Proceedings of the 4th International Conference on Computing Innovation and Applied Physics
DOI: 10.54254/2753-8818/84/2025.21024

Integrating Big Data Analytics and Visualization in Ocean Circulation Modeling for Climate and Ecosystem Insights

Jiongyu Liu[1,a,*]

[1]Qingdao No.2 Middle School, Shandong, China
a. jiongyuliu9@gmail.com
**corresponding author*

Abstract: Ocean circulation plays a key role in the global climate system and marine ecosystems. Accurate monitoring and modelling of ocean circulation not only contributes to an in-depth understanding of climate dynamics, but also has important applications in marine biodiversity conservation, shipping route optimization, and early warning of natural disasters (e.g., storm surges and tsunamis). With the rise of big data technologies, the volume of ocean circulation data has increased dramatically, and traditional analytical methods are difficult to cope with these complex datasets. The efficiency and accuracy of ocean circulation studies can be significantly improved by introducing big data analysis and visualisation techniques. This paper provides an overview of the application of big data analytics and visualisation methods in ocean circulation research, discusses various analysis techniques used to handle big data sets, evaluates their effectiveness, and looks at future trends and challenges in the field, future research will aim to further improve the accuracy and efficiency of ocean circulation data processing.

Keywords: Ocean Circulation, Big Data Analytics, Climate Modeling.

1. Introduction

Ocean circulation plays a crucial role in the global climate system and marine ecosystems. Accurate ocean circulation monitoring and modeling is not only critical for understanding global climate dynamics, but also has an important role in marine biodiversity conservation, shipping route optimization, and early warning of natural hazards (e.g., storm surges and tsunamis) They regulate heat distribution, influence weather patterns, and affect nutrient cycling, which is vital for marine life. Recent studies indicate that ocean currents are becoming increasingly unpredictable due to climate change, necessitating more precise monitoring and modeling.

The advent of big data technologies has led to the generation of vast and complex datasets related to ocean currents. Traditional analytical methods, which often rely on smaller datasets and simpler models, are now inadequate to tackle the nuances of modern oceanographic data. For instance, satellite altimetry can produce global sea surface height data with resolutions of 1–10 cm, leading to enormous datasets that require advanced analytical techniques[1]. Therefore, the integration of big data analytics and visualization techniques is essential to enhance the efficiency and accuracy of ocean current research. This review aims to summarize the current applications of big data analytics and visualization methods in the study of ocean currents. It will explore various analytical techniques

DOI: 10.54254/2753-8818/84/2025.21024

used to handle large datasets, assess their effectiveness, and discuss future trends and challenges in the field.

2. Applications and Challenges of Multi-Source Data in Ocean Circulation Monitoring

Ocean current data is sourced from several key platforms, including satellite remote sensing, buoy observations, numerical model outputs, and ocean station data.

Satellite remote sensing: Instruments such as altimeters and scatterometers provide global coverage with high temporal resolution. For instance, the Jason satellite series has been instrumental in measuring sea surface height with a precision of around 3 cm, contributing significantly to our understanding of ocean dynamics.

Buoy observations: Drifting and stationary buoys collect in-situ measurements of ocean temperature, salinity, and currents. For example, the Global Drifter Program has deployed thousands of buoys that provide real-time data, enhancing our ability to model ocean circulation patterns.

Numerical Model Outputs: Ocean circulation models, such as those run by the National Oceanic and Atmospheric Administration (NOAA), generate vast amounts of data predicting ocean behavior under various climate scenarios. These models help simulate conditions over different time scales, aiding in understanding long-term changes.

Ocean Station Data: Fixed oceanographic stations collect data over extended periods, allowing for long-term monitoring of ocean currents and associated phenomena.

Each data source offers varying spatial and temporal resolutions, presenting unique challenges in data processing and integration. One of the main challenges to multi-source data fusion is the inconsistency of different data sources in terms of temporal and spatial resolution. In addition, the non-linear nature of ocean processes requires complex algorithms to deal with them, while the storage and computational power of large amounts of data also places higher demands on the systemFor instance, satellite data may have a global view but lacks the high resolution found in buoy observations, highlighting the need for multi-source data fusion.

Another problem when analyzing ocean circulation is that ocean current big data exhibits several complex characteristics, including: multisource Heterogeneity and nonlinearity and multiscale dynamics. There are also many challenges in processing ocean current data such as data processing, cleaning, and integration. One of the main challenges to multi-source data fusion is the inconsistency of different data sources in terms of temporal and spatial resolution. In addition, the non-linear nature of ocean processes requires complex algorithms to deal with them, while the storage and computational power of large amounts of data also places higher demands on the system.

3. Traditional Numerical Methods in Ocean Circulation Modeling and Their Limitations

3.1. Traditional Methods and Limitations

Ocean circulation is integral to understanding climate dynamics, marine ecosystems, and global weather patterns. Traditionally, oceanographers have utilized classical numerical simulation methods, primarily finite difference methods (FDM) and finite element methods (FEM), to analyze oceanic processes. While these methods have been effective in various contexts, they exhibit significant limitations, particularly when dealing with the vast datasets produced by modern observational technologies.

3.1.1. Finite Difference Method

The finite difference method (FDM) involves approximating derivatives in differential equations using differences at discrete points. This method is straightforward and computationally efficient for

DOI: 10.54254/2753-8818/84/2025.21024

simple problems. For instance, FDM is commonly employed in the numerical simulation of shallow water equations to model tidal movements. However, when applied to large-scale ocean datasets, FDM encounters several challenges.

Grid Resolution: To capture fine-scale phenomena, such as eddies or coastal upwelling, a finer grid is required. This significantly increases computational demand. A study by O'Brien et al. demonstrated that simulations with a 1 km grid resolution took nearly 10 times longer than those with a 5 km grid [2].

Stability Issues: FDM can suffer from numerical instability, especially in complex, turbulent flows. Instabilities can lead to unrealistic oscillations in the computed solutions, necessitating careful tuning of parameters and time steps.

Scalability: As datasets grow, the need for real-time analysis becomes critical. FDM typically requires substantial computational resources, making it difficult to apply in scenarios requiring quick decision-making, such as responding to environmental emergencies.

3.1.2. Finite Element Method

Finite element methods (FEM) provide a more flexible approach by subdividing the problem domain into smaller, simpler parts (elements). This is particularly beneficial for complex geometries, such as coastal regions. For example, FEM has been successfully applied to model the interaction of ocean currents with varying bathymetry.

Despite its advantages, FEM also faces limitations: Complexity and Setup Time: Setting up an FEM model requires significant effort, especially in defining the mesh and boundary conditions. This can be a barrier in dynamic environments where conditions change rapidly. Computational Intensity: Similar to FDM, FEM can become computationally expensive when high-resolution meshes are required. This restricts its usability in large-scale applications, particularly when combined with vast observational datasets.Integration with Observational Data: FEM often struggles to incorporate real-time data streams effectively, limiting its applicability in scenarios where continuous monitoring and adjustment are necessary.

3.1.3. Limitations in Handling Big Data

The combination of high-resolution models and large datasets from satellite observations, buoys, and numerical model outputs creates a perfect storm of challenges for traditional analysis methods. The volume, velocity, and variety of data produced require advanced processing capabilities that FDM and FEM cannot provide. Consequently, researchers are increasingly turning to big data analytics to unlock insights from ocean circulation data.

3.2. Applications of Big Data Analysis Technologies in Ocean Circulation

The integration of big data technologies has transformed ocean circulation analysis. By leveraging machine learning, deep learning, data mining, and distributed computing frameworks, researchers can effectively analyze and interpret large datasets.

3.2.1. Machine Learning Techniques

Machine learning (ML) techniques have become pivotal in oceanographic studies, particularly for pattern recognition, data-driven predictions, and anomaly detection.

DOI: 10.54254/2753-8818/84/2025.21024

3.2.2. Random Forest

Random Forest is an ensemble learning method that operates by constructing a multitude of decision trees. It has been effectively used in oceanographic research to predict various phenomena. For instance, Gopalakrishnan et al. utilized Random Forest to forecast sea surface temperature anomalies in the Indian Ocean [3]. By training the model on historical satellite data, they achieved an accuracy of over 85%, demonstrating the potential of ML techniques to enhance predictive capabilities in oceanography.

3.2.3. Support Vector Machines

Support Vector Machines (SVM) have also shown promise in classifying oceanographic features based on satellite imagery. A study by Zhang et al. employed SVM to classify chlorophyll-a concentrations in the South China Sea using MODIS satellite data, achieving an impressive classification accuracy of 90% [4]. This application is vital for monitoring algal blooms, which significantly impact marine ecosystems and fisheries.

3.2.4. Neural Networks

Neural networks, particularly multi-layer perceptrons, have been used to predict ocean current patterns. Chen et al. developed a neural network model to forecast surface currents in the Gulf Stream using historical wind and current data [5]. Their model demonstrated a substantial improvement in prediction accuracy compared to traditional methods, highlighting the effectiveness of neural networks in oceanographic applications.

3.2.5. Deep Learning Techniques

Deep learning (DL) algorithms, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, excel in analyzing complex spatiotemporal ocean data.

3.2.6. Convolutional Neural Networks

CNNs are particularly suited for analyzing grid-based data, such as satellite images. For example, Shi et al. employed CNNs to analyze sea surface temperature data and identify patterns associated with El Niño events [6]. Their model successfully detected these events weeks in advance, providing valuable information for climate prediction and preparation.

3.2.7. Long Short-Term Memory Networks

LSTMs are specialized for sequential data, making them ideal for capturing temporal dependencies in oceanographic data. Wang et al. used LSTM networks to predict ocean temperature changes over time [7]. Their model leveraged historical temperature data to forecast future conditions with remarkable accuracy, underscoring the potential of LSTMs in climate modeling and oceanography.

3.2.8. Data Mining Techniques

Data mining techniques are crucial for trend analysis, pattern discovery, and association rule mining in ocean circulation studies.

DOI: 10.54254/2753-8818/84/2025.21024

3.2.9. Trend Analysis

Data mining has proven effective in identifying long-term trends in ocean circulation patterns. Liu et al. analyzed ocean current data from the Gulf of Mexico and identified significant changes linked to climate change [8]. This trend analysis is essential for understanding the implications of changing ocean dynamics on marine ecosystems and fisheries.

3.2.10. Pattern Discovery

Clustering algorithms, such as k-means clustering, are utilized to uncover hidden patterns in ocean data. A study focused on the North Atlantic utilized clustering to identify distinct current regimes associated with different seasonal patterns. This research has critical implications for understanding how changing currents may impact weather systems and marine biodiversity.

3.2.11. Association Rule Mining

Association rule mining helps uncover relationships between different oceanographic variables. Green et al. employed this technique to explore correlations between ocean temperature, salinity, and phytoplankton abundance [9]. Their findings revealed significant associations that inform ecosystem management strategies.

3.2.12. Big Data Processing Frameworks

The scale of ocean circulation data necessitates the use of distributed computing frameworks like Hadoop and Spark, which facilitate efficient storage, processing, and analysis of large datasets. Consequently, distributed computing frameworks like Hadoop and Spark have become essential tools in oceanographic research and data analysis. Hadoop provides a reliable, scalable storage solution through its Hadoop Distributed File System (HDFS), enabling data to be stored across multiple nodes, thus reducing the burden on any single storage resource. This distributed architecture allows researchers to efficiently manage massive datasets while ensuring data redundancy and fault tolerance, which are critical for long-term oceanographic studies.

3.2.13. Hadoop

Hadoop's MapReduce framework enables parallel processing of extensive datasets, making it a popular choice in oceanographic research. Roy et al. utilized Hadoop to process satellite altimetry data for analyzing sea level rise patterns in the Pacific Ocean [10]. By leveraging Hadoop's distributed computing capabilities, they handled large volumes of data efficiently, yielding timely insights.

3.2.14. Apache Spark

Apache Spark, with its in-memory processing capabilities, provides even greater speed and efficiency. A recent study employed Spark to analyze real-time ocean current data from multiple sources, resulting in a significant reduction in computation time and enabling near real-time analysis. This capability is crucial for applications such as early warning systems for extreme weather events.

4. Ocean Circulation Data Visualization Techniques

4.1. Overview of Visualization Techniques

Data visualization is an essential aspect of ocean circulation analysis, allowing researchers and stakeholders to interpret complex datasets effectively. The primary goal of visualization is to

DOI: 10.54254/2753-8818/84/2025.21024

transform abstract data into intuitive visual representations, enhancing understanding and facilitating decision-making processes.

Visualization techniques are particularly important in oceanography, where the dynamics of ocean currents, temperature distributions, and salinity levels can be intricate and multifaceted. By employing various visualization methods, researchers can gain insights into ocean processes, monitor changes over time, and communicate findings to a broader audience.

4.2. 2D and 3D Visualization Methods

4.2.1. 2D Visualization Methods

Common 2D visualization methods include contour plots, vector field diagrams, and heat maps. These techniques effectively represent various oceanographic variables, such as temperature, salinity, and currents.

Contour Plots: These are used to illustrate the spatial distribution of oceanographic variables, such as temperature gradients. For instance, a contour plot can display sea surface temperature variations across the North Atlantic, helping to identify warm and cold fronts that influence weather patterns.

Vector Field Diagrams: Vector fields can visualize ocean currents by representing their direction and magnitude. By plotting vector arrows on a geographical map, researchers can illustrate the flow of currents and identify areas of convergence and divergence.

Heat Maps: Heat maps can represent concentrations of specific variables, such as chlorophyll-a, across a geographical area. These visualizations are particularly useful for monitoring phytoplankton blooms, which are vital indicators of marine ecosystem health.

4.2.2. 3D Visualization Methods

3D visualization techniques, including volumetric rendering and particle tracking, provide a more comprehensive representation of ocean circulation dynamics.

Volumetric Rendering: This technique allows for the visualization of three-dimensional flow structures within the ocean. By rendering data as a volume, researchers can explore complex flow patterns and interactions with the ocean floor, enhancing their understanding of how currents evolve.

Particle Tracking: Particle tracking is used to visualize the movement of water parcels over time. This technique can help illustrate how pollutants disperse in the ocean, which is crucial for environmental monitoring and management.

4.3. Dynamic and Interactive Visualization

Dynamic visualization techniques, such as animated time series and interactive dashboards, showcase the temporal evolution of ocean circulation.

Interactive visualization tools, such as web-based applications and virtual reality interfaces, allow users to explore ocean data actively. These tools enable researchers, educators, and policymakers to engage with data in meaningful ways, fostering a better understanding of ocean dynamics.

Web-based Tools: Platforms like Google Earth Engine enable users to visualize and analyze ocean data interactively. Users can manipulate layers, zoom into specific areas, and analyze changes over time, making data exploration accessible to non-experts.

Virtual Reality: Virtual reality technologies allow for immersive experiences where users can explore three-dimensional representations of ocean currents. Such tools enhance educational outreach and public engagement with oceanographic research.

DOI: 10.54254/2753-8818/84/2025.21024

4.4. Visualization Tools and Software

Several software tools are widely used for visualizing ocean data, providing researchers with robust options for creating informative and engaging visualizations.

4.4.1. R Language

R is a powerful tool for statistical computing and graphics. Packages such as ggplot2 and leaflet are commonly used for creating static and interactive visualizations. For example, ggplot2 allows users to create high-quality plots to represent oceanographic data effectively, while leaflet enables the development of interactive maps to visualize spatial data.

4.4.2. Python

Python, with libraries such as Matplotlib and Plotly, is also widely used for data visualization. Matplotlib is suitable for creating static plots, while Plotly excels in interactive visualizations. Researchers often utilize these libraries to analyze and present complex oceanographic datasets in a user-friendly manner.

4.4.3. Specialized Software

In addition to general-purpose programming languages, specialized ocean data visualization software, such as ParaView and Ocean Data View, offers advanced features for analyzing and presenting oceanographic data. ParaView, for instance, provides tools for handling large datasets and rendering complex 3D visualizations, making it suitable for high-resolution oceanographic studies.

5. Conclusion

In conclusion, the integration of big data analysis and visualization techniques has become essential for a deeper understanding of ocean circulation, particularly in the context of global climate change and its impacts on marine systems [11]. Traditional methods often fall short when confronted with the vast and complex datasets generated by modern oceanographic research. However, leveraging contemporary big data technologies, such as machine learning, deep learning, and distributed computing frameworks, enables researchers to analyze large datasets more effectively.

Visualization techniques further enhance the exploration and comprehension of ocean circulation system, providing intuitive representations of data. The development of 2D and 3D visualizations, dynamic animations, and interactive tools enriches our ability to interpret and communicate complex oceanographic information.

These advancements have led to significant progress in areas such as climate modeling, pollutant dispersion simulations, and the analysis of extreme weather events. As ocean data continues to grow, future innovations in data fusion, visualization technology, and data governance will be vital. These developments not only improve the accuracy of scientific research but also provide valuable insights for climate adaptation, environmental management, and decision-making in smart cities.

References

[1] Wikipedia Contributors. "Greenland Ice Sheet." Wikipedia, Wikimedia Foundation, 24 Aug. 2019, en.wikipedia.org/wiki/Greenland_ice sheet.

[2] O'Brien, J., Smith, P., & Chang, Y. (2019). Numerical simulations of tidal movements in coastal waters: A finite difference approach. Coastal Engineering, 150, 125-135.

[3] Gopalakrishnan, G., et al. (2020). Predicting sea surface temperature anomalies using Random Forest. Remote Sensing, 12(15), 2412.

[4] Zhang, Y., Chen, H., & Liu, M. (2019). Classifying chlorophyll-a concentrations using Support Vector Machines. Remote Sensing of Environment, 231, 111203.
[5] Chen, H., Zhang, L., & Zhao, J. (2021). Predicting ocean surface currents using neural networks. Ocean Dynamics, 71(3), 295-308.
[6] Shi, J., et al. (2020). Early detection of El Niño events using Convolutional Neural Networks. Climate Dynamics, 54(5-6), 1567-1584.
[7] Wang, L., et al. (2021). Long Short-Term Memory networks for ocean temperature prediction. Journal of Atmospheric and Oceanic Technology, 38(2), 195-210.
[8] Liu, Y., Zhang, X., & Liu, Z. (2022). Identifying long-term trends in ocean currents related to climate change. Journal of Marine Systems, 228, 103625.
[9] Green, J., Smith, T., & Lee, R. (2020). Exploring relationships between ocean variables using association rule mining. Marine Ecology Progress Series, 642, 1-12.
[10] Roy, S., et al. (2021). Analyzing sea level rise patterns in the Pacific using Hadoop. Environmental Science & Technology, 55(10), 6891-6900.
[11] Smith, Hayley. "Summer 2024 Was Earth's Hottest on Record." Los Angeles Times, 6 Sept. 2024, www.latimes.com/environment/story/2024-09-06/summer-2024-was-earths-hottest-on-record.