

ИЗТЕГЛЯНЕ, ОБРАБОТКА И ВИЗУАЛИЗИРАНЕ НА РЕАЛНИ МОРСКИ НАБОРИ ОТ ДАННИ: ПРЕВРЪЩАНЕ НА ОКЕАНОГРАФСКИ ДАННИ В СМИСЛЕНИ НАУЧНИ ИСТОРИИ

доц. д-р Димитър Минчев, Бургаски свободен университет, mitko@bfu.bg

Абстракт: Бързият растеж на хранилищата за морски данни с отворен достъп трансформира океанографските изследвания, образованието и мониторинга на околната среда. Трансформирането на големи обеми от сурови сателитни и океански модели данни в интерпретируеми научни знания обаче остава значително предизвикателство. Тази статия представя практичен работен процес за изтегляне, обработка и визуализиране на реални морски набори от данни, получени от големи международни инфраструктури за морски данни, включително Морската служба „Коперник“ на Европейската космическа агенция (ESA), NASA OceanColor и NOAA ERDDAP. Работният процес демонстрира как програмирането, подпомогнато от изкуствен интелект, в комбинация с научни библиотеки на Python, може да ускори прехода от изследователски въпроси на естествен език към възпроизводими аналитични работни процеси. Представени са реални примери, базирани на наблюдения на температурата на морската повърхност (SST) и хлорофил-а, включително казуси от Черно море, Бургаския залив, региона на Балтийско море близо до Рига и образователни примери, свързани с европейски университети. Изследването илюстрира как внимателно проектираните визуализации трансформират сложни морски набори от данни в смислени научни истории, които подпомагат разбирането за околната среда и вземането на решения.

Ключови думи: Морски данни, Океанография, Визуализация на данни, Морска служба на Коперник, NASA OceanColor, NOAA ERDDAP, Python, Температура на морската повърхност, Хлорофил-а, Научно разказване на истории.

DOWNLOADING, PROCESSING, AND VISUALIZING REAL MARINE DATASETS: TURNING OCEANOGRAPHIC DATA INTO MEANINGFUL SCIENTIFIC STORIES

Assoc. Prof. PhD Dimitar Minchev, Burgas Free University, mitko@bfu.bg

Abstract: The rapid growth of open-access marine data repositories has transformed oceanographic research, education, and environmental monitoring. However, transforming large volumes of raw satellite and ocean model data into interpretable scientific knowledge remains a significant challenge. This paper presents a practical workflow for downloading, processing, and visualizing real marine datasets obtained from major international marine data infrastructures, including the European Space Agency (ESA) Copernicus Marine Service, NASA OceanColor, and NOAA ERDDAP. The workflow demonstrates how artificial intelligence-assisted programming, combined with Python scientific libraries, can accelerate the transition from natural-language research questions to reproducible analytical workflows. Real examples based on Sea Surface Temperature (SST) and chlorophyll-a observations are presented, including case studies from the Black Sea, Burgas Bay, the Baltic Sea region near

Riga, and educational examples associated with European universities. The study illustrates how carefully designed visualizations transform complex marine datasets into meaningful scientific stories that support environmental understanding and decision-making.

Keywords: Marine Data, Oceanography, Data Visualization, Copernicus Marine Service, NASA OceanColor, NOAA ERDDAP, Python, Sea Surface Temperature, Chlorophyll-a, Scientific Storytelling.

1. Introduction

Modern oceanography increasingly relies on satellite observations, numerical models, autonomous sensors, and international data infrastructures. Open-access repositories now provide continuous access to physical, biogeochemical, and ecological variables describing ocean conditions at global and regional scales. Examples include the ESA Copernicus Marine Data Store [1,2], NASA OceanColor archives [2,3], and NOAA’s Environmental Research Division Data Access Program (ERDDAP) [5]. These systems provide datasets describing sea surface temperature, salinity, ocean currents, chlorophyll concentration, sea level, dissolved oxygen, and many other parameters.

The challenge is no longer merely obtaining oceanographic data. Researchers must transform raw observations into meaningful visual narratives that reveal environmental processes, temporal trends, and spatial variability. The concept of scientific storytelling combines data acquisition, analysis, visualization, and interpretation into a coherent workflow that enables scientific insights and effective communication.

Recent developments in artificial intelligence (AI) provide new opportunities for accelerating this workflow. Instead of manually writing extensive code, researchers can formulate natural-language questions and use AI-generated Python [8] scripts as starting points for reproducible marine data analysis. However, scientific validation remains essential, requiring verification of dataset identifiers, spatial coordinates, temporal coverage, variables, and measurement units.

2. Marine Data Sources and Infrastructure

Three major marine-data ecosystems form the foundation of the presented workflow.

2.1 ESA Copernicus Marine Service

The Copernicus Marine Service [1,2] provides free and science-based information about the physical ocean (Blue Ocean), sea ice (White Ocean), and marine biogeochemistry (Green Ocean) at global and regional scales. It represents one of the most comprehensive operational marine information systems currently available. Access requires a free account and supports direct Python [8] integration through the `copernicusmarine` package.

Fig. 1. National Aeronautics and Space Administration (NASA) Earth Observation Data

Available variables include: Water temperature (thetao), Salinity (so), Eastward current velocity (uo), Northward current velocity (vo), Sea surface height (zos), Chlorophyll-a (chl) and Dissolved oxygen (o2)

2.2 NASA OceanColor

NASA OceanColor [3,4] products provide satellite observations of marine biological and optical properties. Particularly valuable are chlorophyll-a products derived from MODIS Aqua and other Earth-observing sensors. These datasets support investigations of phytoplankton dynamics, primary productivity, and ecosystem health.

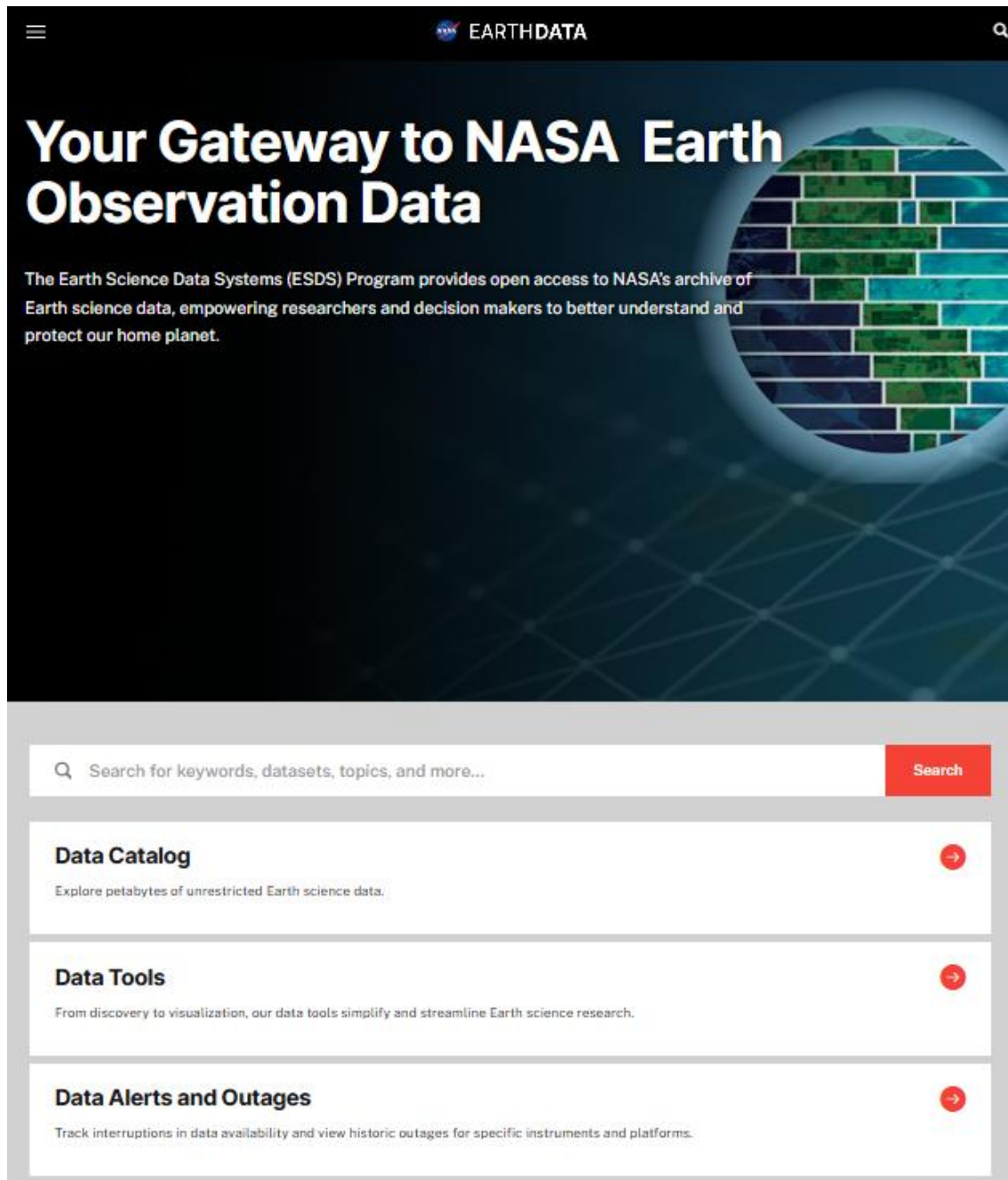


Fig. 2. National Aeronautics and Space Administration (NASA) Earth Observation Data

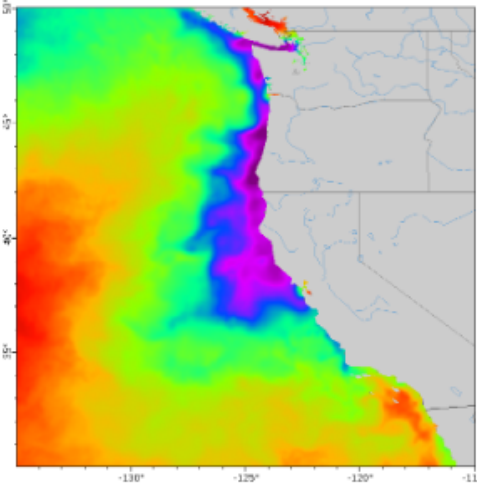
2.3 NOAA ERDDAP

NOAA ERDDAP [5] offers standardized web-based access to numerous marine datasets through reproducible URL queries. Unlike traditional download systems, ERDDAP allows direct retrieval of small spatial and temporal subsets, making it particularly suitable for educational applications and lightweight scientific workflows.

[HOME](#) [DATA ACCESS ▾](#) [TOOLS & TRAINING ▾](#) [ABOUT ▾](#) [PROJECTS](#)

**NOAA** COASTWATCH
WEST COAST REGIONAL NODE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Providing satellite data for the coastal ocean



Environmental Data

View and download over 800 regional and global datasets, including satellite data, model output, and in situ measurements from field sensors. The data are intended for experimental use only and are not supported on an operational basis ([disclaimer](#))

[Data Catalog](#) [Coastal Conditions](#)

ERDDAP Data Server

The ERDDAP data server provides a simple, consistent way to subset and download environmental datasets in common file formats with options to make graphs and maps.

[Features](#) [Get Data](#)

Software

The Environmental Data Connector (EDC) and Xtractomatic data extraction scripts make it easy to discover and extract data from online servers and download them directly into ArcGIS, R, MatLab, and Excel.

[EDC](#) [Xtractomatic](#)

Fig. 3. National Oceanic and Atmospheric Administration (NOAA) Environmental Research Division's Data Access Program (ERDDAP)

3. Methodology

The proposed workflow consists of five stages:

1. Define the scientific question
2. Select an appropriate dataset
3. Download data through Python
4. Visualize spatial and temporal patterns
5. Validate and interpret results

The required software environment includes Visual Studio Code, Python [8], and scientific libraries such as copernicusmarine [1,2], earthaccess [3,4], xarray [6], matplotlib [7], numpy [9], pandas, and netCDF4. These tools support downloading, processing, and visualization of NetCDF-based marine datasets.

AI can assist by generating initial Python [8] workflows from prompts specifying: Data source, Geographic region, Variable, Time period and Visualization requirements.

However, resulting scripts must always be validated against official metadata before scientific use.

4. Case Studies and Real-Time Examples

4.1 Black Sea Sea Surface Temperature Mapping

One example demonstrates downloading a June 2026 Sea Surface Temperature dataset for the Black Sea using Copernicus Marine. The workflow downloads a regional subset, reads the NetCDF product, and generates a thematic map illustrating spatial SST variability.

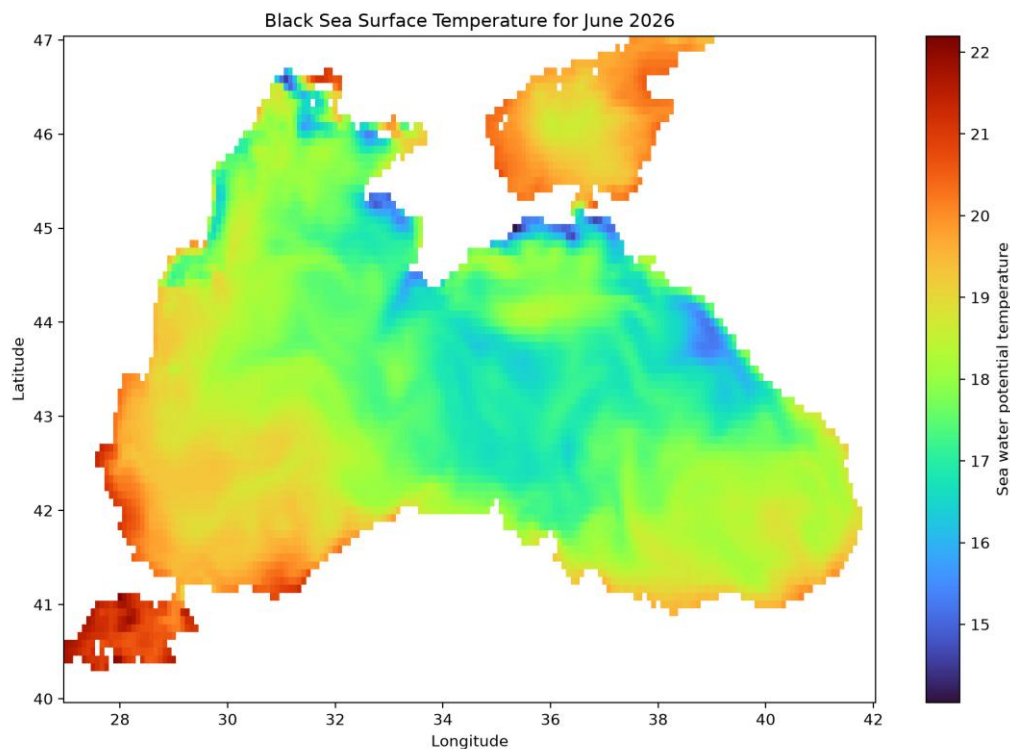


Fig. 4 Black Sea Surface Temperature (SST)

The resulting visualization reveals warmer southern sectors and relatively cooler northern areas, illustrating how spatial mapping immediately transforms numerical data into interpretable environmental information.

4.2 Sea Surface Temperature Near Burgas

A second example focuses on local ocean conditions near Burgas during June 2026. Instead of visualizing an entire basin, SST values are extracted from a regional grid cell and plotted as a time series. This approach highlights daily variability and supports studies of seasonal warming trends, marine heat events, and local coastal processes.

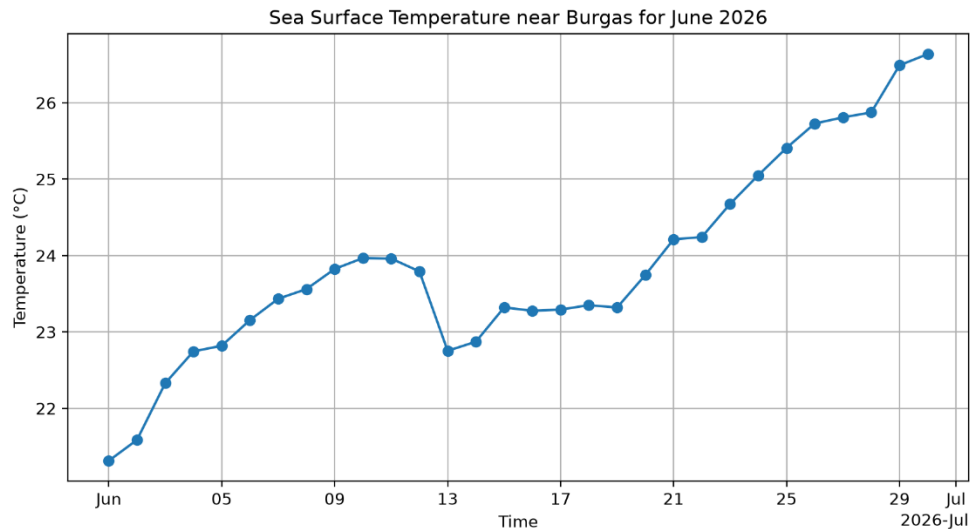


Fig. 5 Sea Surface Temperature (SST) near Burgas

The time-series representation demonstrates how temporal aggregation can reveal environmental dynamics that are difficult to identify in static maps.

4.3 Chlorophyll-a Around Burgas

Using NASA OceanColor MODIS Aqua Level-2 products, chlorophyll-a concentrations around Burgas were visualized for June 2026. Chlorophyll-a serves as an indicator of phytoplankton biomass and marine productivity.

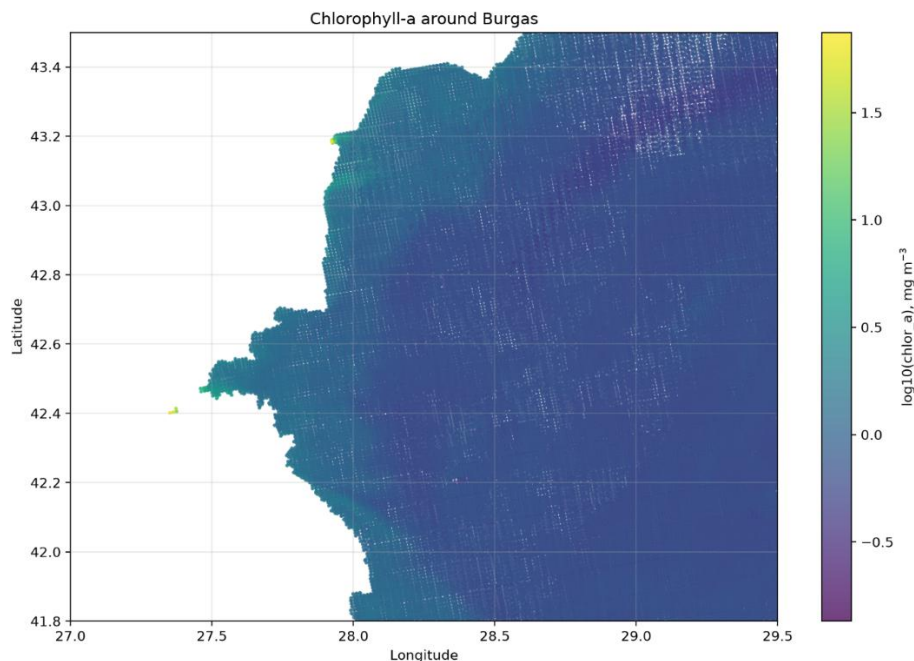


Fig. 6 Chlorophyll-a Around Burgas

This example demonstrates how biogeochemical variables complement physical ocean observations, enabling researchers to examine ecosystem responses alongside environmental forcing.

4.4. Daily Median Chlorophyll-a Trends

Spatial chlorophyll measurements can be aggregated into daily median values, producing simple trend visualizations that support scientific interpretation. Median statistics reduce the influence of outliers caused by clouds, sensor noise, or quality-control issues.

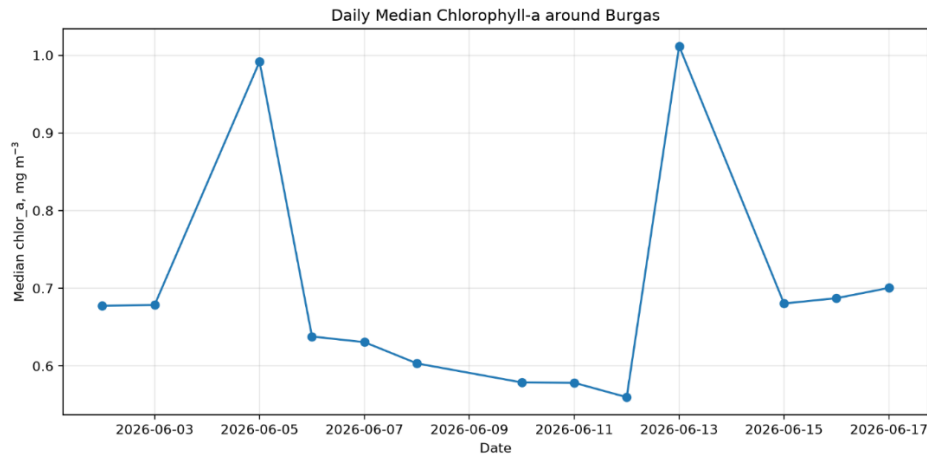


Fig. 7. Daily Median Chlorophyll-a around Burgas

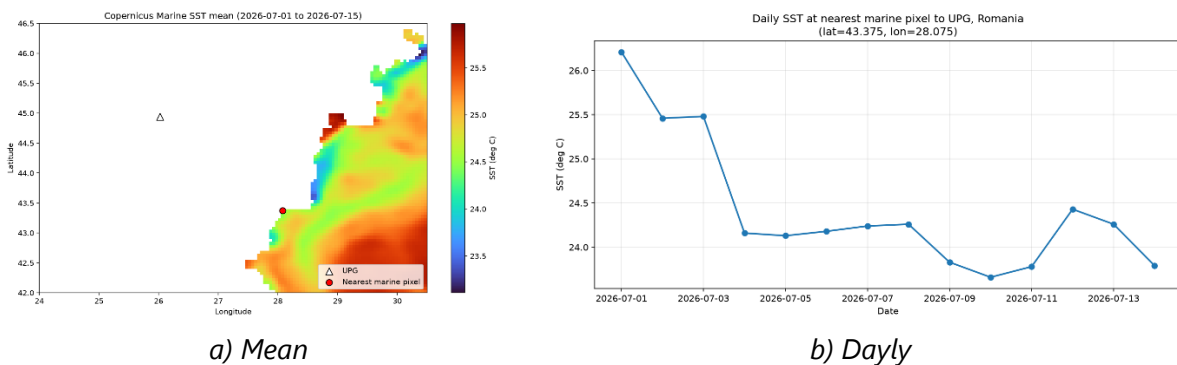
The resulting graph transforms thousands of satellite measurements into a concise environmental narrative describing biological variability through time.

4.5 Multi-Regional Educational Examples

Additional educational examples demonstrate SST extraction for study regions associated with:

- University of Petroleum-Gas, Romania
- Bogdan Voda University, Romania
- University of Trnava, Slovakia
- College of Law, Riga, Latvia

These case studies apply identical workflows while emphasizing the importance of careful geographic validation and bounding-box selection. Notably, inland locations require explicit specification of representative marine study areas rather than relying solely on institution names. The Riga example demonstrates application within the Baltic Sea and Gulf of Riga environment, where marine conditions can be meaningfully analyzed.



a) Mean *b) Daily*
Fig. 8. Satellite Sea Surface Temperature (SST) near University of Petroleum-Gas, Romania (1-15 July 2026)

5. From Raw Data to Scientific Stories

Scientific storytelling emerges when technical workflows are combined with meaningful interpretation. The presented examples illustrate three complementary visualization approaches:

1. **Spatial maps** reveal geographic patterns.
2. **Time-series** plots reveal temporal dynamics.
3. **Multi-variable comparisons** reveal environmental relationships.

For example, simultaneous analysis of SST and chlorophyll-a may help investigate links between thermal conditions and biological productivity. Similarly, integration of currents, salinity, oxygen, and sea-level observations enables broader ecosystem analyses.

Effective scientific stories require: Reproducible workflows, Transparent metadata, Proper validation, Clear graphics and Domain expertise.

AI can accelerate implementation but cannot replace scientific judgment. Researchers remain responsible for ensuring correct dataset selection, geographic applicability, and interpretation of environmental processes.

6. Conclusion

The growing availability of marine datasets from Copernicus Marine [1,2], NASA OceanColor [3,4], and NOAA ERDDAP [5] provides unprecedented opportunities for oceanographic research and education. By combining Python-based workflows with AI-assisted code generation, researchers can rapidly access, process, and visualize real marine observations.

The examples presented demonstrate that meaningful scientific stories emerge when spatial maps, temporal analyses, and domain knowledge are combined into reproducible workflows. Whether analyzing Black Sea temperatures, chlorophyll variability around Burgas, or Baltic Sea conditions near Riga, the same methodological framework applies: define a question, obtain data, visualize results, validate assumptions, and interpret findings scientifically.

As marine observation systems continue to expand, the ability to transform raw datasets into clear, reproducible, and compelling scientific narratives will become an increasingly important skill for both researchers and students.

References

- [1] Copernicus Marine Service. (2025). Copernicus Marine Data Store. Available at: <https://data.marine.copernicus.eu>
- [2] European Commission. (2025). Copernicus Programme Marine Service Documentation. Brussels, Belgium.
- [3] NASA Ocean Biology Processing Group. (2025). NASA OceanColor Web. Available at: <https://oceancolor.gsfc.nasa.gov>
- [4] NASA Earth Science Data Systems (ESDS). (2025). Earthdata Documentation. Available at: <https://earthdata.nasa.gov>
- [5] NOAA Environmental Research Division. (2025). ERDDAP Documentation and User Guide. Available at: <https://coastwatch.pfeg.noaa.gov/erddap>
- [6] Hoyer, S., & Hamman, J. (2017). xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software*, 5(1), 10.
- [7] Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90-95.
- [8] McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, 51-56.
- [9] Harris, C. R., et al. (2020). Array Programming with NumPy. *Nature*, 585, 357-362.

- [10] Roemmich, D., et al. (2019). On the Future of Argo: A Global, Full-Depth, Multi-Disciplinary Array. *Frontiers in Marine Science*, 6, 439.
- [11] OceanPredict. (2024). Global Ocean Forecasting and Marine Services. Available at: <https://oceanpredict.org>
- [12] GHRSSST Science Team. (2024). Group for High Resolution Sea Surface Temperature (GHRSSST). Available at: <https://www.ghrsst.org>
- [13] EMODnet. (2025). European Marine Observation and Data Network. Available at: <https://emodnet.ec.europa.eu>
- [14] International Argo Program. (2025). Argo Float Data and Products. Available at: <https://argo.ucsd.edu/data/>