pairs of polyextremophiles thriving in the *Hephaestus* interface have yielded an interesting insight into how extremely halophilic prokaryotes can prosper in anoxic, hypersaline, and highly chaotropic ecosystems, maintain intracellular turgor, withstand osmotic stress, and produce reduced equivalents at near-saturation salinities (Yakimov *et al.* 2013; La Cono *et al.* 2015, 2020).

In conclusion, we can say that deep-sea anoxic brine lakes represent particular, and in the case of $MgCl_2$-saturated lakes *Discovery*, *Hephaestus* and *Kryos*, unique 3D visuals, located at the bottom of the Mediterranean Ridge. The gradient in the form of an interface occurring between the overlying seawater column and the body of the lakes is nothing but the boundary of multiple physicochemical parameters and can be considered as an oxy-, chemo-, thermo-, and halocline. Despite the polyextremophilic environmental settings, the upper half of the interface of $MgCl_2$-saturated lakes is highly productive, densely populated by very specific organisms, harboring yet-untapped diversity of unique phenotypes, metabolisms, extremozymes, and extremolytes, which await future biomolecular studies and cultivation attempts.

References

Alexander, H., Wanger, G., Malmstrom, R. R. *et al.* (2009). Microbial selection in the euxinic chemocline of the Black Sea and the Mediterranean Sea DHABs. *Environmental Microbiology*, 11(10), 2419–2433.

Alves, F. de L., Stevenson, A., Baxter, E., *et al*. (2015). Concomitant osmotic and chaotropicity-induced stresses in *Aspergillus wentii*: compatible solutes determine the biotic window. *Current Genetics*, 61 (4): 457–477.

Andrei, A.S., Banciu, H.L., Oren, A. (2012). Living with salt: Metabolic and phylogenetic diversity of archaea inhabiting saline ecosystems. *FEMS Microbiology Letters*, 330 (1): 1–9.

Bell, A.N.W., Magill, E., Hallsworth, J.E., *et al*. (2013). Effects of alcohols and compatible solutes on the activity of ß-galactosidase. *Applied Biochemistry and Biotechnology*, 169 (3): 786–796.

Bhaganna, P., Volkers, R.J., Bell, A.N., *et al*. (2010). Hydrophobic substances induce water stress in microbial cells. *Microbial Biotechnology*, 3 (6): 701–716.

Borin, S., Brusetti, L., Mapelli, F., *et al*. (2009). Sulfur cycling and methanogenesis primarily drive microbial colonization of the highly sulfidic Urania deep hypersaline basin. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (22): 9151–9156.

Chamot-Rooke, N., Rabaute, A., Kreemer, C. (2005). Western Mediterranean Ridge mud belt correlates with active shear strain at the prism-backstop geological contact. *Geology*, 33 (11): 861–864.

Cita, M.B. (2006). Exhumation of Messinian evaporites in the deep-sea and creation of deep anoxic brine-filled collapsed basins. *Sedimentary Geology*, 188-189: 357–378.

Cray, J.A., Bell, A.N., Bhaganna, P., *et al*. (2013b). The biology of habitat dominance; can microbes behave as weeds? *Microbial Biotechnology*, 6 (5): 453–492.

Cray, J.A., Russell, J.T., Timson, D.J., *et al*. (2013a). A universal measure of chaotropicity and kosmotropicity. *Environmental Microbiology*, 15 (1): 287–296.

Daffonchio, D., Borin, S., Brusa, T., *et al*. (2006). Stratified prokaryote network in the oxic-anoxic transition of a deep-sea halocline. *Nature*, 440 (7081): 203–207.

De Lange, G.J., Catalano, G., Klinkhammer, G.P., *et al*. (1990). The interface between oxic seawater and the anoxic Bannock brine; its sharpness and the consequences for the redox-related cycling of Mn and Ba. *Marine Chemistry*, 31 (3-4): 205–217.

De Lange, G.J. and Ten Haven, H.L. (1983). Recent sapropel formation in the eastern Mediterranean. *Nature*, 305 (5937): 797–798.

Edgcomb, V., Orsi, W., Leslin, C., *et al*. (2009). Protistan community patterns within the brine and halocline of deep hypersaline anoxic basins in the eastern Mediterranean Sea. *Extremophiles*, 13 (1): 151–167.

Grant, W.D. (2004). Life at low water activity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359 (1448): 1249–1266.

Hallsworth, J.E. (2019). Wooden owl that redefines Earth's biosphere may yet catapult a fungus into space. *Environmental Microbiology*, 21 (6): 2202–2211.

Hallsworth, J.E., Heim, S., Timmis, K.N. (2003). Chaotropic solutes cause water stress in *Pseudomonas putida*. *Environmental Microbiology*, 5 (12): 1270–1280.

Hallsworth, J.E., Nomura, Y., Iwahara, M. (1998). Ethanol-induced water stress and fungal growth. *Journal of Fermentation and Bioengineering*, 86 (5): 451–456.

Hallsworth, J.E., Yakimov, M.M., Golyshin, P.N., *et al*. (2007). Limits of life in $MgCl_2$-containing environments: chaotropicity defines the window. *Environmental Microbiology*, 9 (3): 801–813.

King, G.M. (1988). Methanogenesis from methylated amines in a hypersaline algal mat. *Applied and Environmental Microbiology*, 54 (1): 130–136.

La Cono, V., Arcadi, E., La Spada, G., *et al*. (2015). A three-component microbial consortium from deep-sea salt-saturated anoxic lake *Thetis* links anaerobic glycine betaine degradation with methanogenesis. *Microorganisms*, 3 (3): 500–517.

La Cono, V., Bortoluzzi, G., Messina, E., *et al*. (2019). The discovery of Lake *Hephaestus*, the youngest athalassohaline deep-sea formation on Earth. *Scientific Reports*, 9 (1): 1679.

La Cono, V., Messina, E., Rohde, M., *et al*. (2020). Symbiosis between nanohaloarchaeon and haloarchaeon is based on utilization of different polysaccharides. *Proceedings of the National Academy of Sciences*, 117 (33): 20223–20234.

La Cono, V., Smedile, F., Bortoluzzi, G., *et al*. (2011). Unveiling microbial life in new deep-sea hypersaline Lake *Thetis*. Part I: prokaryotes and environmental settings. *Environmental Microbiology*, 13 (9): 2250–2268.

Lazar, C.S., Dinasquet, J., L'Haridon, S., *et al*. (2011). Distribution of anaerobic methane-oxidizing and sulfate-reducing communities in the G11 Nyegga pockmark, Norwegian Sea. *Antonie van Leeuwenhoek*, 100 (4): 639–653.

Lee, C.J., McMullan, P.E., O'Kane, C.J., *et al*. (2018). NaCl-saturated brines are thermodynamically moderate, rather than extreme, microbial habitats. *FEMS Microbiology Reviews*, 42 (6): 672–693.

McGenity, T.J. (2010). Methanogens and methanogenesis in hypersaline environments. In: Timmis, K.N. (Ed.), *Handbook of Hydrocarbon and Lipid Microbiology*. Springer-Verlag: Berlin, Germany, pp. 665–680.

MEDRIFF Consortium. (1995). Three brine lakes discovered in the seafloor of the eastern Mediterranean. *Eos, Transactions American Geophysical Union*, 76 (33): 313–318.

Nigro, L.M., Hyde, A.S., MacGregor, B.J., *et al*. (2016). Phylogeography, salinity adaptations and metabolic potential of the candidate division KB1 bacteria based on a partial single cell genome. *Frontiers in Microbiology*, 7: 1266.

Oremland, R.S. and King, G.M. (1989). Methanogenesis in hypersaline environments. In: Cohen, Y. and Rosenberg, E. (eds), *Microbial Mats: Physiological Ecology of Benthic Microbial Communities*. American Society for Microbiology: Washington, DC, USA, pp. 180–189.

Oren, A. (1983). *Halobacterium sodomense* sp. nov., a Dead Sea Halobacterium with an extremely high magnesium requirement. *International Journal of Systematic Bacteriology*, 33 (2): 381–386.

Oren, A. (1990). Formation and breakdown of glycine betaine and trimethylamine in hypersaline environments. *Antonie van Leeuwenhoek*, 58 (4): 291–298.

Oren, A. (2011). Thermodynamic limits to microbial life at high salt concentrations. *Environmental Microbiology*, 13 (7): 1908–1923.

Stevenson, A., Cray, J.A., Williams, J.P., *et al*. (2015). Is there a common water-activity limit for the three domains of life? *The ISME Journal*, 9 (6): 1333–1351.

Stevenson, A., Hamil, P.G., O'Kane, C.J., *et al*. (2017). *Aspergillus penicillioides* differentiation and cell division at 0.585 water activity. *Environmental Microbiology*, 19 (2): 687–697.

Van der Wielen, P.W.J.J., Zaid, M., Bolhuis, H. *et al.* (2005). The diversity of bacterial and archaeal communities in a deep-sea hypersaline anoxic lake (DHAB) of the Eastern Mediterranean. *Environmental Microbiology*, 7(12), 2038–2046.

Vengosh, A., Chivas, A.R., Starinsky, A. *et al*. (1994). Chemical and isotopic evidence for the origin of the East Mediterranean brines (Dead Sea and Jordan Rift Valley). *Geochimica et Cosmochimica Acta*, 58(23), 5407–5421.

Wallmann, K.J., Aghib, F.S., Castradori, D., *et al*. (2002). Sedimentation and formation of secondary minerals in the hypersaline Discovery Basin, eastern Mediterranean. *Marine Geology*, 186 (1-2): 9–28.

Wallmann, K.J., Suess, E., Westbrook, G.H., *et al*. (1997). Salty brines on the Mediterranean Sea floor. *Nature*, 387 (6628): 31–32.

Werner, J., Ferrer, M., Michel, G., *et al*. (2014). *Halorhabdus tiamatea*: proteogenomics and glycosidase activity measurements identify the first cultivated euryarchaeon from a deep-sea anoxic brine lake as potential polysaccharide degrader. *Environmental Microbiology*, 16 (8): 2525–2537.

Wessel, P., Smith, W.H.F., Scharroo, R., *et al*. (2013). Generic mapping tools: improved version released. *Eos, Transactions American Geophysical Union*, 94 (45): 409–410.

Williams, J.P. and Hallsworth, J.E. (2009). Limits of life in hostile environments: no barriers to biosphere function? *Environmental Microbiology*, 11 (12): 3292–3308.

Yakimov, M.M., Giuliano, L., Cappello, S., *et al*. (2007). Microbial community of a hydrothermal mud vent underneath the deep-sea anoxic brine lake Urania (Eastern Mediterranean). *Origins of Life and Evolution of Biospheres*, 37 (2): 177–188.

Yakimov, M.M., La Cono, V., La Spada, G., *et al*. (2015). Microbial community of the deep-sea brine Lake *Kryos* seawater-brine interface is active below the chaotropicity limit of life as revealed by recovery of mRNA. *Environmental Microbiology*, 17 (2): 364–382.

Yakimov, M.M., La Cono, V., Slepak, V.Z., *et al*. (2013). Microbial life in the Lake *Medee*, the largest deep-sea salt-saturated formation. *Scientific Reports*, 3: 3554.

Zhilina, T.N. and Zavarzin, G.A. (1990). Extremely halophilic, methylotrophic, anaerobic bacteria. *FEMS Microbiology Reviews*, 87 (3-4): 315–322..

Exploring Microbial Metabolic Networks in a Metagenomics Context; From Raw Data to 3D Modelling of Marine Microbiomes

Mariana Reyes-Prieto[1], Mercè Llabrés[2], Pere Palmer-Rodríguez[2], Kaveh Rassoulzadegan[3], Giuseppe D'Auria[1]

[1]Sequencing and Bioinformatics Service, Foundation for the Promotion of Sanitary and Biomedical Research of the Valencia Region, 46020 Valencia, Spain
[2]Department of Mathematics and Computer Science, University of Balearic Islands, 07122 Palma de Mallorca, Spain
[3]Mediterranean Science Commission (CIESM), Monaco

Abstract

The marine ecosystem covers over 70% of Earth's surface and harbors immense microbial diversity, essential for environmental stability. Microbes drive fundamental ecological processes, yet much of their biodiversity remains unexplored, often referred to as "microbial dark matter." In this context, metagenomics has revolutionized microbial research by enabling direct genomic analysis of environmental samples, thereby bypassing the need for cultivation and providing valuable insights into taxonomic and functional profiling.

Microbial metabolic networks regulate energy flow and ecosystem dynamics, evolving under selective pressures that shape their organization and functional signatures. Moreover, computational analyses have revealed how environmental interactions influence microbial adaptation, offering insights into evolutionary processes and ecosystem resilience.

To support metabolic pathway reconstruction, several resources providing comprehensive databases linking genes, compounds, and biochemical reactions already exist. Here, we present a case study of the MetaDag tool, which enables hierarchical visualization of metabolic dependencies and maps biochemical interactions across microbial communities. Furthermore, for the first time, we present a prototype of the MetaCube application, facilitating multidimensional analysis of metabolic pathways by integrating taxonomic and functional data to reveal ecosystem-wide metabolic trends.

By combining metagenomic data with advanced computational tools, researchers can uncover the intricate relationships between marine microbes and their environment. These approaches improve our understanding of oceanic biogeochemical cycles, microbial interactions, and the role of microbes in sustaining marine ecosystems.

Keywords: metagenomics, metabolic networks, marine microbiome, bioinformatics, MetaDAG, 3D modeling, ocean gradients

Introduction

Over 70% of Earth's surface is covered by the marine ecosystem, and due to the diversity of their surroundings, oceans host a highly varied array of species. Microbes possess an extensive assemblage of physiological properties that are largely unknown but fundamental for the environment to thrive. Their emergence is a result of long-term evolutionary processes involving physiological adaptations under diverse environmental conditions and selective

pressures. Oceans are an ideal environment for biodiversity; they are vast and well-connected, boast an endless water supply, maintain milder temperatures than land, and exhibit a higher diversity of phyla and classes compared to freshwater and terrestrial environments. Within the ocean, species richness typically decreases with depth, yet deep-sea species often have wider geographic ranges than those in coastal areas (Costello and Chaudhary 2017).

Among natural biomes, microbial diversity and the highly variable relative frequency of each player characterize metabolic processes and interactions. It's worth noting that cultivable organisms represent only a tiny fraction of the world's known diversity. Consequently, much microbial biodiversity remains hidden or undescribed, forming an enormous reservoir of diversity and metabolic capabilities, commonly referred to as "dark matter" (Marcy et *al.* 2007).

Metagenomics stands as one of the most advanced approaches to microbial characterization. This field of molecular biology involves studying genetic material recovered directly from environmental samples without the need to isolate and culture individual organisms. This allows for the analysis of the collective genomes of all microorganisms present in a particular habitat, providing a comprehensive view of their overall diversity and functional profiles (Stein *et al*. 1996).

The focus on microbial metagenomics began as researchers started to understand the crucial role microbes play, participating directly or indirectly in environmental or host-related metabolisms and homeostasis. From a general perspective, the microbiota—including bacteria, archaea, protists, fungi, and viruses—forms a deeply interconnected network of metabolic processes where tight interactions maintain a dynamic equilibrium across all systems. Natural niches are typically populated by specific microbiota structures that interact amongst themselves and with their associated hosts. Moreover, in microbial-host interactions, hosts provide shelter, metabolites, and nutrients necessary for the microbial fraction to survive, grow, and maintain their metabolic networks. In turn, microbiota metabolizes compounds for which the host lacks degradation capacity, producing byproducts of great value for the host's metabolism (Jansma and El Aidy 2021).

Marine microbiome characterizations are based on defining their taxonomic and/or functional profiles, for example, descriptions within and between different environments. Standard profiling approaches employ either amplicon-based (metataxonomy) or metagenomic strategies. Amplicon-based strategies, which target the 16S ribosomal gene (for bacteria and archaea), 18S (for higher eukaryotes), or the internal transcribed sequences (ITS, mainly for fungi), offer the fastest and most cost-effective proxy for microbiome profiling. In this framework, the PCR product is considered a representative subsample of the original environment, and the obtained DNA is then sequenced. Conversely, the metagenomic approach relies on the random fragmentation and sequencing of environmental DNA, providing, after sequence annotation, a general overview of the taxonomy and potential functional profiles of that specific environment.

The entire range of potential biological processes an organism can perform is referred to as its metabolic network (Parter *et al.* 2007). The sum of these metabolic networks, constructed across the complex set of protists, bacteria, archaea, and viruses, forms an "ecosphere" in which energy dynamics and equilibria maintain the basic energy homeostasis of every natural environment.

Metabolic networks are built upon energy flows through enzymes and compounds, conforming to the metabolic capabilities of each organism. Generally, these networks are typically expressed as nodes and edges with a specific structure called "topology" (Figure 1).

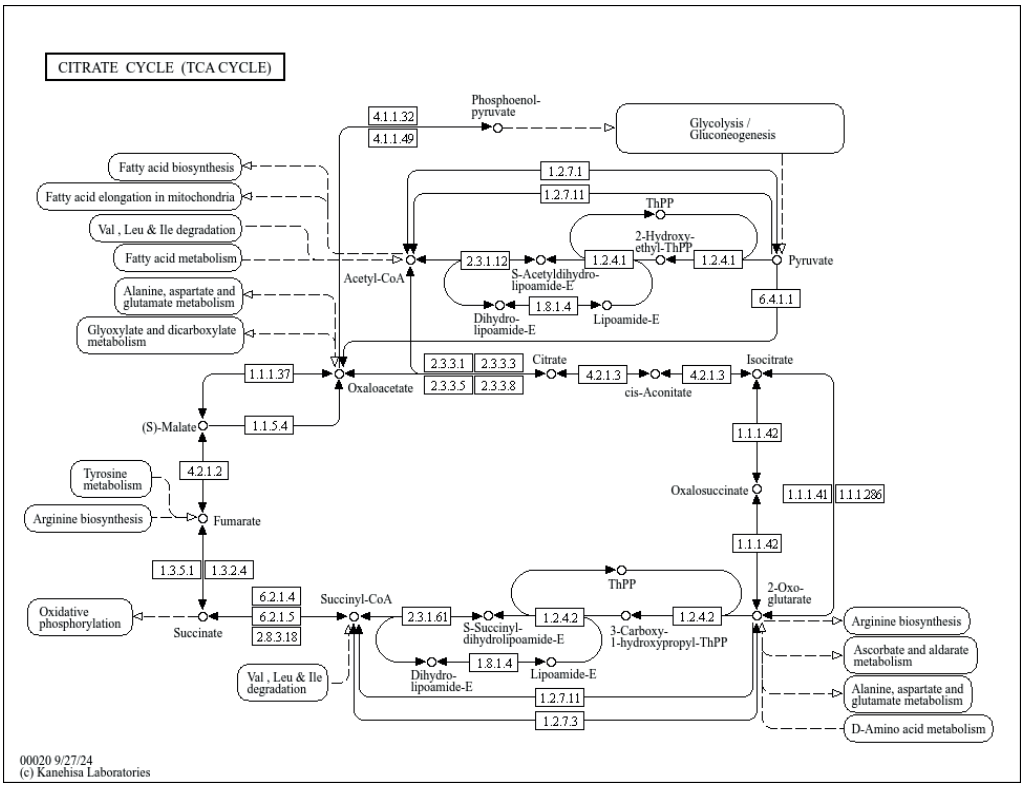


Figure 1. Reference pathway from KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa *et al.* 2025) for the citrate cycle. One simple metabolic pathway depicted as a metabolic network.

Evolutionary studies on various metabolic networks have demonstrated that due to diverse topological features, alterations in their nodes and edges are subject to different selection pressures, thereby capturing the organism's functional aspects. Additionally, environmental interactions through evolution, coupled with selection pressures from the environment on the metabolic network, can change the network's organization and produce a topological "signature." Network analyses can also shed light on the environment in which a species evolved (Yamada and Bork 2009).

Prior research has concentrated on computational methods to examine the topological characteristics of small-scale metabolic networks, typically involving one or a few species at a time. Other research has focused on the context in which each network operates, investigating the impact of biochemical settings and examining multiple metabolic pathways concurrently within certain models (Ibarra *et al.* 2002; Almaas *et al.* 2004).

To aid the metabolic reconstruction of pathways, the metabolism of specific organisms, or even complex networks such as those involving multiple organisms or an entire ecosystem, we rely on KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa *et al.* 2025). KEGG is a comprehensive database resource for understanding the high-level functions and utilities of biological systems from molecular-level information. This database comprises all information on metabolic pathways in maps, including a catalogue of genes, compounds, reactions, and orthologs. Figure 2 shows the complete reference metabolic pathway included in their database, describing the global association between reactions and compounds across all organisms.

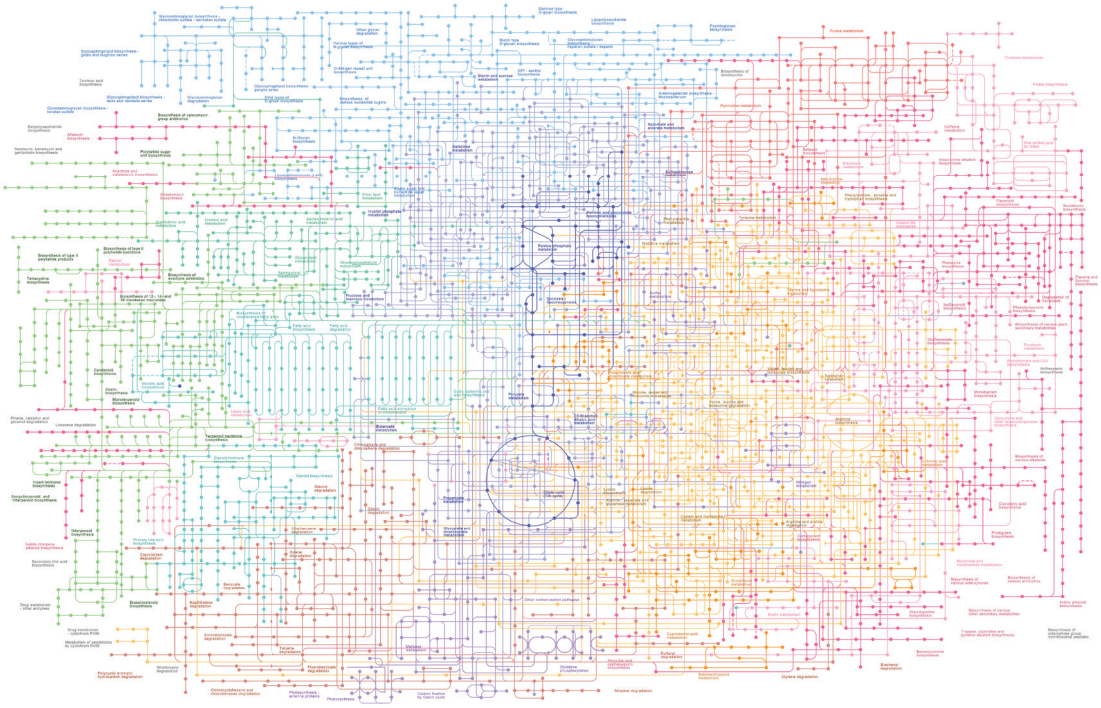


Figure 2. Complete reference metabolic pathway from Kanehisa *et al.* 2025).

Microbial Networks in the Oceans

The ocean's surface is continuously collecting organic matter from the environment, which sinks through the water column. The microbial fraction uses this infinite carbon source, structuring a complex "network of networks" where every actor is tightly linked to the following peer and just standing or leaning their product to the following network.

The networks do not work in closed compartments but follow interleaved gradients of metabolic compounds and biochemical activities through the physicochemical shapes offered by the surrounding environment. Moreover, the medium in which the microbial fraction orchestrates these complex relationships is not a continuum liquid medium but is composed of a continuously changing matrix, offering a fractal of conditions and biochemical possibilities. Thus, a part of free-living bacteria, dark matter, and attached bacteria are part of this continuum, increasing the interface landscapes from macro- to microscales.

Therefore, being able to account for the interfaces, the possibilities of interactions among microbes become enormously complex. This brings us to the impossibility of getting a continuum description of microbiomes, where the boundaries of the diversity or metabolic data descriptions are not always defined but belong to the "slice" or "snapshot" of the collected sample.

Still, the best we can do regarding the sampling strategies is about collecting a sample snapshot from which the microbial DNA is extracted and sequenced through a metagenomic approach, reporting the genetic potential of a given sample from a spot in a three-dimensional space. Despite the apparent triviality of this sentence, the physicochemical continuum permeating the sample is the force driving the metabolic pathways we can observe and expect.

The KEGG atlas may provide a queue of the available metabolisms that can be found in a given sample. Thus, the possibility of microbiome configurations inhabiting a sample is reduced according to the spatial niche.

In general, there are central metabolism pathways that may be recognized by all taxonomic groups, independent of whether these are free-living or attached lifestyles. Things may change in the case of symbiotic, or strictly symbiotic, organisms where a process of genome reduction may bring about the loss of pathway steps where the enzyme/metabolite couple may be provided by one of the actors of the symbiosis.

Currently, the metabolism ontologies collect hundreds of different metabolic maps, and very few of them may be considered uniquely used by a taxonomic organism. For example, the case of methanogenic archaea, where the production of methane may be considered fairly unique. Thus, while it is clear that the microbiome's asset possibilities are limited by environmental constraints and by the buffering and redundancy power offered by the proper community, drafting a catalogue of microbiomes is still a challenge. For example, we can name organisms by some specific metabolic capabilities (singularity), such as sulfate-reducing bacteria (SRB), Anammox bacteria, etc. This kind of classification acts as a way to identify a bacterium by its genomic traits.

In a more general and environmental context, many kinds of bacteria with different metabolic "abilities" may survive in the same environment. Thus, a more general vision centred on the environment may rely on the ability to identify the players of a given environmental asset. Thus the metabolite, energy, and by-product fluxes at a given point may be the key for representing and generalizing a given geographically defined environment. Thus, a more objective approach is probably the metabolomic one, or at least its inference from metagenomic data.

Topological Dynamics in Metabolic Networks of Microbial Communities

Microbial communities inhabiting extreme or fluctuating environments—such as hypersaline, anoxic, or high-pressure ecosystems—exhibit remarkable adaptability in their metabolic organization. Environmental stressors in these systems can drive both gradual and abrupt changes in the topological structure of metabolic networks. Gradual adaptations often involve the progressive rewiring of metabolic pathways to optimize energy efficiency, nutrient acquisition, or stress tolerance, reflecting a form of metabolic plasticity. In contrast, abrupt topological shifts may entail large-scale reconfigurations, where entire pathways are reorganized or replaced in response to sudden environmental perturbations or evolutionary pressures (Figure 3).

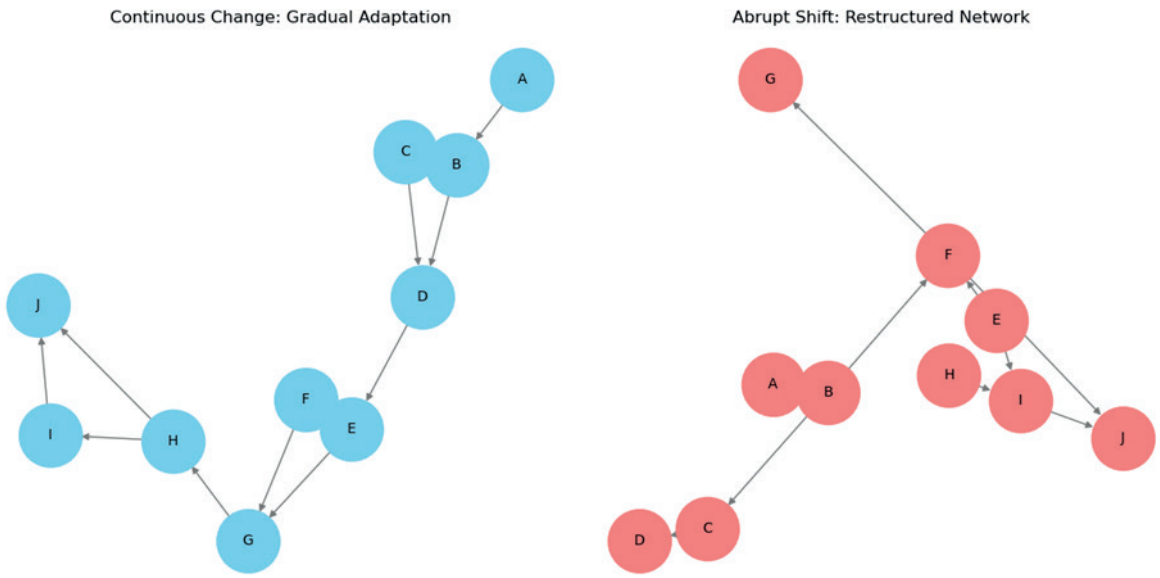


Figure 3. Conceptual diagram illustrating continuous metabolic adaptation through gradual pathway rewiring (left) and abrupt topological shifts involving pathway removal and new connections (right) in microbial metabolic networks.

These dynamic responses underscore the resilience of microbial life and provide a framework for understanding how metabolic systems evolve under environmental constraints. However, our ability to capture these dynamics is inherently limited by the temporal resolution of microbiome analyses. Most studies rely on discrete sampling points—essentially snapshots in time—that may miss transient states or rapid transitions in network topology. As a result, distinguishing between continuous adaptation and abrupt reorganization remains challenging, especially in environments where microbial turnover and environmental fluctuations occur on short timescales. Addressing these limitations will require more refined temporal sampling strategies and integrative approaches that combine metagenomics, transcriptomics, and metabolic modeling to reconstruct the trajectories of microbial network evolution with greater fidelity (Chalancon *et al*. 2013; Schreiber *et al*. 2021; Ye *et al*. 2022).

The MetaDAGs Approach to Marine Microbiome Representations as a Reductive Approach to Metabolic Networks

Metabolic network reconstructions from complete metagenomic data are currently scarce in the literature. It's incredibly complex to construct and visualize all metabolic pathways within a metagenomic dataset, making it extremely difficult to identify differences in metabolic networks across various samples. To tackle this challenge, we're using a methodology that infers metabolic networks from metagenomic data with MetaDAG, a web tool first described in Alberich *et al.* (2017) and implemented in Palmer-Rodríguez *et al.* (2025).

MetaDAG operates by using the notion of strongly connected components of a reaction graph, known as metabolic building blocks (MBBs), and contracting them into single nodes. The resulting graph is a directed acyclic graph called a metabolic DAG (m-DAG). This significantly reduces the number of nodes while highlighting a background graph topology that reveals the metabolic network's connectivity, including bridges, isolated nodes, and cut nodes—all

crucial information for discovering functional metabolic relationships. Additionally, MetaDAG can generate metabolic profiles and pan and core metabolic frames from a group of samples. Panmetabolism refers to the set of reactions present in at least one organism in a community of organisms, while core metabolism is the set of common reactions shared by all organisms within that community. These pan and core metabolic frames are useful for easily visualizing differences between samples, which can help classify the metabotypes of marine microbiomes.

Building on this concept, we're constructing a reference metabolism in a reaction graph, analogous to the KEGG database's reference pathway (Figure 2). We've parsed each reaction and pathway into a global 3D environment we call MetaCube. MetaDAG can then infer the metabolic networks of specific metagenomic samples by directly applying its methodology, calculating all reactions and metabolites found in the genomes of each marine microbiome's bacterial load. Finally, we used MetaCube to create a differential visualization of metagenomic samples from various marine contexts. The reductionist approach of MetaDAG and the 3D metabolic pathway reconstructions allowed us to identify and visualize the main energy flux differences among the niches we studied.

Material and Methods

A total of 30 marine metagenomic samples were retrieved from the MGnify database (https://www.ebi.ac.uk/metagenomics; Mitchell *et al*. 2020). These samples correspond to the project associated with the study "Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation" by Ustick *et al*. (2021). The accession numbers for the selected samples are ERR598949, ERR598974, ERR599150, SRR12479657, SRR12480024, SRR12480136, SRR12480284, SRR12480297, SRR12480319, SRR12480345, SRR12480351, SRR13122117, SRR13122118, SRR13122185, SRR13122207, SRR13122230, SRR13122271, SRR5720227, SRR5720233, SRR5720241, SRR5720257, SRR5720271, SRR5720276, SRR5787989, SRR5788001, SRR5788022, SRR5788035, SRR5788042, SRR5788047, and SRR5788054. All samples consisted of paired-end Illumina sequencing reads derived from marine water. Metadata associated with each sample included geographic origin (Indian, Pacific, or Atlantic Ocean), sampling depth (in meters), and season of collection.

De-novo metagenomic processing was conducted using SqueezeMeta v1.6.3 (Tamames and Puente-Sánchez 2019). The functional annotation included in this software was carried out using DIAMOND (Buchfink *et al*. 2015) to perform similarity searches against multiple databases, including GenBank (Clark *et al*. 2016), CARD, MEGARES, and the KEGG database (Kanehisa and Goto 2000). Read mapping against assembled contigs was performed using Bowtie2 (Langmead and Salzberg 2012).

KEGG orthologs (KOs) identified in each sample were used to reconstruct metabolic networks and generate metabolic-directed acyclic graphs (metabolic DAGs or mDAGs) with the software MetaDAG (Palmer-Rodriguez *et al*. 2025) for each microbiome. We used Query 6 of the MetaDAG web interface to extract and compare the metabolic networks associated with each microbiome.

These networks were then modeled using the software MetaCube, a platform designed to visualize the three-dimensional spatial organization of metabolic networks interactively within a web browser. MetaCube is built on top of the WebGL2 Three.js render engine, using lil-gui as a Graphical User Interface (GUI). Reactions and pathways are represented using instanced buffer geometry. MetaCube enables the exploration of interactions among metabolic reactions based on shared input and output compounds, providing a spatially resolved view of pathway

integration and metabolic connectivity within each microbial community. It allows users to compare, load, save, import, and export sample datasets from different sources, and provides a search engine to highlight specific reactions and pathways.

Results

The annotation and MetaDAG analyses of each sample yielded 30 individual reaction graphs, each representing a single sample's metabolic landscape. We also obtained their corresponding metabolic-directed acyclic graphs (mDAGs), which offer a topological abstraction of the underlying metabolic pathways. To assess potential ecological or biogeographical patterns, we further stratified the samples by oceanic origin and examined group-level differences in their metabolic architecture.

As a first step in our comparative analysis, we characterized the dataset's collective metabolic repertoire by computing both the pan- and core metabolisms. The pan-metabolism, defined as the union of all metabolic reactions identified across the 30 samples, comprises 5,734 reactions distributed among 1,638 metabolic building blocks (MBBs). In contrast, the core metabolism—representing the set of reactions shared across all samples—consists of 5,042 reactions organized into 1,502 MBBs. These results indicate a high degree of metabolic conservation among the sampled microbial communities.

To further explore the consistency of metabolic functions across environmental gradients, we examined the number of shared reactions and MBBs under various stratifications, including ocean basin, sampling depth, and season of collection. Table 1 summarizes these comparisons. Across all groupings, the core metabolism consistently accounts for more than 90% of the total reactions, suggesting that these communities' fundamental metabolic capabilities are largely preserved despite spatial and temporal variability. This high level of functional redundancy points to a robust and conserved metabolic framework, likely reflecting essential biochemical processes required for microbial survival and ecological function in marine environments.

Dataset	Total reactions	Core reactions	Pan MBBs	Core MBBs
All	5732	5042	1638	1502
Depth 0-5 m	5717	5517	1633	1618
Depth 6-14 m	5732	5562	1638	1618
Depth 15-25 m	5734	5042	1638	1502
Spring	5734	5579	1638	1623
Summer	5684	5517	1621	1618
Autumn	5732	5562	1638	1618
Winter	5732	5042	1638	1502
Atlantic Ocean	5732	5617	1638	1630
Indian Ocean	5649	5042	1631	1502
Pacific Ocean	5734	5562	1638	1618

Table 1. Number of reactions and MBBs resulting from the groups of samples included in this study.

Subsequently, we analyzed the outputs generated by MetaDAG, as a key challenge in big data analysis is the precise identification and interpretation of subtle yet significant variations among samples. In scientific research, isolating the most informative elements from complex datasets is critical. To facilitate this, we applied the sparse variant of Partial Least Squares Discriminant Analysis (sPLS-DA), a technique designed to identify the most predictive or discriminative features (MBBs) for effective sample classification (Lê Cao *et al*. 2011). This analysis was conducted using the MixOmics package in R (Lê Cao *et al*. 2016).

Figure 4 presents a heatmap generated from the sPLS-DA, illustrating the degree of discrimination for each MBB across the different sample groups. The color gradient, ranging from dark blue to dark red, reflects the discriminative or predictive strength of each MBB. Each row corresponds to a sample, labeled on the right, while the columns represent individual MBBs. Notably, for the location metadata, the 25 MBBs on the left and the 35 on the right effectively cluster the Indian

Ocean samples, with the exception of a single outlier. In contrast, the MBBs located in the central region of the heatmap group the remaining samples, though without a clear separation between those from the Pacific and Atlantic Oceans.

The dendrogram at the top of the figure clusters MBBs based on their contribution to group discrimination, while the dendrogram on the left organizes the samples (mDAGs) according to their similarity. Metadata annotations are also included in the upper left corner of the figure, providing contextual information for each sample. To further explore the MBBs that appear as thin vertical lines in dark red—indicating high predictive value—we extracted the most significant features for each group.

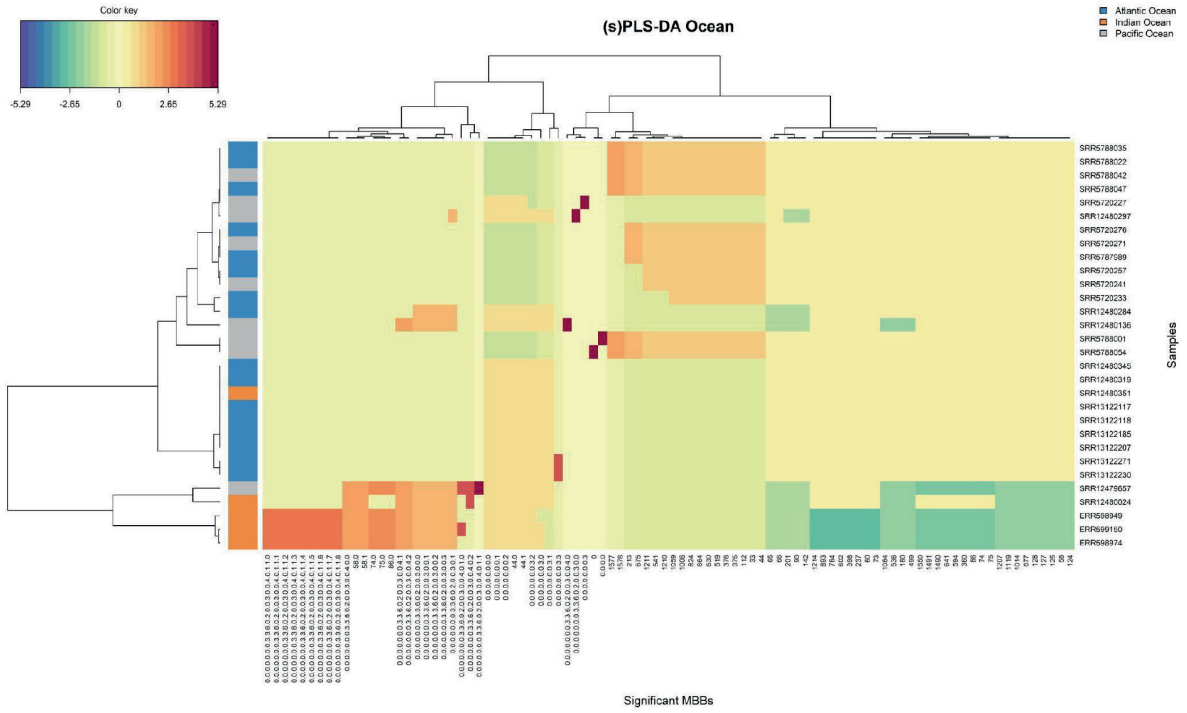


Figure 4. Heatmap of sparse Partial Least Squares Discriminant Analysis (sPLS-DA). Rows represent individual samples (labelled on the right), while columns display metabolic building blocks (MBBs). The colour gradient indicates discriminative strength, from most predictive (dark red) to most discriminative (dark blue) MBBs. Dendrograms illustrate the clustering of MBBs (top) and the classification of m-DAGs by similarity (left).

These key MBBs are examined in greater detail in Figure 5, which displays the top 50 loadings plot derived from the sPLS-DA. Each box in the plot represents an individual MBB, with its line reflecting its loading value in the discriminant components. The color of each point indicates the group for which the MBB has the highest mean value, thereby highlighting its relative importance in distinguishing that group from the others. This visualization not only identifies the most influential metabolic features but also provides insight into their potential biological relevance, supporting the interpretation of the clustering patterns observed in Figure 4.

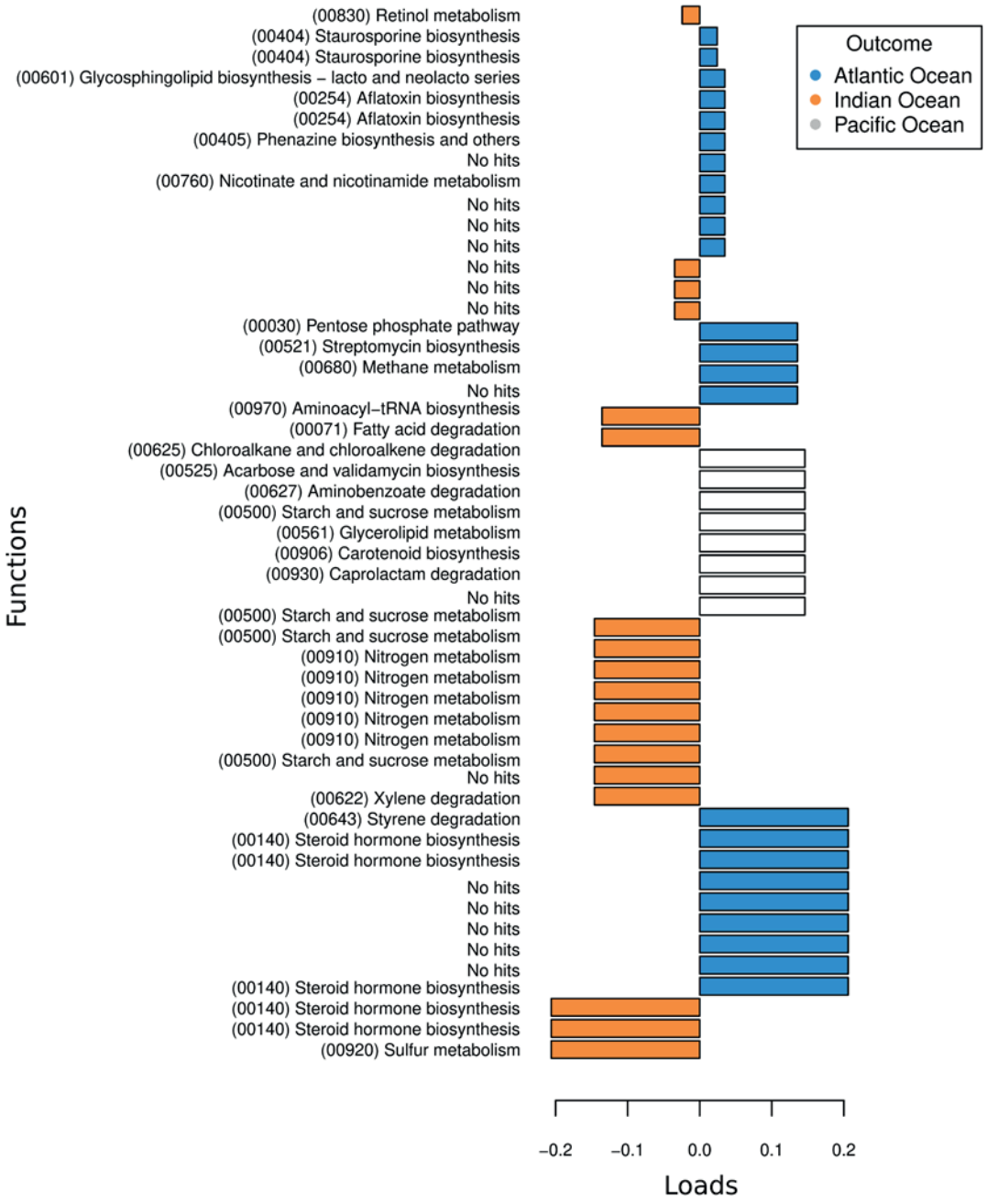


Figure 5. sPLS-DA Loading Plot of Metabolic Building Blocks (MBBs). This plot highlights the contribution of individual MBBs to group discrimination. Each point represents an MBB and its associated metabolic pathway(s), with its position reflecting its loading value in the discriminant components. Points are coloured by the group for which the MBB has the highest mean value, emphasizing its relative importance in distinguishing that group. This visualization identifies key metabolic features driving the observed sPLS-DA separation, offering insights into biological differences between sample groups.

Figure 5 reveals interesting differential metabolic capabilities among microbiomes from the Atlantic, Indian, and Pacific Oceans. While these results primarily serve as a proof of concept for using these applications to identify environmental signatures, we observe clear distinctions. Notably, some steroid-related MBBs show a negative correlation with the Indian Ocean, in contrast to a positive correlation in the Atlantic Ocean.

Interestingly, Pacific Ocean samples appear enriched in pathways for caprolactam degradation, carotenoid biosynthesis, glycerolipid metabolism, starch and sucrose metabolism, aminobenzoate degradation, acarbose and validamycin biosynthesis, and chloroalkane and chloroalkene degradation. These findings may suggest communities adapted to high energy loads and/or those involved in degrading anthropogenic compounds, hinting at significant anthropocentric activity.

The original study from which these samples were extracted (Ustick *et al*. 2021) provided a comprehensive analysis of nutrient limitation across major ocean regions, including the Pacific Ocean. Their findings indicate that microbial communities adapt to nutrient stress through genomic changes, particularly in genes associated with nitrogen, phosphorus, and iron assimilation. Our analysis here delves into the details of common metabolic blocks shared among bacteria within a given environment, employing a pangenomic approach that summarizes metabolic redundancy among microbial populations.

Similarly, another study on nutrient distribution in the subtropical southern Indian Ocean reported that surface waters are nutrient-depleted, with low nitrate (< 3 µmol/kg) and phosphate (< 0.3 µmol/kg) concentrations (Harms *et al*. 2019). Our analysis corroborates this by showing that sulfur, starch, xylene, sucrose, and nitrogen metabolism are negatively correlated in the Indian Ocean compared to the other oceans. We further utilized nitrogen-related MBBs as an example to demonstrate MetaCube viewer functionalities (Figure 6).

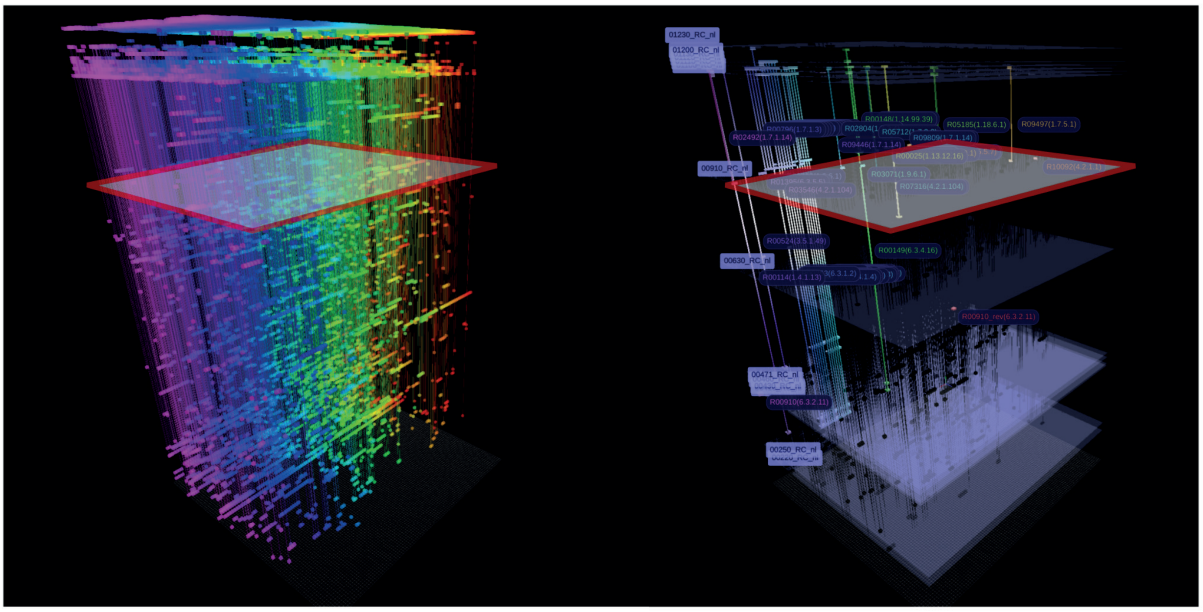


Figure 6. MetaCube Interface for Metabolic Network Visualization. The interface displays metabolic pathways as distinct planes, which can be highlighted via mouse interaction. Colored blocks represent enzymes, with identical enzymes connected by vertical lines across planes. The left panel shows the complete enzymatic dataset (pangenome) for the Indian Ocean samples. The right panel highlights pathway 00910, corresponding to nitrogen metabolism within the same samples.

At the time of writing, the MetaCube application allows importing Metabolic Building Block (MBB) reaction datasets for visualization in a 3D immersive environment. This visualization approach draws inspiration from the graphical identification of areas in Computerized Axial Tomography (CAT scan) or X-ray visualizations, where expert personnel are trained to pinpoint problematic regions. Though in its early stages, MetaCube provides a free, 3D navigable

environment that enables investigation into the presence and absence (in the current version) of specific reactions from metagenomic annotation data.

In the near future, we plan to introduce boolean operation filtering, comparative tools, and pattern recognition tools based on machine learning (ML) modeling approaches. These enhancements will improve the recognition power and representation of metagenomics data. As the volume and complexity of such data grow, ML will become an essential tool to upgrade this exploratory process from passive visualization to active pattern recognition and hypothesis generation. This includes automatically detecting clusters or functional zones, predicting ecological or disease states, discovering logical feature combinations, analyzing and predicting metabolic network structures, comparing entire samples or environments, enhancing navigation and spatial representation, incorporating user feedback to improve predictions, and adapting models from public datasets.

Conclusions

The comparative tools presented in this study effectively reduce the complexity of environmental metagenomics data, revealing distinct functional patterns across samples from five different marine environments. As the complexity and volume of metagenomic data continue to grow, there's an increasing need for robust summarization and analytical frameworks capable of identifying differential patterns. While still in its early stages, the ability to reduce the redundancy of the genomic potential of microbial communities is crucial for improving our capacity to understand the core aspects of what's truly happening within a given microbial community.

Our findings demonstrate that reducing genomic redundancy and conducting subsequent differential mDAG analysis across geographic regions or environmental gradients not only highlights the key functioning features distinguishing these ecosystems but also directly helps uncover potential molecular signatures. These signatures can serve as ecological markers or early warning indicators, offering valuable insights into environmental shifts and ecosystem health. Finally, a new immersive approach to achieve this level of resolution is proposed, which will form the basis for an advanced method of delving into the more intricate nature of microbial communities.

References

Alberich, R., Castro, J.A., Llabrés, M., *et al*. 2017. Metabolomics analysis: Finding out metabolic building blocks. *PLoS One*, 12(5): e0177031.

Almaas, E., Kovács, B., Vicsek, T., *et al*. 2004. Global organization of metabolic fluxes in the bacterium Escherichia coli. *Nature*, 427(6977): 839–843.

Buchfink, B., Xie, C., Huson, D.H. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1): 59–60.

Chalancon, G., Kruse, K., Babu, M.M. 2013. Metabolic networks, structure and dynamics. *In* Dubitzky, W., Wolkenhauer, O., Cho, K-H. and Yokota, H. (eds), *Encyclopedia of Systems Biology*: 1263–1267. Springer.

Clark, K., Karsch-Mizrachi, I., Lipman, D.J., *et al*. 2016. GenBank. *Nucleic Acids Research*, 44(D1): D67–D72.

Costello, M.J. and Chaudhary, C. 2017. Marine Biodiversity, Biogeography, Deep-Sea Gradients, and Conservation. *Current Biology*, 27(13): R511–R527.

Harms, N.C., Lahajnar, N., Gaye, B., *et al*. 2019. Nutrient distribution and nitrogen and oxygen isotopic composition of nitrate in water masses of the subtropical southern Indian Ocean. *Biogeosciences*, 16: 2715–2732.

Ibarra, R.U., Edwards, J.S., Palsson, B.Ø. 2002. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420(6912): 186–189.

Jansma, J. and El Aidy, S. 2021. Understanding the host-microbe interactions using metabolic modeling. *Microbiome*, 9(1): 16.

Kanehisa, M., Furumichi, M., Sato, Y., *et al*. 2025. KEGG: Biological systems database as a model of the real world. *Nucleic Acids Research*, 53(D1): D672–D677.

Kanehisa, M., Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1): 27–30.

Lê Cao, K-A., Rohart, F., Gonzalez, I., *et al*. 2016. mixOmics: Omics Data Integration Project. R package version 6.1.1.

Langmead, B. and Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4): 357–359.

Lê Cao, K-A., Boitard, S., Besse, P. 2011. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinform*, 12: 253.

Marcy, Y., Ouverney, C., Bik, E.M., *et al*. 2007. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proceedings of the National Academy of Sciences*, 104(29): 11889–11894.

Mitchell, A.L., Almeida, A., Beracochea, M., *et al*. 2020. MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Research*, 48(D1): D570–D578.

Palmer-Rodríguez, P., Alberich, R., Llabrés, M. 2025. Metadag: A web tool to generate and analyse metabolic networks. *BMC Bioinformatics*, 26(1): 31.

Parter, M., Kashtan, N., Alon, U. 2007. Environmental variability and modularity of bacterial metabolic networks. *BMC Evolutionary Biology*, 7(1): 169.

Schreiber, F., Grafahrend-Belau, E., Kohlbacher, O., *et al*. 2021. Visualising metabolic pathways and networks: Past, present, future. *In Integrative Bioinformatics*: 237–267. Springer.

Stein, J.L., Marsh, T.L., Wu, K.Y., *et al*. 1996. Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology*, 178(3): 591–599.

Tamames, J. and Puente-Sánchez, F. 2019. SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Frontiers in Microbiology*, 9: 3349.

Ustick, L.J., Larkin, A.A., Garcia, C.A., *et al*. 2021. Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation. *Science*, 372(6539): 287-291.

Yamada, T. and Bork, P. 2009. Evolution of biomolecular networks: Lessons from metabolic and protein interactions. *Nature Reviews Molecular Cell Biology*, 10(11): 791–803.

Ye, C., Wei, X., Shi, T., *et al*. 2022. Genome-scale metabolic network models: From first-generation to next-generation. *Applied Microbiology and Biotechnology*, 106: 4907–4920.

WORKSHOP COMMUNICATIONS

D) IT TOOLS AND DATABASES

Tara Pacific Metadata: a Powerful Tool for a Holistic Study of the Adaptation and Resilience of Coral Reefs in the Anthropocene

Paola Furla[1,2] and the Tara Pacific Consortium[3]

[1]Université Côte d'Azur, CNRS, Inserm, IRCAN, Nice, France
[2]LIA ROPSE, Laboratoire International Associé Université Côte d'Azur - Centre Scientifique de Monaco, Monaco, France
[3]Tara Pacific Consortium: https://doi.org/10.5281/zenodo.3777760

Coral reefs are biodiversity hotspots of the planet (Reaka-Kudla 1997). They are not only of crucial ecological importance, serving as part of the food chain and providing shelter for animals, but they also represent an important economic asset (*e.g.*, fishing, material resources, recreation) that is estimated to be more than 30 Billion $/year. They are formed by colonial Cnidarians that live in symbiosis with unicellular photosynthetic algae from the *Symbiodiniaceae* family, which supply them energy, facilitate skeletal growth rate (Allemand *et al.* 2004; Iwasaki *et al.* 2016), and stress resistance (Richier *et al.* 2005). In the context of ocean warming, successive heat waves have already proven to be the main cause of large-scale coral bleaching events (*i.e.*, symbiosis breakdown), possibly leading to coral death and endangering the entire reef ecosystem (Bindoff *et al.* 2019). While coral bleaching episodes have been recorded worldwide, regional-scale responses, variations in susceptibility among coral taxa, and changes in the reef coral species assemblage before, during, and after bleaching have also been observed.

Between 2016 and 2018, the research schooner Tara sailed more than 100,000 km across the Pacific Ocean, aiming to carry out a holistic study of reef coral adaptation and resilience in the face of climate change (Planes *et al.* 2019). During its two-year voyage, through 3,000 scuba dives, the Tara Pacific expedition sampled coral ecosystems from 32 islands across the Pacific Ocean and ocean surface waters at 249 locations, resulting in the collection of nearly 58,000 samples using various approaches (*ibid.*). Specifically, this expedition allowed the building of an unprecedented database and community resource on coral reef complexity, including metagenomics, metatranscriptomics, metabarcoding, stress biomarkers, and telomere DNA length for more than 2,000 colonies. These data were matched to a large dataset of contextual and long-term environmental variables obtained from a suite of methods, including collected samples, automated on-board measurements, biogeochemical models, and satellite imagery (Lombardi *et al.* 2023; Belser *et al*. 2023; Galand *et al.* 2023).

This ambitious project aimed to provide a reference about the biological state of modern coral reefs in the Anthropocene. Among the new insights coming from the Tara Pacific program, we assessed the entire microbial and chemical diversity of the coral holobionts at a basin-wide scale (Galand *et al.* 2023; Hochart *et al.* 2023; Veglia *et al.* 2023). It also highlighted the multiple strategies of corals selected during evolution to respond to environmental fluctuations (Armstrong *et al.* 2023; Canesi *et al.* 2023; Voolstra *et al.* 2023; Rouan *et al.* 2023; Poro *et al.* 2023). It then evaluated the plasticity of coral reef ecosystems facing climate change and contributed to predicting their response in the future."

This ambitious project aimed to provide a reference of the biological state of modern coral reefs in the Anthropocene. Among the new insights coming from The Tara Pacific program, we raised the entire microbial and chemical diversity of the coral holobionts at a basin-wide scale (Galand *et al.* 2023, Hochart *et al.* 2023, Veglia *et al.* 2023). It also highlighted the multiple strategies of corals selected during the evolution to respond to the environment fluctuation (Armstrong

et al. 2023; Cansi *et al.* 2023; Voolstra *et al.* 2023; Rouan *et al.* 2023; Porro *et al.* 2023). It then evaluated the plasticity of coral reef ecosystems facing climate change and contributed to predict their response in the future.

Keywords: *coral reefs, climate change, metagenomics, Tara Pacific expedition*

References

Allemand, D., Ferrier-Pagès, C., Furla, P., et al. (2004). Biomineralization in Corals: The Role of the Symbiotic Algae. In: Biomineralization (pp. 71–100). Springer, Berlin, Heidelberg.

Armstrong, E.J., Lê-Hoang, J., Carradec, Q., *et al*. (2023). Host transcriptomic plasticity and photosymbiotic fidelity underpin *Pocillopora* acclimatization across thermal regimes in the Pacific Ocean. *Nature Communications*, 14 (1): 3056.

Belser, C., Poulain, J., Labadie, K., *et al*. (2023). Integrative omics framework for characterization of coral reef ecosystems from the Tara Pacific expedition. *Scientific Data*, 10 (1): 326.

Bindoff, N.L., Cheung, W.W.L., Kairo, J.G., *et al*. (2019). IPCC special report on the ocean and cryosphere in a changing climate. IPCC, pp. 477–587.

Canesi, M., Douville, É., Montagna, P., *et al*. (2023). Differences in carbonate chemistry up-regulation of long-lived reef-building corals. *Scientific Reports*, 13 (1): 11589.

Galand, P.E., Ruscheweyh, H.J., Salazar, G., *et al*. (2023). Diversity of the Pacific Ocean coral reef microbiome. *Nature Communications*, 14 (1): 3039.

Hochart, C., Paoli, L., Ruscheweyh, H.J., *et al*. (2023). Ecology of *Endozoicomonadaceae* in three coral genera across the Pacific Ocean. *Nature Communications*, 14 (1): 3037.

Iwasaki, S., Inoue, M., Suzuki, A., *et al*. (2016). The role of symbiotic algae in the formation of the coral polyp skeleton: 3-D morphological study based on X-ray microcomputed tomography. *Geochemistry, Geophysics, Geosystems*, 17 (9): 3629–3637.

Lombard, F., Bourdin, G., Pesant, S., *et al*. (2023). Open science resources from the Tara Pacific expedition across coral reef and surface ocean ecosystems. *Scientific Data*, 10 (1): 324.

Planes, S., Allemand, D., Agostini, S., *et al*. (2019). The Tara Pacific expedition—A pan-ecosystemic approach of the "-omics" complexity of coral reef holobionts across the Pacific Ocean. *PLoS Biology*, 17 (9): e3000483.

Porro, B., Zamoum, T., Forcioli, D., *et al*. (2023). The Oak and the Reed: Two Environmental Response Strategies in Corals from Pacific Islands. *Communications Earth & Environment*, 4 (1): 311.

Reaka-Kudla, M.J. (1997). The global biodiversity of coral reefs: a comparison with rain forests. In: Reaka-Kudla, M., Wilson, D.E. and Wilson, E.O. (eds), *Biodiversity II: Understanding and Protecting our Biological Resources*. National Academies Press: Washington, DC, Chapter 7.

Richier, S., Furla, P., Plantivaux, A., *et al*. (2005). Symbiosis-induced adaptation to oxidative stress. *Journal of Experimental Biology*, 208 (2): 277–285.

Rouan, A., Pousse, M., Djerbi, N., *et al*. (2023). Telomere DNA length regulation is influenced by seasonal temperature differences in short-lived but not in long-lived reef-building corals. *Nature Communications*, 14 (1): 3038.

Veglia, A.J., Bistolas, K.S., Voolstra, C.R., *et al*. (2023). Endogenous viral elements reveal associations between a non-retroviral RNA virus and symbiotic dinoflagellate genomes. *Communications Biology*, 6 (1): 566.

Voolstra, C.R., Hume, B.C., Armstrong, E., *et al*. (2023). Disparate patterns of genetic divergence in three widespread corals across a pan-Pacific environmental gradient highlights species-specific adaptation trajectories. *npj Biodiversity*, 2 (1): 15.

Bayesian Approaches for Imputing Missing Data in Environmental Studies: Hierarchical Models and Kalman Filter Techniques

Zeki Bora Ön and Sena Akçer-Ön

Muğla SK Üniversitesi, Jeoloji Mühendisliği Bölümü, Türkiye

Abstract

Missing data present significant challenges in environmental sciences, particularly when accurately characterizing environmental dynamics. Imputing missing values in nonlinear systems, common in environmental measurements, requires incorporating all relevant and available data. Without a comprehensive approach, imputed measurements risk failing to capture the true environmental dynamics or introducing biases into subsequent analyses. In this paper, we present three Bayesian approaches to impute missing data, explicitly incorporating measurement uncertainties and uncertainties arising from model assumptions while utilizing all available data. First, we demonstrate Bayesian hierarchical regression, which effectively accounts for the nested structure typical of environmental datasets. Next, we explore hierarchical regression models incorporating change points to represent gradients or boundaries within the data. Finally, we examine Bayesian implementations of Kalman filters enhanced by Bayesian variable selection methods.

Keywords: Bayesian models, hierarchical models, missing data, environmental science, uncertainty quantification, imputation

Introduction

Measurement data often contains missing values, and imputation becomes particularly complex for data sampled in nonlinear mediums or gradients (see d'Auria *et al.*, this volume) that defy easy mathematical characterization, such as those encountered in environmental studies.

Statistical approaches to imputation should leverage all relevant data, but this inevitably increases model dimensionality and complexity, making it essential to quantify uncertainty accurately. Since statistics is the art of quantifying uncertainty, the Bayesian framework is particularly well suited to this task. Bayesian methods that are typically implemented via hierarchical (multilevel) models incorporate assumptions and prior information into probability distributions to offer a coherent, nuanced representation of complexity and uncertainty.

Bayesian statistics, notable for its ability to update event probabilities based on prior information and observed data, offers a versatile framework for data analysis. Its incorporation of prior information distinguishes it from frequentist statistics, facilitating the integration of outcomes from previous studies or expectations into analyses. Particularly advantageous for relatively small datasets, such as those typical in environmental studies, the Bayesian approach allows for dynamic updating of results with new data, enhancing adaptability and robustness.

The unified framework of Bayesian statistics addresses a wide range of statistical challenges, such as parameter estimation, model comparison, and hypothesis testing. It excels in handling complex models, including hierarchical structures and missing data problems, while offering a natural approach to quantifying uncertainties through probability distributions. Additionally, the ability to define all components of a model within the same framework provides a holistic approach, facilitating comprehensive assessment of uncertainties associated with both

measurement datasets and the model itself. Readers unfamiliar with Bayesian approaches and hierarchical models can refer to Kruschke (2015), Gelman *et al.* (2013) or Gelman and Hill (2006) for further insights.

In this context, we briefly present three examples from our previous works that apply Bayesian techniques to impute missing data. These methods assume that missing values depend on available information and treat each observation as a random sample contributing to the overall result. First, we describe two multilevel Bayesian approaches, which naturally arise when modeling missing data as conditional on observed covariates. A multilevel model specifies hierarchical distributions at both upper and lower levels that inform one another, integrating information across strata and improving predictive power by shrinking extreme low-level estimates toward the group mean (Kruschke and Liddell, 2018). Finally, we showcase a Bayesian Kalman-filter solution for predicting missing values.

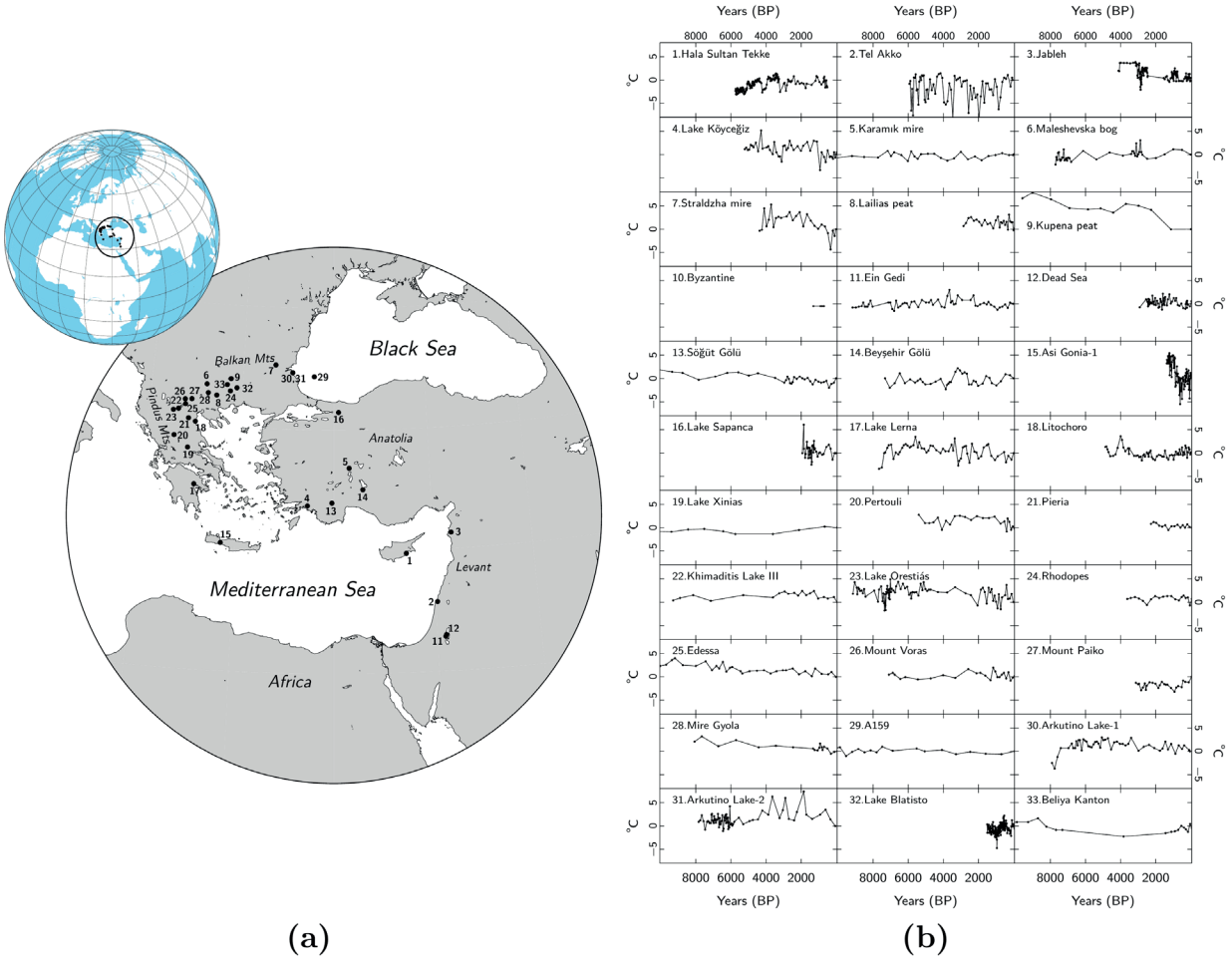
Examples of Bayesian Models

Hierarchical Models

Bayesian Hierarchical Regression

Consider a dataset of direct or proxy measurements of the same underlying quantity, collected from multiple locations over time, or depths, or across gradients of another variable. For the purpose of this study, let's assume that these measurements are expected to exhibit similar patterns, potentially due to collection in comparable environments or clustering within regions anticipated to display analogous trends. It is important to note that some of these measurements may be more complete than others, characterized by higher resolution in time or depth, or simply possessing more data points. For example, different temperature reconstructions of the Holocene in the eastern Mediterranean from pollen counts depicted in Figure 1 (Ön *et al.*, 2023) exemplify such variations in completeness.

In our study, our objective was to reconstruct an "average" temperature anomaly in the eastern Mediterranean region over the past 10,000 years through a meta-analysis, acknowledging the different resolutions, variabilities, and trends observed in each dataset when viewed collectively (Ön *et al.*, 2023). We utilized temperature reconstructions derived from pollen counts from various sites (Figure 1). To align with our research objectives, we segmented the data into 300-year intervals. Subsequently, we applied a Bayesian hierarchical linear regression model to analyze the dataset. In this hierarchical regression model, regressions are calculated for all individual samples (Figure 2a), while also incorporating an upper-level regression within the same model (Figure 2b).



(a) (b)

Figure 1. (a) Locations of pollen sites in the eastern Mediterranean. (b) Various temperature reconstructions derived from pollen counts. For details and references regarding each dataset, please consult Ön *et al.* (2023), from which the figure is adapted.

A key advantage of this multilevel model is partial pooling, where information flows bidirectionally between individual-level and population-level parameters. For example, units with sparse data (*e.g.,* Unit 5 and Unit 20 in Figure 2a) borrow strength from the overall distribution, yielding reasonable regression fits and well-calibrated uncertainty estimates. Units with outlying trends (*e.g.,* Unit 31) can still deviate when their data provide strong evidence, since the hierarchical prior shrinks—but does not override—their estimates. Intermediate cases (*e.g.,* Unit 16) strike a balance between local variation and the population mean, avoiding overfitting while preserving genuine signal. In a setup like this (see Figure 2), the hierarchical model naturally regularizes noisy or under-sampled datasets through shared regression parameters and can, if necessary, impute missing observations seamlessly within the same inference framework.

Bayesian Hierarchical Change Point Analysis

Consider again multiple measurements collected within a localized region or across similar environments exhibiting comparable patterns over time, depth, or other dimensions. These patterns manifest as shifts in the mathematical functions describing the data. For example, refer to the $^{(239+240)}$Pu measurements presented in Figure 3, obtained from various environments (for dataset references, see Waters *et al.*, 2023 and references therein).

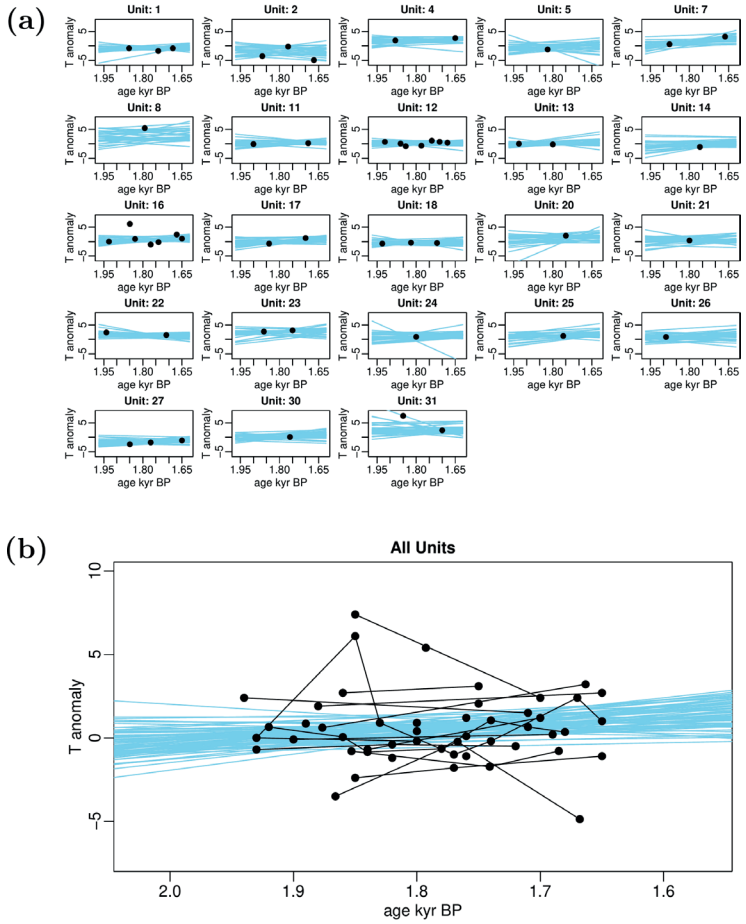


Figure 2. Linear regression curves of the hierarchical model for the period between 1650–1950 calibrated years before present. (a) For low-level regression parameters, (b) For high-level regression parameters. Each plot in (a) represents a different measurement in the region depicted in Figure 1a. The blue bunch of lines represents random regression solutions generated from posterior draws. The figure is adapted from Ön *et al.* (2023).

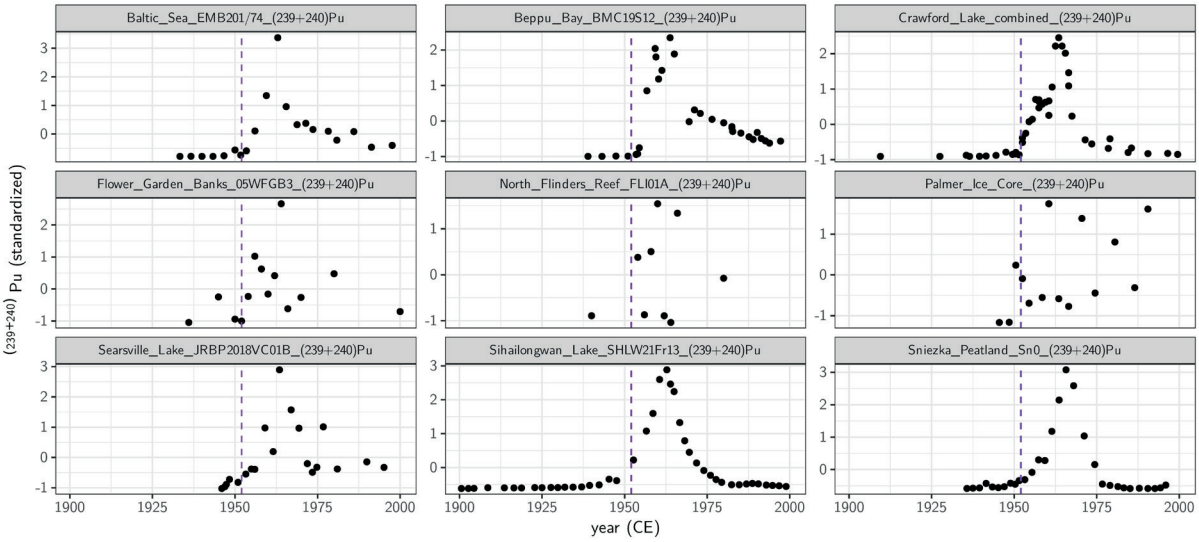


Figure 3. $^{(239+240)}$Pu measurements from various environments and locations serve to pinpoint the onset of the Anthropocene. The purple dashed line denotes the year 1952 CE, marking the occurrence of the first H-bomb testing. The figure is adapted from Ön *et al.* (2025).

Our objective with these data was to critically evaluate the designation of 1952 CE as the onset of a potential Anthropocene epoch using a Bayesian hierarchical change point model.
As anticipated, the measurements are imperfect due to several factors, including sample quality and resolution, uncertainties inherent in the measurement processes, and variations in the preservation of proxy signals under differing environmental conditions. While some records clearly show an abrupt increase followed by a somewhat gradual decline, others exhibit less distinct patterns.

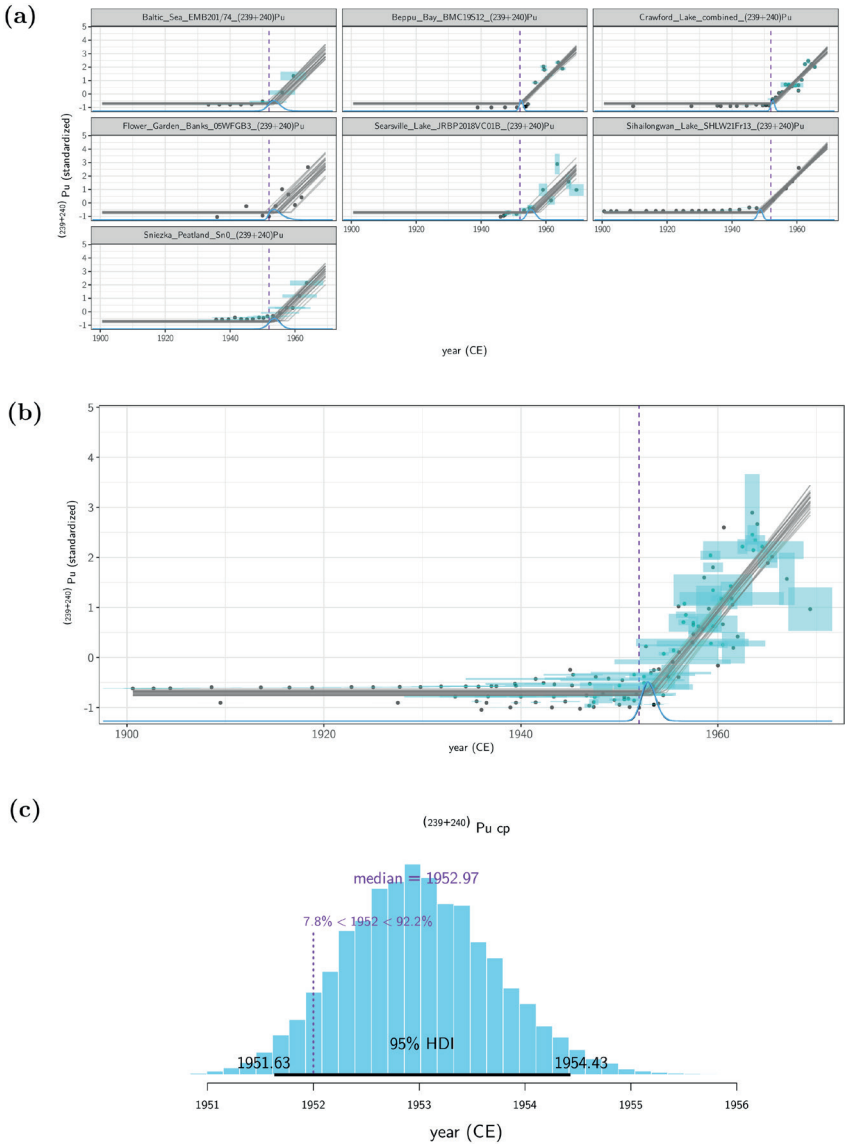


Figure 4. The hierarchical change point analysis results for $^{(239+240)}$Pu data are presented across distinct datasets (a), for the overall change point (b), and the posterior distribution of the overall change point (c). Our assumption for this dataset was a constant level followed by a ramp. Each dataset was truncated after approximately 1965 CE, aligning with the model described in Ön *et al.* (2025). Cyan boxes in (a) and (b) denote one standard deviation uncertainty intervals associated with dates and measurements. Gray lines represent random solutions generated from posterior draws. The purple dashed lines indicate 1952 CE, while the black bar below the posterior distribution in (c) denotes the 95% highest density interval. The figure is adapted from Ön *et al.* (2025).

To delineate the start of the increase, we removed the data after around 1970 CE (for details, consult Ön *et al.*, 2025). Applying a hierarchical change point analysis using an errors-in-

variables approach, which accounts for the associated uncertainties in both proxy and dating measurements, can effectively regularize the data (Figure 4). Consequently, it can serve as a useful tool for imputing missing data in this type of dataset.
Notice how uncertainties are integrated into the model and how their impact is reflected in the increasing uncertainty of the regression and change-point parameters for more variable records (see Figure 4). Furthermore, even when the exact location of the change point is unclear in certain datasets, the hierarchical model helps pinpoint its location by shrinking the posterior distributions toward the group mode through shared information between higher-level and lower-level parameters.

Kalman Filter Through Variable Selection

Missing data in environmental time series present a significant challenge, particularly for time series analyses, which often require evenly spaced data. For example, missing data may arise in certain gauge or climate stations, while neighboring stations possess complete datasets. Imputation methods based solely on the inherent structure of the time series with missing data may overlook the non-stationary nature of environmental data. Therefore, filling these gaps through a unified approach that considers both the inherent structure of the data with missing values and the available complete data should be the preferred method.

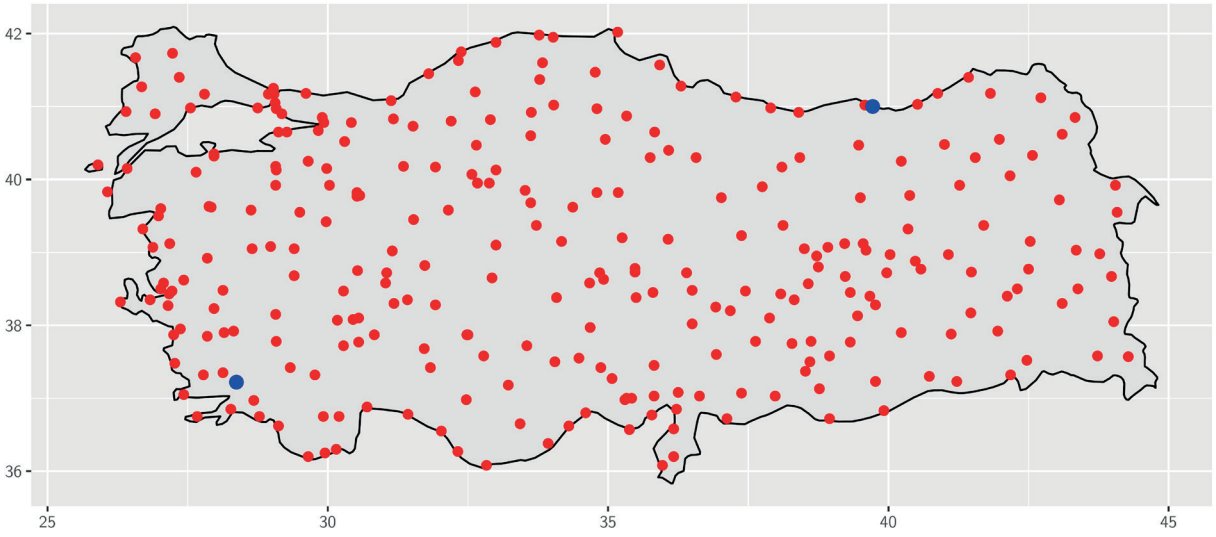


Figure 5. The locations of the meteorology stations used in the study. Muğla and Trabzon stations are given in blue color.

Turkey can be described as a high plateau peninsula, with a mean altitude of approximately 1,132 m, covering a relatively large land area of approximately 785,000 km^2. Primarily, tectonically induced topographic gradients, coupled with proximity to water bodies of varying characteristics, result in diverse climatic conditions throughout the country (Ünal *et al.*, 2003; Erinç, 1984).

In our study, we utilized monthly precipitation and temperature station data from Turkey, sourced from Göktürk (2005)[1]. These meteorological stations offer comprehensive spatial

[1] The data may have been derived from daily records, with missing values potentially imputed using expectation maximization, a widely employed technique in climatology (Schneider, 2001). However, this preprocessing step does not affect the objectives of our study. Please note that we cannot definitively confirm the specific imputation method used; this footnote serves as an acknowledgment of uncertainty regarding the imputation procedure.

coverage of the entire country (Figure 5). Specifically, we selected two stations, Muğla and Trabzon, as case studies. Muğla, located in Turkey's southwestern corner and surrounded by high mountains, predominantly receives precipitation from Mediterranean cyclones during winter and experiences a typical Mediterranean climate (wet, mild winters and dry, hot summers). Trabzon, situated on the northern slopes of the east-west trending Eastern Pontides, experiences a Black Sea climate characterized by maritime influences, relatively warm winters, and consistent precipitation year-round.

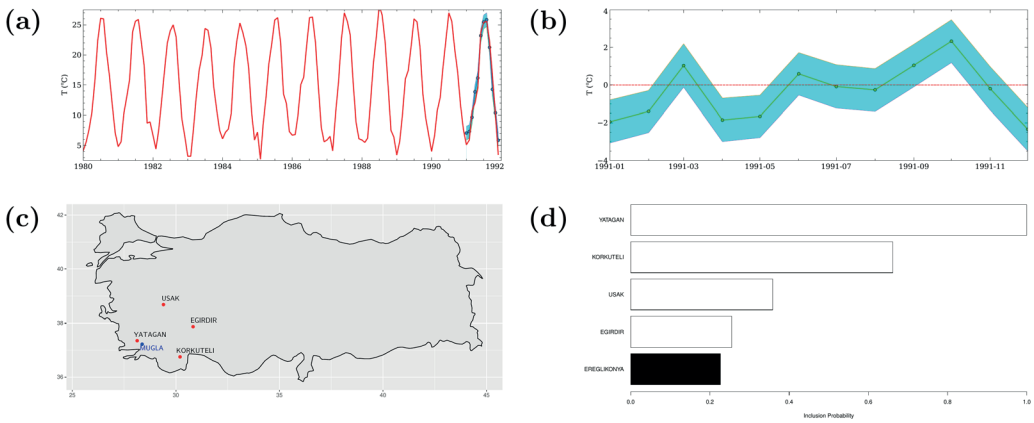


Figure 6. Results of temperature measurements at Muğla station. (a) Imputed data (blue curve) overlaid on measurement data (red curve), with 95% credible interval (cyan shading); (b) Difference between original and imputed data; (c) Map showing Muğla station (blue dot) and predictors with highest inclusion probabilities (red points); (d) Posterior inclusion probabilities of predictors assigned by Markov chain Monte Carlo in the regression model. White bars represent positive coefficients, and black bars represent negative coefficients.

For data imputation, we implemented a time series model along with a regression model using covariates derived from the remaining contemporaneous station data. The time series model incorporates a seasonal process to account for yearly oscillations and an autoregression process of order three.

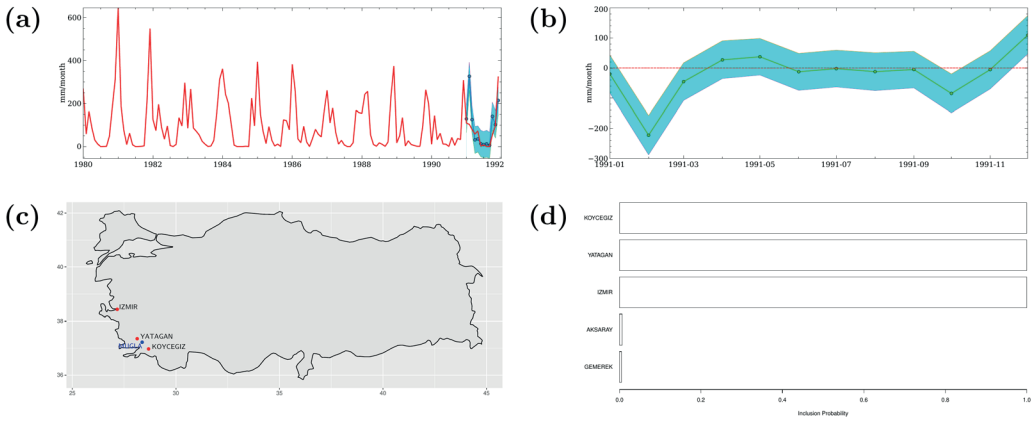


Figure 7. Results for precipitation measurements from Muğla station, similar to Figure 6.

Additionally, we applied the spike-and-slab variable selection (SSVS) technique (Scott and Varian, 2014; George and McCulloch, 1997; O'Hara and Sillanpää, 2009) , which serves to reduce the size of the covariate set and effectively cluster covariates showing similar trends

with the target series. In this analysis, the solution is handled through a fully Bayesian model, and therefore the components of the model and weights of each covariate are found through the Bayesian model itself.

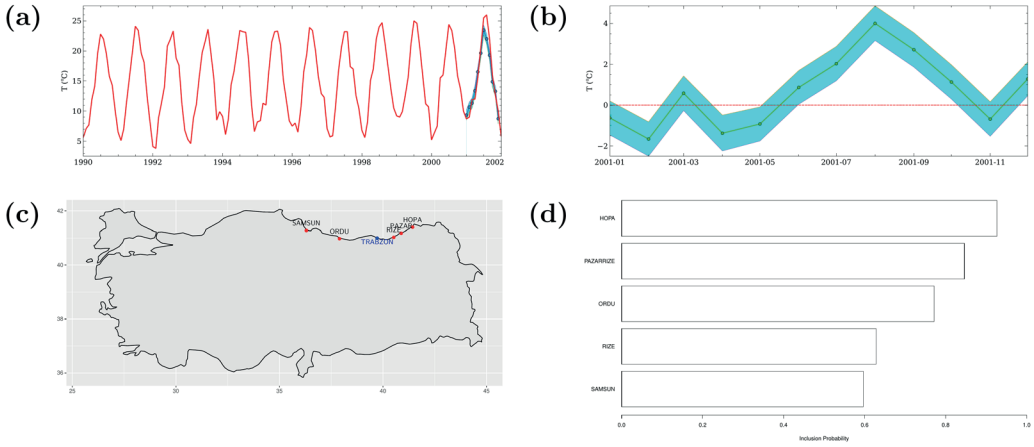
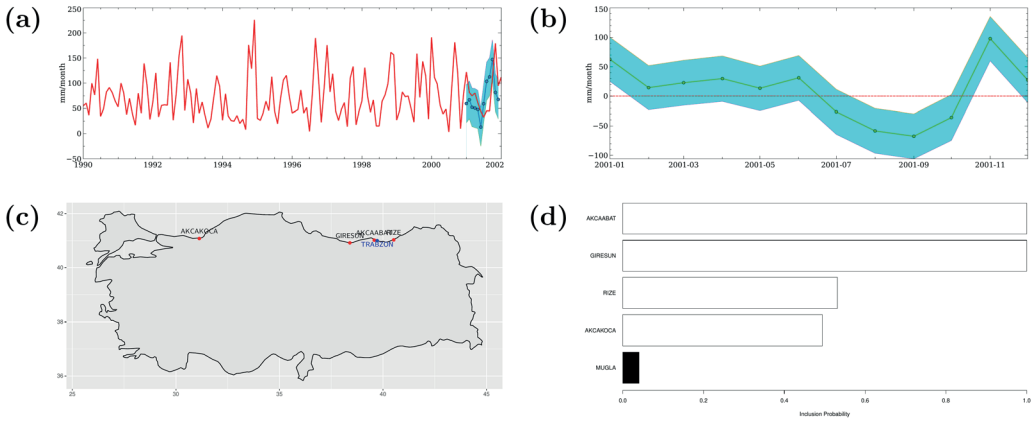


Figure 8. Results for temperature measurements from Trabzon station, similar to Figure 6.

For imputation validation, we deliberately excluded the entire year 1991 CE from the Muğla monthly temperature and precipitation records. Specifically, for the model to adapt the coefficients, we utilized Muğla data from 1980 to 1990, while employing data from 1980 to 1991 for the remaining stations. The results of the imputation for the entire year of 1991 are depicted for temperature in Figure 6a and for precipitation in Figure 7a.

Similarly, we excluded the entire year 2001 CE from the Trabzon monthly temperature and precipitation data for imputation purposes. Specifically, for the model to adapt the coefficients, we utilized Trabzon data from 1990 to 2000, while employing data from 1990 to 2001 for the remaining stations. The results of the imputation for the entire year of 2001 are presented for temperature in Figure 8a and for precipitation in Figure 9a.



Fiure 9. Results for precipitation measurements from Trabzon station, similar to Figure 6.

The results of the yearly imputations produced noteworthy results (Figures 6b, 7b, 8b, and 9b). Furthermore, the outcomes of the SSVS method provided insightful findings regarding covariate clustering. The primary covariates identified by SSVS for the regression of Muğla temperature and precipitation (see Figures 6c, 6d, 7c, and 7d) align closely with climatological

clusters identified by Ünal *et al.* (2003), specifically corresponding to the Eastern Mediterranean and Aegean climate zones. Similarly, results from the Trabzon analysis (Figures 8c, 8d, 9c, and 9d) identified a covariate cluster confined to the northern slopes of the Pontides, corresponding to the Black Sea climate zone described by Ünal *et al.* (2003).

Concluding Remarks

There are already dedicated software packages for missing data imputation. However, in the case of more complex datasets, such as those encountered in environmental studies, Bayesian techniques are recommended. These techniques effectively incorporate multiple sources of uncertainty and fully leverage all available data.

Finally, the utilization of Bayesian techniques, particularly those that incorporate total evidence through hierarchical structures, represents a pivotal step forward in addressing the persistent challenges posed by missing data in environmental studies. By explicitly integrating uncertainties and facilitating information sharing across datasets, these sophisticated approaches provide exceptional insights into complex environmental systems.

References

D'Auria, G., El Bour, M., Furla, P., *et al.* (2025). An ocean of gradients - towards a 3D mapping and visualization of high-resolution marine data – An overview. In: Giuliano, L. and Rodriguez y Baena, A. (Eds.) *CIESM Monograph 53*, pp. 7–29. CIESM Publisher, Paris, Monaco.

Erinç, S. (1984). *Klimatoloji Ve Metodları*. İstanbul Üniversitesi Yayınları, İstanbul Üniversitesi.

Gelman, A., Carlin, J.B., Stern, H.S. *et al.* (2013). *Bayesian data analysis* (3rd ed.). CRC Press.

Gelman, A., Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

George, E.I., McCulloch, R.E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7(2), 339–373.

Göktürk, O.M. (2005). North Sea – Caspian Pattern and its influence on the hydrometeorological parameters over Turkey. Master's Thesis, İstanbul Teknik **Üniversitesi**, Avrasya Yer Bilimleri Enstitüsü, İstanbul.

Kruschke, J.K. (2015). *Doing Bayesian Data Analysis* (2nd ed.). Academic Press, Boston.

Kruschke, J.K., Liddell, T.M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25(1), 178–206.

O'Hara, R.B., Sillanpää, M.J. (2009). A review of Bayesian variable selection methods: what, how and which. *Bayesian Analysis*, 4(1), 85–117.

Ön, Z.B., Ateş, M.E., Kaiser, J. (2025). A conceptual and statistical framework for delineating the timing of a stratigraphic transition: Holocene-Anthropocene boundary as a case study. *Progress in Physical Geography: Earth and Environment*, 49(1-2), 63–83.

Ön, Z.B., Macdonald, N., Akçer-Ön, S. *et al.* (2023). A novel Bayesian multilevel regression approach to the reconstruction of an eastern Mediterranean temperature record for the last 10 000 years. *The Holocene*, 33(7).

Schneider, T. (2001). Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate*, 14(5), 853–871.

Scott, S.L., Varian, H.R. (2014). Predicting the present with Bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, 5(1-2), 4–23.

Ünal, Y., Kindap, T., Karaca, M. (2003). Redefining the climate zones of Turkey using cluster analysis. *International Journal of Climatology*, 23(9), 1045–1055.

Waters, C.N., Turner, S.D., Zalasiewicz, J. *et al.* (2023). Candidate sites and other reference sections for the Global boundary Stratotype Section and Point of the Anthropocene series. *The Anthropocene Review*, 10(1), 3–24.

Web Based 3D Research Data Visualization

Kaveh Rassoulzadegan

Mediterranean Science Commission (CIESM), Monte-Carlo, Monaco

Abstract

The display of 3D scientific datasets within web browsers can be enhanced through automated techniques designed to balance ergonomic data access, interactive rendering performance, and visual clarity. This article presents a comprehensive overview of web 3D research data visualization methods, utilizing a proof-of-concept main application to address oceanographic dataset access and rendering complexities across various scenarios. A series of more specific ocean science prototypes are dedicated to the visualization of datasets from distinct disciplines (physical, seismic, metagenomic), demonstrating development modularity via an open-source and permissive software toolchain. The results indicate a clear potential for 3D, interactive, and accessible multidisciplinary scientific visualizations directly on the web, employing JavaScript Web Graphics Library 2 (WebGL2) without the need for any third-party software component installation, thus facilitating improved understanding and analysis of marine environments.

Keywords: Web-based 3D scientific visualization, GPU-accelerated volume rendering, out-of-core data management, interactive data querying and processing, multiresolution and adaptive sampling techniques

Introduction

Earth science research using digital twins relies on sophisticated data viewers capable of requesting and interactively displaying n-dimensional datasets from various sources, utilizing their respective data transfer protocols. These viewers should also feature capabilities such as translating time-series data into animated representations.

Massive data visualization presents various challenges in scientific analysis applications, particularly when employing web-based 3D render engines due to raw performance considerations. However, these solutions offer significant advantages. The primary advantage of web-based solutions lies in their ease of deployment compared to native applications. Native applications often require the program to run on multiple hardware targets, which can be tedious to optimize for different processor architectures and their specific computational resources. Targeting a web browser frees developers from these concerns, benefiting from well-standardized codebases and the support of growing communities.

A few decades ago, the computing horsepower of browser JavaScript engines[2] could barely render complex geometries efficiently. Nowadays, however, the support of graphical processing units (GPU) substantially boosts performance.

Previously, browsers had quite modest access to hardware acceleration offered by processing units, but this is less of a concern now.

2 JavaScript engines are the set of functions letting compile JavaScript instructions within web browsers.

Massive scientific models, based, for example, on oceanographic or atmospheric datasets, require efficient acceleration structures[3] to navigate them fluidly. In this context, WebGL[4] has become a pivotal technology, essentially transforming 3D rendering on the internet to be almost as capable as high-end medical visualization equipment or latest-generation gaming consoles in terms of performance. Running more computationally intensive transfer functions in real time is now possible thanks to direct access to the GPU, eliminating the need for third-party browser software components such as early Virtual Reality Modeling Language (VRML[5]) viewers or older Java-based 3D engines.

Thanks to the multidisciplinary nature of the Mediterranean Science Commission and its six scientific committees, a discussion was initiated regarding the development of an innovative tool. This tool aims to enable more intuitive visualization of any dataset from any discipline within a single application, while minimizing both human and material resources for its development. The purpose of this article is to describe the development of a new CIESM tool (Figure 1) from its conceptualization and inception through to its initial testing phases, and to present some early applications using relevant scenarios.

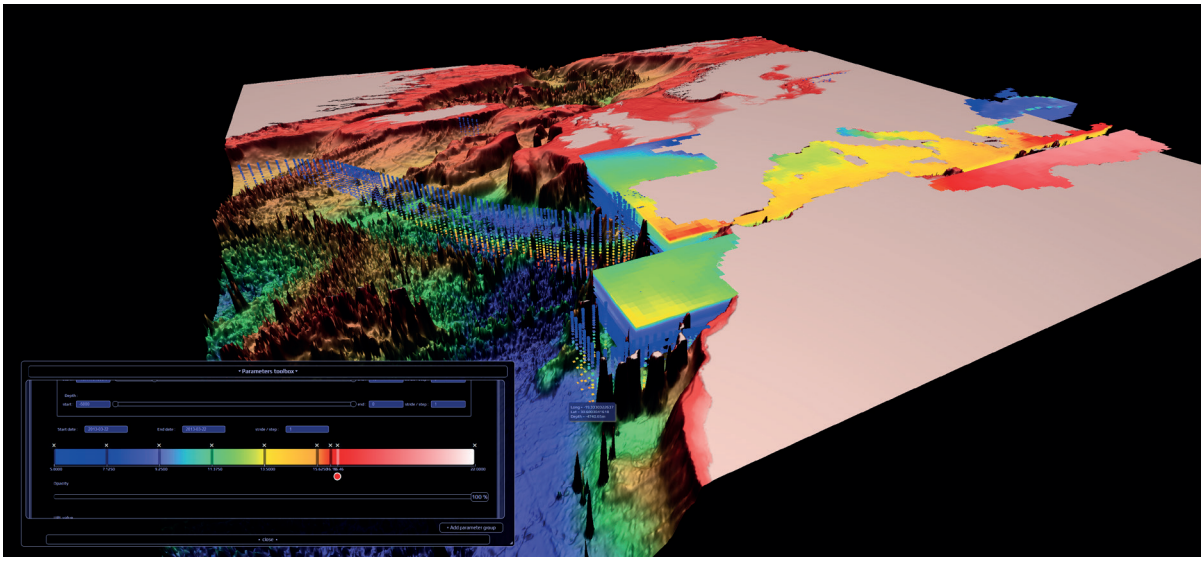


Figure 1. Example of the CIESM tool showing combined bathymetric grid (GEBCO and EMODNET 2020), gridded 3D aggregations of March 2013 temperature and salinity (NOAA National Centers for Environmental Information, 2013), and a tabulated collection of sampled silicate concentrations during different oceanographic campaigns (EMODNET Physics, 2020).

Background and Related Work

Early rendering techniques were mostly ray-based approaches before the latest hardware-friendly parallel polygon rasterization[6] methods, which provide more predictable performance in extreme scenarios.

3 Acceleration structures are spatial subdivision and indexing strategies designed to progressively load regions of interest (ROIs) instead of an entire, overly heavy dataset at once.
4 WebGL (Web Graphics Library) is a JavaScript API for rendering interactive 3D graphics in web browsers using the GPU.
5 VRML (Virtual Reality Modeling Language) is an older standard for describing 3D scenes and objects in a text-based format, primarily used for web-based virtual environments.
6 Rasterization converts projected primitives into discrete fragments (potential pixels), which are then shaded and displayed on the screen.

Since the mid-90s (when online interactive 3D visualization was rare), native real-time 3D applications have remained the most performant. On the other hand, thanks to modern processors, the performance gap between native and web-based applications is becoming almost negligible when it comes to displaying 3D data, among other capabilities.

For desktop solutions, lower-level languages such as Assembler, C, and C++ (possibly with the use of specific CPU intrinsics), are largely preferred for maximizing hardware resources and efficiently dispatching jobs across available units, especially when targeting embedded systems rather than consumer devices. These toolchains become substantially heavier when targeting and maintaining solutions on multiple platforms.

Higher-level languages were initially less performant due to processing resources reserved for intermediate components like virtual machines or other command interpreters. However, they allowed for more human-friendly instruction writing, with the benefit of providing highly reliable execution on most platforms and the convenience of a single source code base to maintain.

Among these higher-level languages, JavaScript (or European Computer Manufacturers Association – ECMAScript[7]) compilers within web browsers have particularly improved, making them viable development targets in terms of performance. Many non-browser environments also execute this scripting language.

In contrast to native applications, and despite their initially inferior processing capabilities, the ease of instantaneous deployment on any web browser and the maintenance comfort of a unique source code make it an obvious choice to drastically reduce programming efforts by targeting web browsers with such technology.

The local desktop software installation process is gradually diminishing alongside the performance gap between native and web applications. The latter solutions also require less storage on both client and server sides and allow direct sharing of a complete and complex application via a single web address, without the need for any prior compilation and linking[8] steps.

Another major advantage of JavaScript is its ever-growing community, providing many libraries, including several very efficient interactive 3D ones. This offers a flexible choice of rendering techniques and easily allows for the design and fine-tuning of specific ones.

Regarding trends in rendering performance, the arrival of WebGL was a significant advancement for developers needing full GPU access to display massive 3D geometries inside a web browser.

Methods and Implementation

Design Considerations

Finding the optimal and most cost-effective balance between high rendering performance, code modularity, user-friendly interactivity, and platform portability is a complex challenge. Exchanging data between hosts and users (Figure 2) for 3D visualization involves fast conversion steps to properly render geometries reconstructed from point cloud models.

7 ECMAScript is the standardized scripting language specification underlying JavaScript, defining its core syntax, types, and runtime behavior.

8 In traditionally developed applications, the linking step joins separately compiled modules, resolving external dependencies to produce the final binary executable.

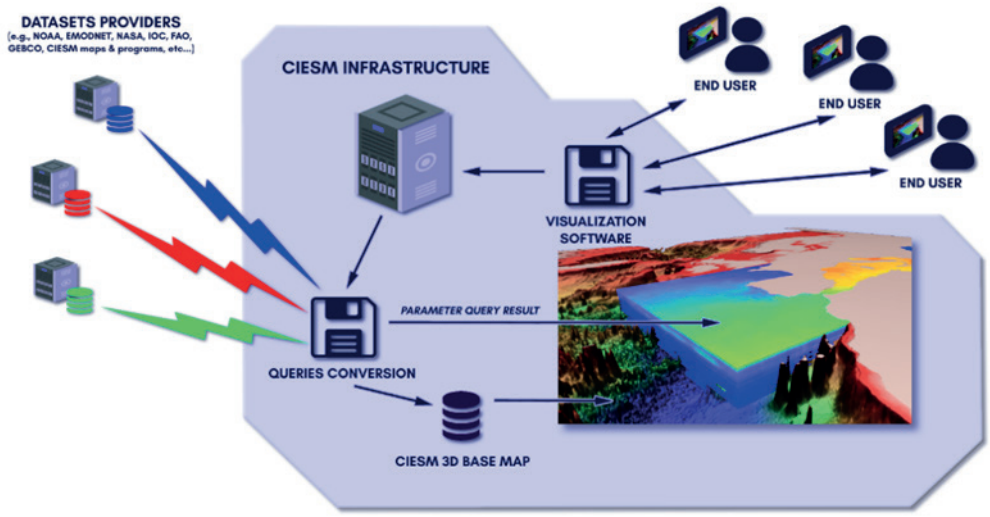


Figure 2. Framework overview illustrating the integration of required hardware and software resources. This diagram shows how end users connect with CIESM-proposed data querying, conversion, and rendering functions, and the available data host infrastructures.

While still far from being able to process/transform/reconstruct raw data in real-time across various research areas, modern domestic computers and workstations are increasingly capable of handling heavy client-side computation through different standalone or more distributed strategies, depending on the required computational resources.

After an extensive testing phase of various JavaScript 3D engines, the 'Three.js' library was selected because it offers a well-abstracted Application Programming Interface (API) and supports numerous popular interchange formats, enabling the loading of a wide range of variously structured geometries within the main scene. This library is a mature solution, developed over 15 years, implementing a WebGL renderer and operating under an MIT license scheme. These permissive license schemes are crucial for guaranteeing complete freedom in code usage and modification.

Such a toolchain allows for efficient computation of rendering operations in the vertex and fragment pipelines through WebGL 2.0 (OpenGL ES 3.0), as well as utilizing WebGPU[9] compute shaders[10] for advanced modification of the topology before the final fragment pass.

Proof of Concept Application

To prepare streams (vertex positions, normal vectors, face index buffers, texture coordinates, etc.) from query results received from various science data servers before transmitting them to the GPU vertex pipeline[11], the response data must be interpreted according to dataset metadata. This ensures the received data is accurately converted into a suitable and efficient 3D rendering asset.

9 WebGPU is a low-level API for high-performance graphics and parallel computing.
10 Shaders contain code sections (generally vertex and fragment shaders) intended to be executed on the GPU.
11 The vertex pipeline transforms vertex data (e.g., positions, normals) from model space to clip space, often applying operations like skinning or lighting, before passing them on for rasterization.

Once the conversion step is completed, the object is ready for further vertex topology and projection operations, before the final fragment pipeline[12] code is executed for applying textures to the geometry.
Bathymetric data serves as an example for producing an oceanographic 3D basemap. For effective rendering of polygonal grid displacements (Figure 4), the received data is converted into tiles (Figure 3) to enable out-of-core collections optimized for spatial indexing.

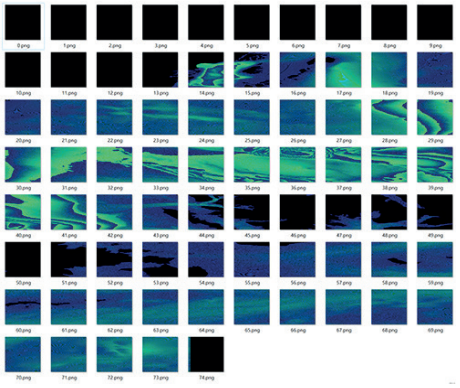


Figure 3. Tiles containing bathymetric raw values.

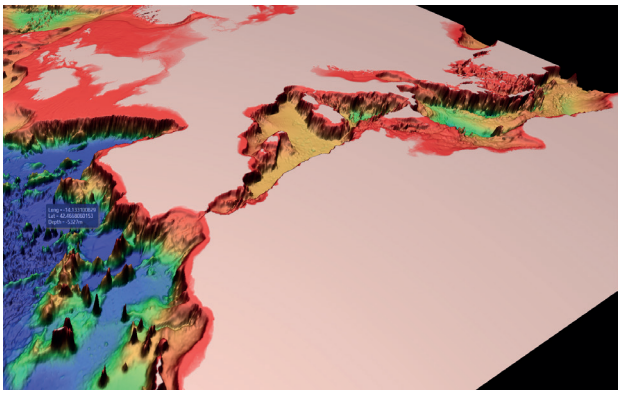


Figure 4. Resulting elevated grid.

Different types of primitives can be combined and shaded using various reasonably performant techniques. 3D layers composed of various primitives can be formed using methods ranging from simple geometry instancing setups to more advanced surface reconstruction techniques, such as Marching Cubes (Lorensen *et al.* 1987), Dual-Contouring (Ju *et al.* 2002), Dual Marching Cubes (Schaefer *et al.* 2004) and their corresponding rendering approaches, for wrapping and shading the received point clouds.

For each entry in the received query result, a user-defined primitive is cloned (*e.g.*, points, sprites[13], boxes, spheres, etc.), and its coordinates are set in the 3D scene graph[14]. The color of the instance is set according to the desired variable value by mapping its texture coordinates to a user-defined gradient through an editor. This editor is bound to a custom WebGL shader, which allows for stop color management and provides real-time updates in the 3D view for a more intuitive user experience.

Data Handling Pipeline

A user interface section for intuitively querying different data hosts, such as NOAA's Environmental Research Division's Data Access Program (ERDDAP) servers, is presented as a form (Figure 5).

12 The fragment pipeline processes fragments (potential pixels) generated by rasterization, applying shading, texturing, and depth/alpha tests to produce the final pixel color.
13 A sprite is a geometry that always faces the camera, typically used to render a bitmap texture or a coloring function.
14 A 3D scene graph is a hierarchical data structure that organizes and manages the spatial and logical relationships among objects in a 3D scene.

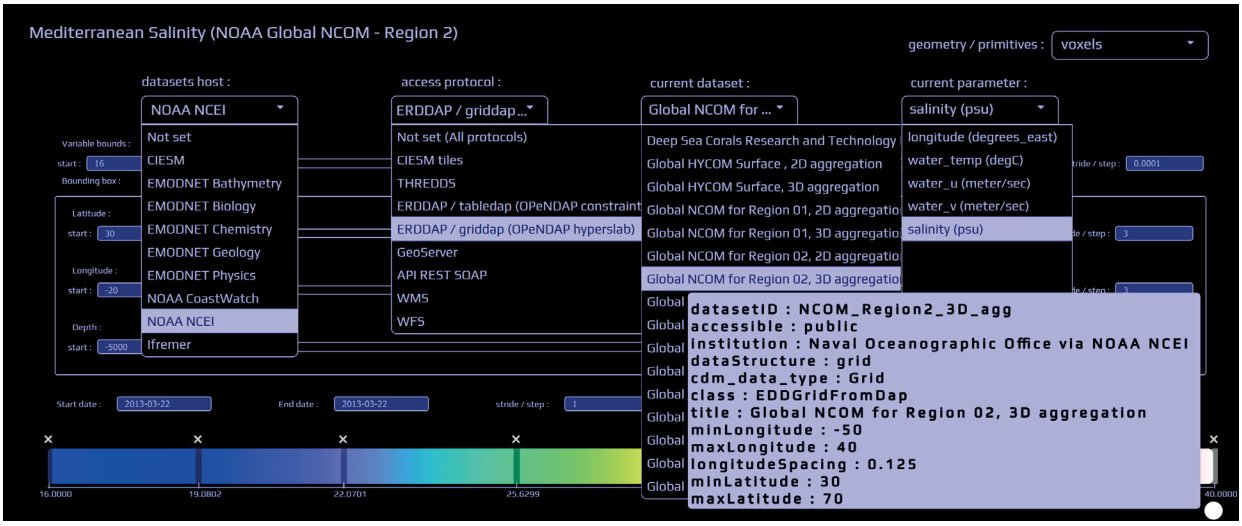


Figure 5. Query builder input fields.

This form enables users to easily build URL queries by selecting key fields necessary to issue a request, specifically:

- The data host.
- The data storage type and corresponding transfer protocol.
- The dataset of interest within the selected host (by consulting its metadata).
- The environmental variable(s) of interest within the dataset.
- The desired range of geographic and temporal coordinates.

With data-sharing portals capable of returning query results in various formats, such as JavaScript Object Notation (JSON) streams, the requested environmental variable is directly loaded into a JavaScript object, thereby enabling immediate access and/or evaluation, and facilitating subsequent filtering operations.

Both gridded and tabulated datasets can be queried, and tiles of different formats can be streamed through various web APIs.

Rendering techniques

Aside from traditional polygon rasterization, some datasets can benefit from better representation clarity and performance, depending on the type of primitive and data structure processed by the unit. For example, the geometry instancing capacity of GPUs has, over the years, enabled drawing an ever-increasing amount of cloned geometry for the cost of a single draw call.

In certain scenarios, Direct Volume Rendering (DVR) by raymarching[15] a 3D texture (stored as an array of 8- to 32-bit precision integer or floating-point values) can also be an efficient option. This technique is useful as the geometry relies only on arrays of scalar values, with spatial coordinates induced by the offset or position in the array.

15 Raymarching is a rendering technique that iteratively advances a ray through a scene or volume to sample data, commonly used for volume rendering and signed distance field visualization.

Another advantageous aspect of DVR is the modern GPUs' built-in hardware bilinear interpolation capability for either iso-surfaces or Maximum Intensity Projection (MIP[16])-based representations. Despite not being fully accelerated currently, other interpolation methods can also be easily implemented at the expense of a few extra samples gathered in the fragment pipeline, resulting in a relatively small increase in computational overhead (tokens[17]).

These methods also have hardware resource limitations, particularly when handling extremely high resolutions. Recent out-of-core approaches now enable high and stable framerates (*e.g.*, over 60 FPS at UHD resolutions) on both workstations and consumer platforms with large data arrays (*e.g.*, 1024^3 voxels[18]), although performance may degrade significantly at even higher resolutions.

Efficient Octree acceleration structures (Knoll *et al.* 2006; Laine *et al.* 2010; Crassin 2011) and consistent LOD strategies successfully solve memory limitation issues for surface reconstruction from point clouds. These solutions often involve nodes that can be compressed, potentially using a combination of advanced packing approaches (*e.g.*, Lefebvre *et al.* 2006) that vary depending on the sparsity level of specific dataset areas.

Dedicated Prototypes and Case Studies in Marine Science

Dedicated prototype 1: Seismic data visualization using DVR

Once seismic tomographic data is inverted through significant computations, models can be represented using a 3D texture, similar to how Nearly Raw Raster Data (NRRD[19]) files are viewed in the medical field. The idea was to represent the Earth's subduction zones from a global temperature anomaly model, and to quantify rigid versus softer areas from the surface down to 2815 kilometers below the surface.

A compact user interface was added to allow the user to isolate desired ranges of temperature anomalies and perform geometry clipping along the three axes, thereby enabling the user to narrow Regions of Interest (ROIs) within the volume. These clipping planes were also designed to be rotatable to occlude parts of the geometry from any azimuth.

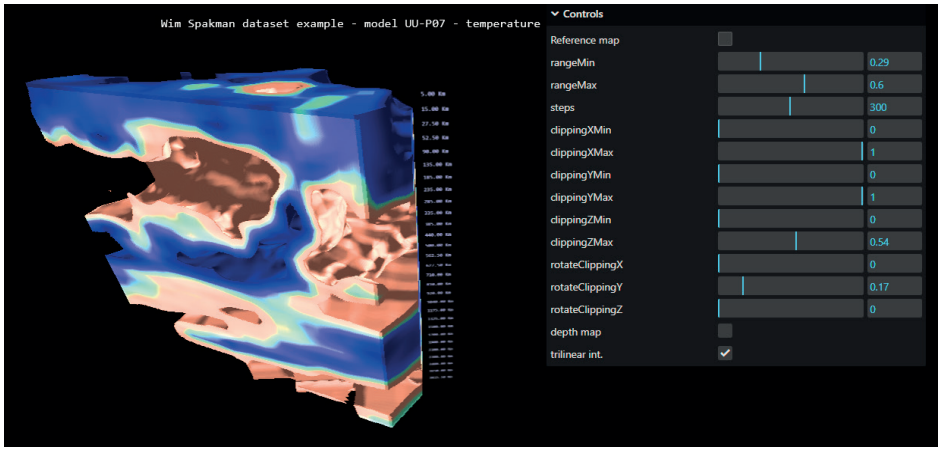


Figure 6. Calabrian subduction zone representation obtained by ray-marching a 3D texture derived from the global temperature anomalies tomographic UU-P07 model (Amaru *et al.* 2007; Hall *et al.* 2015).

16 The Maximum Intensity Projection (MIP) technique continues sampling through the entire volume, instead of abruptly stopping when the data value meets a chosen threshold, like with iso-surfaces representations.
17 In programming, a token is the smallest meaningful unit of code of a compiler or an interpreter during lexical analysis.
18 A voxel (volumetric pixel) is the smallest unit of a 3D grid representing spatial data, analogous to a pixel in 2D imaging.
19 Nearly Raw Raster Data (NRRD) is a file format often used in medical imaging and scientific visualization.

Dedicated Prototype 2: Metagenomic Data Visualization ("Metacube")

This sample application maps the microbial metabolism genomic catalogue by importing metagenomic data from a Kyoto Encyclopedia of Genes and Genomes (KEGG) database (*e.g.*, Kanehisa and Goto 2000), enabling immediate 3D representation and highlighting pathways that share the same reactions. It also allows for selecting and saving subsets of the loaded datasets via mouse clicks or through a search engine that enables quick entry and retrieval of specific entries. A specific shader is also attached to allow for various opacity levels, as well as clipping planes similar to those in Prototype 1.

Modularity And Open-Source Toolchain

As a modern and constantly improving programming language, JavaScript provides significant development modularity by eliminating the prior compilation and linking steps typical of more traditional compilers. This advantage enables the inclusion and/or loading of external and complementary code, which can also be done asynchronously where needed. The libraries can be loaded either locally or remotely, facilitating testing before further integration into specific solutions. In the proposed solution, "Three.js" (the 3D engine) is combined with "jQuery" for JavaScript syntax simplification, "jQuery UI" for user interface design, and a few other libraries for more specific tasks (*e.g.*, compression, 3D functionality, image processing extensions).

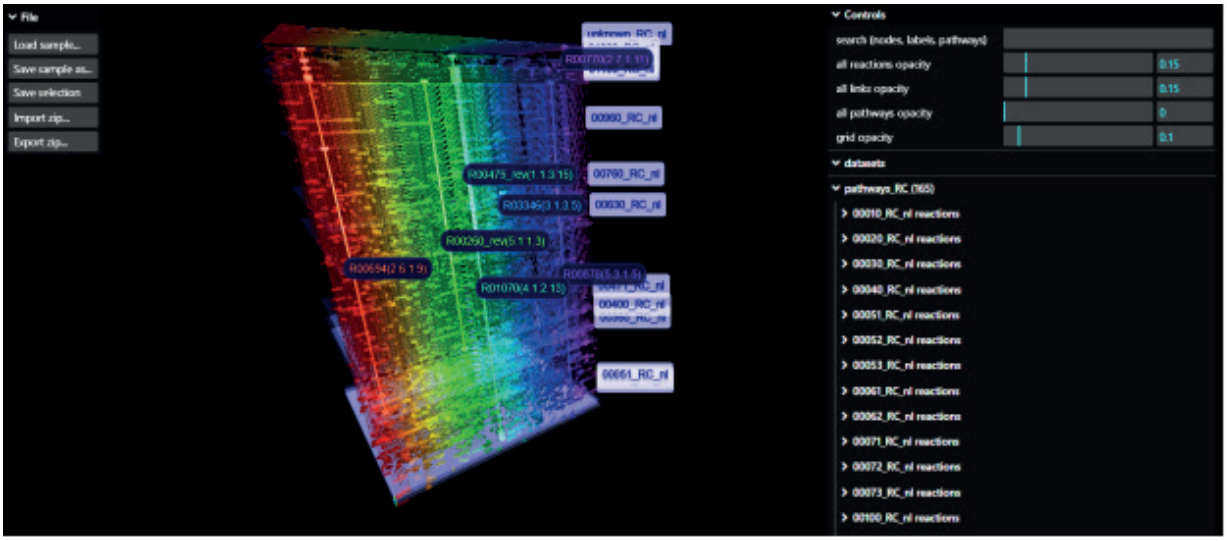


Figure 7. "Metacube" rendering output of reactions and pathways exported from the KEGG metagenomic database (Kanehisa *et al.* 2025). The prototype directly imports/ exports KEGG archives and loads/saves whole representations or subsets.

The open-source nature of the toolchain is central to ensuring ease of maintenance and evolution of the codebase. Permissive license schemes such as MIT, Apache, or FreeBSD ensure complete freedom in terms of code modification, evolution, and commercial distribution. This kind of license scheme is generally preferred among other open-source licenses, as it quickly clarifies to developers that any source code usage will not infringe upon any legal prerequisites.

Finally, through various open-source versioning platforms or repositories (such as GitHub), candidate revisions are easily screened by relevant communities, which significantly enhances maintenance and future evolution or extensions.

Results

Performance Evaluation

The presented main application and prototypes' rendering performance exceeds 60 frames per second (FPS) for 4K output when ray-marching volumes between 256^3 and 512^3 voxels with 8- to 16-bit precision 3D textures, using Mozilla Firefox or Google Chrome on a 2021 entry-level laptop computer. This performance increases to over 140 FPS when rendering 1024^3 volumes with 8- to 32-bit value arrays on current (2025) hardware setups, without additional advanced optimizations. The following profiles were conducted on three different Nvidia GPU models with 4K output (3840 x 2160).

Profiles in various performance areas show that the approach is effectively scaling across hardware generations and consequently features regular improvements.

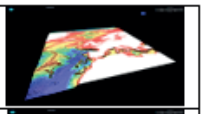
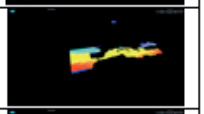
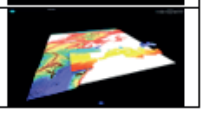
3D basemap	3.3	1048576	2093058	28 FPS@4K 36ms	141 FPS@4K 7ms	278 FPS@4K 3.6ms	
3D features	8.5	3337776	1676208	38 FPS@4K 26ms	170 FPS@4K 5.9ms	303 FPS@4K 3.3ms	
basemap + features	11.8	4386352	3769266	24 FPS@4K 41ms	102 FPS@4K 9.8ms	208 FPS@4K 4.8ms	

Table 1. Geometry construction and rendering performance (main application / rasterization).

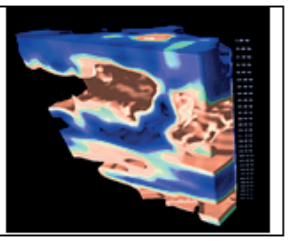
domain grid	constr. (s)	GTX 1050Ti	RTX 3070Ti	RTX 5080	
256^3	4	637 FPS@4K 1.57ms	1431 FPS@4K 0.69ms	1913 FPS@4K 0.52ms	
512^3	22	212 FPS@4K 4.72ms	573 FPS@4K 1.75ms	963 FPS@4K 1.04ms	
1024^3	72	26 FPS@4K 38.5ms	74 FPS@4K 13.51ms	141 FPS@4K 7.09ms	

Table 2. Volume geometry construction and rendering performance (DVR prototype).

The end-user experience in terms of data transfer highly depends on the user's available bandwidth and can impact streaming and conversion speeds. Depending on data hosts, significant data transfer differences can be observed from one server to another, for various reasons inherent to the compression methods and available server hosting formats.

Visual Fidelity and User Feedback

After various demonstrations and showcases, the precision and interactivity of the representations were judged to be excellent by a specialized scientific audience from different disciplines, with features demonstrated in the main application and both dedicated prototypes meeting their requirements.

Data Formats

The presented solution demonstrated results consistent with initial expectations regarding streaming either compressed image tiles for displacement or other forms of reconstructions, as

well as retrieving raw point cloud data via JSON network streams or from uncompressed local or remote text files (*e.g.*, *.XYZ, *.CSV, *etc.*).

Decoding files containing compressed data (such as NRRD or Network Common Data Form (NetCDF)) is also possible, and format compatibility is regularly extended in accordance with their respective specifications, to achieve robust data interchange capabilities.

Regarding data formats for rendering, various strategies, ranging from older traditional to modern GPU-friendly approaches (*e.g.*, Khronos Group's KTX2, which supports raw compressed data), minimizing CPU-side float texture decompression, were tested. These yielded satisfactory results even on previous generations of hardware.

Data Formats

Being web-based solutions, the application and prototypes proved to run successfully across various platforms, including consumer and workstation laptops and desktops, smartphones, tablets, and VR headsets (even smartwatches), without any modification of the code base.

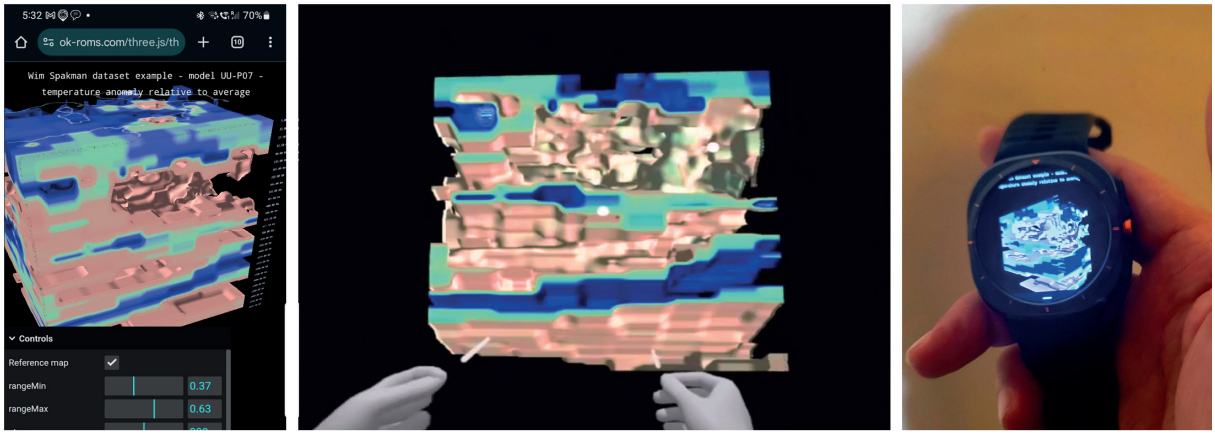


Figure 8. Browser output exceeding 60 frames per second on smartphones, VR headsets, and smartwatches.

Discussion

Enhancing Oceanographic Education

In the context of marine sciences, the need for 3D-4D fast data visualization is increasing for models and datasets of significant sizes. Relieving users of cumbersome tasks like software or extension installation by providing instantaneous representations in any web browser, rather than requiring dedicated local installations, can expedite the initiation of specific data analysis environments and provide an efficient, cost-effective, and portable educational platform.

Strengths and Limitations

Accessing the GPU through browsers remains more limited and constrained than desktop solutions for very large-scale data. However, the convenience afforded by the modularity of web development libraries and tools, combined with the ability to deploy a single source code base, drastically reduces and alleviates maintenance and evolution efforts.

The use of GPU-friendly data structures ensures satisfactory rendering performance.

"Although still outperformed by lower-level languages, web-based solutions continue to

improve their direct access to hardware resources (*e.g.*, via Web Workers[20] or more manual approaches). However, they remain less mature than locally installed software, which more effectively matches rendering specifications when it comes, for example, to dispatching threads across available processing units, whether managed (through frameworks like CUDA or OpenCL) or via custom Direct Memory Access (DMA engine[21]) transactions.

Comparison with Existing Tools

Some very efficient web 3D data visualization solutions have reached a mature stage in terms of rendering performance and data interchange capabilities but are still mostly provided as more abstracted APIs. These tools, such as Paraview Glance or those manually integrating the VTK library, introduce additional abstraction layers to contend with when building more custom visualizations.

These solutions often lack a fully flexible scene graph and an animation system for cameras, transforms, or keyframes. Furthermore, scene composition with external libraries is often more complex.

The vtk.js shader pipeline is abstracted away (derived from VTK), consequently making it challenging to directly override GLSL shaders and limiting control over buffers, instancing, and GPU pipelines.

The control is also limited for advanced rendering techniques (*e.g.,* procedural noise, custom raymarching).

These VTK-tied solutions can also have a large bundle size, exceeding 2MB, compared to 300-700KB with Three.js, even when including proper modular imports, assets, shaders, GUI, and libraries.

Offering less abstraction, Three.js provides a rapid prototyping capability for standard or non-standard applications across various disciplines and is maintained by a very large community offering a broad extension ecosystem.

Other solutions designed for medical imaging (XTK) or geospatial viewers (Cesium.js) are highly domain-specific, therefore not always suitable for facilitating prototyping with sufficient design freedom.

Cesium.js, for example, is excellent for globe-based data, but requires specialized formats (3D Tiles, CZML).

High-quality engines such as Babylon.js are more feature-rich than Three.js but can appear less tailored to scientific contexts.

Jupyter and PyThreejs are Great for Python-heavy workflows but have limited performance and interactivity.

20 Web Workers enable background JavaScript execution for parallel tasks.

21 A Direct Memory Access (DMA) engine is a hardware controller that transfers data between memory and peripherals without involving the CPU, reducing overhead and improving data throughput.

Conclusion and Future Work

Key objectives

The first milestone was to obtain a fast yet concise bathymetric representation of the Mediterranean and Black seas from GEBCO and EMODNET datasets, yielding solid 3D basemap examples.

After ensuring the data could be reliably streamed locally or remotely with properly sanitized queries, the user interface was enhanced with various forms and functions, enabling users to stream and superimpose numerous other environmental layers for comparison through different scientific data hosting protocols such as NOAA's ERDDAP or Thematic Real-time Environmental Distributed Data Services (THREDDS).

Further studies were then conducted to determine the most suitable types of accurate geometric representations, appropriate for datasets from various scientific disciplines, including geology/ seismology, physical and chemical parameters, and metagenomic catalogues.

Planned Extensions

A robust file interchange support is planned with client and server-side tools for formats translation tasks (*e.g.*, with the Geospatial Data Abstraction Library- GDAL) integration is in discussion, with potentially off-loading where possible, conversion tasks to the client-side with the help of WebAssembly[22] (WASM) modules.

Among other general code management tasks and future plugin integrations, a clear interest has emerged in establishing a more machine learning-friendly architecture, thereby providing improved, AI-assisted insights for both teaching and application use. These additions could pave the way for learning models trained in areas such as 'before query' metadata analysis, optimizing host server communication patterns, suggesting anomalies or key regions in datasets, or automatically applying optimal visualization modes (*e.g.,* using WebGPU or WASM-based ML models).

Further advancements also include the integration of Web Real-Time Communication (WebRTC[23]) and WebSockets[24] (Fette *et al.* 2011) to support real-time shared dataset navigation, enabling users to annotate and comment directly within the data structure.

Closing Notes

Identifying and implementing efficient methods for faster display of real-time rendered multidimensional scientific datasets enhances the illustration of research and educational projects.

Thoroughly optimizing development and deployment with web technologies and protocols consistently saves significant time and resources.

More traditional and low-level programming tools maintain the performance lead for embedded, computationally intensive tasks, while their deployment and compatibility advantages gradually diminish over time, thanks to network-based approaches.

22 WebAssembly (WASM) is a fast, binary-format web runtime for portable code execution.
23 WebRTC enables real-time peer-to-peer audio, video, and data communication.
24 WebSockets enable full-duplex communication between a client and server

Acknowledgements

The author extends his sincere thanks to the Scientific Committee Chairs for their extensive advice and consultations, which made the development of this tool possible: the Chairs of the Marine Geosciences Committee: Professors Jean Mascle (GeoAzur, France) and Luis Menezes Pinheiro (Aveiro University, Portugal), and the Chair of the Physics and Climate of the Ocean Committee: Dr Katrin Schroeder (CNR-ISMAR, Italy).
The author would also like to thank:

- Professor Wim Spakman (Utrecht University, Netherlands) for sharing his seismic temperature anomaly models and his advice on efficiently rendering them.
- Dr David Sallis (NOAA National Centers for Environmental Information) for his valuable availability and support with NOAA ERDDAP servers for testing physical layers.
- Professor Giuseppe D'Auria and Dr Mariana Reyes (FISABIO's Bioinformatics, Spain) for their metagenomic samples and detailed insights.

References

Amaru, M.L., 2007. *Global travel time tomography with 3-D reference models*. Geologica Ultraiectina, 274, 174p.

Crassin, C., 2011. *GigaVoxels: A Voxel-Based Rendering Pipeline For Efficient Exploration Of Large And Detailed Scenes*. Ph.D. Thesis, Université de Grenoble. Available at: https://maverick.inria.fr/Publications/2011/Cra11/CCrassinThesis_EN_Web.pdf.

EMODnet Bathymetry Consortium, 2020. *EMODnet Digital Bathymetry (DTM 2020)*. Available at: https://doi.org/10.12770/bb6a87dd-e579-4036-abe1-e649cea9881a.

EMODnet Physics, 2020. *Collection of Silicate (SIO4-SI) (SLCW) profiles (Multi Point Profile Observation)*. Available at: https://ercompwebapps.emodnet-physics.eu/erddap/tabledap/TS_SLCA.html.

Fette, I., Melnikov, A., 2011. *The WebSocket Protocol*. IETF RFC 6455. Available at: https://www.rfc-editor.org/rfc/rfc6455.html.

Hall, R., Spakman, W., 2015. Mantle structure and tectonic history of SE Asia. *Tectonophysics*, 658, 14–45.

Ju, T., Losasso, F., Schaefer, S., Warren, J., 2002. Dual Contouring of Hermite Data. *ACM Transactions on Graphics*, 21(3), 339–346.

Kanehisa, M., Furumichi, M., Sato, Y., *et al.* (2025). KEGG: biological systems database as a model of the real world. *Nucleic Acids Research*, 53(D1), D672–D677.

Knoll, A., Wald, I., Parker, S., & Hansen, C., 2006. *Interactive Isosurface Ray Tracing of Large Octree Volumes*. Scientific Computing and Imaging Institute, University of Utah, Technical Report No UUSCI-2006-026. Available at: https://www.sci.utah.edu/publications/SCITechReports/UUSCI-2006-026.pdf.

Lefebvre, S., & Hoppe, H., 2006. Perfect Spatial Hashing. *ACM Transactions on Graphics*, 25(3), 579–588.

Laine, S., Karras, T., 2010. Efficient Sparse Voxel Octrees. *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, 55–63.

Lorensen, W.E., Cline, H.E., 1987. Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 21(4), 163–169.

Schaefer, S., Warren, J., 2004. Dual Marching Cubes: Primal Contouring of Dual Grids. *Proceedings of the 12th Pacific Conference on Computer Graphics and Applications*, 70–76.

External Resources

Cabello, R., 2010. three.js. Available at: https://threejs.org/.

GDAL Development Team. n.d. *GDAL - Geospatial Data Abstraction Library*. doi:10.5281/zenodo.5884351. Available at: https://gdal.org/en/stable/.

GEBCO Compilation Group, 2020. *GEBCO_2020 Grid* [Data set]. GEBCO. Retrieved from https://www.gebco.net

Khronos Group, 2021. *Khronos KTX 2.0 Textures Enable Compact, Visually Rich glTF 3D Assets*. Available at: https://www.khronos.org/news/press/khronos-ktx-2-0-textures-enable-compact-visually-rich-gltf-3d-assets.

Khronos Group. n.d. *OpenCL - The Open Standard for Parallel Programming*. Available at: https://www.khronos.org/opencl/.

Mozilla Developer Network (MDN) Web Docs. n.d. *Using Web Workers*. Available at: https://developer.mozilla.com/en-US/docs/Web/API/Web_Workers_API/Using_web_workers.

NOAA National Centers for Environmental Information (NCEI), 2013. *Global Navy Coastal Ocean Model (NCOM Region 2 3D aggregation)*. Available at: https://www.ncei.noaa.gov/products/weather-climate-models/global-navy-coastal-ocean.

NVIDIA Corporation, 2007. *CUDA: Compute Unified Device Architecture*. Available at: https://developer.nvidia.com/cuda-toolkit.

W3C, 2024. *WebGPU*. W3C Working Draft, 2 April 2024. Available at: https://www.w3.org/TR/webgpu/.

W3C, 2024. *WebRTC: Real-Time Communication in Browsers*. W3C Recommendation, 15 October 2024. Available at: https://www.w3.org/TR/webrtc/.

WebAssembly Community Group. n.d. *WebAssembly*. Available at: https://webassembly.org/.

LIST OF PARTICIPANTS

Giuseppe d'Auria
Fisabio, Spain
Giuseppe.dauria@fisabio.es

Monia El Bour
(Chair, CIESM Committee 4)
INSTM, Tunisia
moniaelbour@gmail.com

Paola Furla
Univ. Nice, France
paola.furla@unice.fr

Donato Giovannelli
CNR, Italy
donato.giovannelli@gmail.com

Laura Giuliano
CIESM, Director General
lgiuliano@ciesm.org

Mohamed Jebbar
Univ. Brest, France
mohamed.jebbar@univ-brest.fr

Paula Moschella
CIESM Headquarters
pmoschella@ciesm.org

Z. Bora Ön
Univ. Mugla, Turkey
zekiboraon@gmail.com

Luis Pinheiro
(Chair, CIESM Committee 1)
Univ. Aveiro, Portugal
lmp@ua.pt

Kaveh Rassoulzadegan
CIESM Headquarters
kvr@ciesm.org

Mariana Reyes Prieto
Fisabio, Spain
mariana.reyes@fisabio.es

Lorna Richardson
Univ. Cambridge, UK
lornar@ebi.ac.uk

Thierry Schmitt *(remote)*
Emodnet, SHOM, France
thierry.schmitt@shom.fr

Javier Tamames
CNB CSIC, Spain
jtamames@cnb.csic.es

Virginie Tassin Campanella
VTA Tassin, Switzerland
virginie.tassin@vta-tassin.com

Alberto Vitale Brovarone
Univ. Bologna, Italy
alberto.vitaleb@unibo.it

Mikhail Yakimov
CNR, Italy
mikhail.iakimov@cnr.it

Mustafa Yucel
MET, Turkey
muyucel@metu.edu.tr

Noureddine Zaaboub
INSTM, Tunisia
muyucel@metu.edu.tr

Printed by

IMPRIM'VERT®

MULTIPRINT
MONACO

FOURNISSEUR BREVETÉ DE S.A.S. LE PRINCE DE MONACO

Mediterranean Science Commission
CIESM

An intergovernmental organization
- 23 Member states -
promoting since 1910 research cooperations in the Mediterranean and Black Seas in the many sectors of oceanographic and coastal sciences as well as marine policy. Supported by a large pluridisciplinary network: more than 500 institutes and 10 000 researchers.

Une organisation intergouvernementale - 23 Etats membres - intervenant dans les domaines de la recherche océanographique pour favoriser les coopérations entre scientifiques ainsi qu'entre chercheurs et décideurs. Avec le concours d'un vaste réseau pluridisciplinaire : plus de 500 instituts spécialisés et de 10 000 chercheurs.

www.ciesm.org

** CIESM Workshop volumes have been published since 1997 as such under:*
- CIESM Workshop Series, n°1-18
- CIESM Workshop Monographs, n° 19-onward

The full list is available on our website as well as an order form.

*Already published **

*Déjà parus **

34
Towards an integrated system of Mediterranean marine observatories
La Spezia (Italy), 16-19 January 2008

35
Climate warming and related changes in Mediterranean marine biota
Helgoland (Germany), 27-31 May 2008

36
Impacts of acidification on biological, chemical and physical systems in the Mediterranean and Black Seas
Menton (France), 1-4 October 2008

37
Economic valuation of natural coastal and marine ecosystems
Bodrum (Turkey), 22-25 October 2008

38
Dynamics of Mediterranean deep waters
Malta, 27-30 May 2009

39
Climate forcing and its impacts on the Black Sea marine biota
Trabzon (Turkey), 3-6 June 2009

40
Phytoplankton responses to Mediterranean environmental changes
Tunis (Tunisia), 7-10 October 2009

41
Marine Peace Parks in the Mediterranean - a CIESM proposal
Syracuse (Italy), 18-20 November 2010

42
Marine geo-hazards in the Mediterranean
Nicosia (Cyprus), 2-5 February 2011

43
Designing Med-SHIP: a Program for repeated oceanographic surveys
Supetar, Brac Island (Croatia), 11-14 May 2011

44
Molecular adaptations of marine microbial life to extreme environments - Mediterranean opportunities
Island of Samos (Greece), 12-15 October 2011

45
Marine extinctions - patterns and processes
Valencia (Spain), 10-13 October 2012

46
Marine litter in the Mediterranean and Black Seas
Tirana (Albania), 18-21 June 2014

47
Submarine canyon dynamics in the Mediterranean and tributary seas - An integrated geological, oceanographic and biological perspective
Sorrento (Italy), 15-18 April 2015

48
Marine connectivity - migration and larval dispersal
Sòller (Spain), 9-12 March 2016

49
Searching for bacterial pathogens in the Digital Ocean
Paris (France), 27-30 September 2017

50
Engaging marine scientists and fishers to share knowledge and perceptions - Early lessons
Paris (France), 18-21 April 2018

51
Marine heatwaves in the Mediterranean Sea and beyond -
Paris (France), 27-30 November 2023

52
Marine hazards, coastal vulnerability, risk (mis) perceptions – a Mediterranean perspective