remote sensing

Article

PoSDMS: A Mining System for Oceanic Dynamics with Time Series of Raster-Formatted Datasets

Lianwei Li [1], Cunjin Xue [2,3,*], Yangfeng Xu [1,2,3], Chengbin Wu [4,5] and Chaoran Niu [2,3,4]

[1] College of Oceanography and Space Informatics, China University of Petroleum, Qingdao 266580, China; 20050046@upc.edu.cn (L.L.); z19160008@s.upc.edu.cn (Y.X.)
[2] International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China; 1801050222@s.upc.edu.cn
[3] Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
[4] University of Chinese Academy of Sciences, Beijing 100049, China; wucb.19s@igsnrr.ac.cn
[5] State Key Laboratory of Resource and Environmental Information System, Institute of Geographical Science and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
* Correspondence: xuecj@aircas.ac.cn; Tel.: +86-10-82178126; Fax: +86-10-82178177

Abstract: Many effective and advanced methods have been developed to explore oceanic dynamics using time series of raster-formatted datasets; however, they have generally been designed at a scale suitable for data observation and used independently of each other, despite the potential advantages of combining different modules into an integrated system at a scale suited for dynamic evolution. From raster-formatted datasets to marine knowledge, we developed and integrated several mining algorithms at a dynamic evolutionary scale and combined them into six modules: a module of raster-formatted dataset pretreatment; a module of process-oriented object extraction; a module of process-oriented representation and management (process-oriented graph database); a module of process-oriented clustering; a module of process-oriented association rule mining; and a module of process-oriented visualization. On the basis of such modules, we developed a process-oriented spatiotemporal dynamic mining system named PoSDMS (Process-oriented Spatiotemporal Dynamics Mining System). PoSDMS was designed to have the capacity to deal with at least six environments of marine anomalies with 40 years of raster-formatted datasets, including their extraction, representation, storage, clustering, association and visualization. The effectiveness of the integrated system was evaluated in a case study of sea surface temperature datasets during the period from January 1982 to December 2021 in global oceans. The main contribution of this study was the development of a mining system at a scale suited for dynamic evolution, providing an analyzing platform or tool to deal with time series of raster-formatted datasets to aid in obtaining marine knowledge.

Keywords: marine mining system; spatiotemporal dynamics; process-oriented; marine anomaly variation; raster-formatted datasets

check for updates

Citation: Li, L.; Xue, C.; Xu, Y.; Wu, C.; Niu, C. PoSDMS: A Mining System for Oceanic Dynamics with Time Series of Raster-Formatted Datasets. *Remote Sens.* **2022**, *14*, 2991. https://doi.org/10.3390/rs14132991

Academic Editors: Peng Liu, Guojin He, Kie B. Eom and Mohd Anul Haq

Received: 21 April 2022
Accepted: 21 June 2022
Published: 22 June 2022

1. Introduction

The world is geographically dynamic [1,2], and this dynamism has drawn increasing attention in recent years. This attention has focused on dynamic object extraction and analysis, dynamic mining methods, mining frameworks and tools [3–6] and especially on oceanic dynamics [3,7–10]. Series of images taken by advanced Earth-observing technologies over long periods of time, combined with historical climate records, constitute the main source of continuous and consistent information about the marine environment [11,12] and offer new opportunities for monitoring oceanic dynamics and understanding their evolutionary patterns [10]. These evolutionary patterns generally have lifespans ranging over generations through development, merging, splitting and dissipation [13,14], playing significant roles in regional and global climate change [15,16].

In recent decades, a large number of models and methods with time series of raster-formatted datasets have been proposed to obtain geographical dynamics in the form of objects, events or processes [17–21]. These have been proposed to analyze their dynamic characteristics [22–24] and explore their clustering patterns, association patterns and evolutionary patterns [8,25,26]. Differing from geographical dynamics, however, these methods and models lacked effective and useful platforms or tools with which to obtain oceanic dynamic knowledge directly from time series of raster-formatted datasets. Regarding this hurdle, many object/event/process-oriented methods were proposed. Object-oriented image analysis technologies were widely used to identify instantaneous geographical change as a snapshot object [21,27]. As they were discrete, there were no evolutionary relationships between successive snapshot objects. To overcome this deficiency, event-oriented models were designed to handle the geographical dynamics, e.g., a rainstorm event [17,28], a flood event [29], a 4-dust storm [18], a marine heatwave [20], etc. Generally, a scale of geographical dynamic evolution is out of step with the scale of data observation; some evolutionary relationships among geographical dynamics will be lost. To obtain true evolutionary relationships, a process-oriented idea was proposed to obtain marine anomalies [13,19,26]; to represent and analyze such dynamic changes, the graph-based model was used for urban hot islands [30], rainstorms [17,31], land use and cover changes [12] and oceanic eddies [3,9].

As an important component of dynamic mining, spatiotemporal clustering analysis aims to find clusters of the same properties in both time and space, and recently gained attention as a means to discover oceanic patterns [32–34]. Popular clustering algorithms, such as K-Mean, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and SRNN (Shared Reciprocal Nearest Neighborhood), were expanded and widely used to obtain oceanic clustering patterns [35–38]. To obtain oceanic dynamic information within successive time snapshots, Liu et al. proposed a process-oriented clustering method of treating spatiotemporal dynamics as a trajectory [26], with the trajectory clustering representing an evolutionary pattern. Regarding geographical association rule mining: the quantitative Apriori algorithm is a classical algorithm for obtaining richer information [39]. Combining this algorithm with geographical spatiotemporal characteristics, several algorithms were expanded on the basis of the quantitative Apriori, e.g., a cluster-based association rule (CBAR) [40] and a mutual-information-based quantitative association rule-mining algorithm (MIQarma) [41]. Meanwhile, FP-Tree (a non-Apriori algorithm) and its expansions were also developed for mining association patterns [42]. For oceanic dynamic association patterns, Saulquin et al. designed an event-based mining algorithm to deal with SST anomalies (SSTA) relative to El Niño–Southern Oscillation (ENSO) events [7].

In the field of spatiotemporal mining frameworks, Lee and Lee's proposal included a two-tier knowledge discovery model that integrated a foundation model and an executing model [43]. Compieta et al. and Bertolotto et al. designed a three-layer mining and visualizing architecture that included a data layer, an application layer and a visualization layer to reveal spatial and temporal patterns of natural phenomena [44,45]. Yoo and Bow designed different mining frameworks to deal with two-parameter constraints and find spatially interesting colocation patterns [46]. Xue et al. discussed pixel- and object-based spatiotemporal mining frameworks to deal with marine abnormal association patterns [8]. To reveal the dynamic characteristics of geographic phenomena and discover their association patterns, they proposed an event-based spatiotemporal association rule mining framework with two strategies; a sequence and an episode [47]. To visualize marine dynamic environments, the three-dimensional temporal-spatial process visualization component based on particle system was designed [48], also the interactive multi-scale, multivariate visualization system with a unified visual data service and a component-based visualization structure was developed [49].

On the basis of the aforementioned mining models, algorithms and frameworks, several considerable mining systems and operational tools have been developed to transform raster-formatted datasets into geographical knowledge. For example, Korting et al. proposed and designed a new toolbox, GeoDMA (Geographic Data Mining Analyst), which integrated a series of modules, including segmentation, feature extraction, feature selection, landscape and multi-temporal features, as well as data mining for pattern recognition and multi-temporal analysis of remote sensing imagery [50]. GeoDMA used decision-tree strategies adapted for spatial data mining and connected remotely sensed imagery with other geographic data types using access to local or remote database. The Argo-MIS designed a service module, a knowledge discovery module, an operational storge module, a notificaion module, a graphical user interface and an environmental decision module to support marine environmental monitoring [51]. Romani et al. developed the RemoteAgri system to discover the Plateau–Valley–Mountain (P–V–M) association patterns, used for monitoring sugar cane fields via time series of remote sensing images [52]. The main function of RemoteAgri consisted of time series of extraction and time series of pattern exploration. Xue et al. developed the image-driven remote-sensing mining system RSMapMining to explore marine knowledge from remote sensing images, which integrated an image preprocessing module, a pattern mining module and a knowledge visualization module [53].

Oceanic dynamics reveal not only when and where marine environmental parameters change, but also how they evolve in space and time [3,19,25], and these dynamics can help to better understand global climate change, e.g., via an evolution of SSTA in space to define a new ENSO index for identifying ENSO types [54,55]. In spite of the considerable achievements in mining frameworks, methods and tools [8,43,44,50], there is still the great challenge of addressing when, where and how the evolution of oceanic dynamics occurs with time series of raster-formatted datasets. Thus, this paper developed a mining system at a scale of dynamic evolution by integrating existing popular techniques and methods. The integrated system was called PoSDMS (Process-oriented Spatiotemporal Dynamics Mining System). A scale of dynamic evolution refers to the lifespan of oceanic phenomena, or an object, from production through development to dissipation. This is a time duration and differs from a scale of data observation, which is generally a time snapshot. The main aim of PoSDMS as a platform was for the end user to be able to explore oceanic dynamic knowledge from raster-formatted datasets. The three main contributions of PoSDMS to marine spatial information science were as follows:

- Using a scale of dynamic evolution, rather than a scale of data observation, as a unit to integrate popular mining algorithms and models, PoSDMS ensured the integrity of spatial structure, temporal evolution and thematic characteristics when dealing with oceanic dynamics.
- PoSDMS developed an automatic/semi-automatic technical workflow of obtaining oceanic dynamics knowledge from time series of raster-formatted datasets.
- Providing an analyzing platform capable of dealing with marine anomalies at a scale of dynamic evolution, PoSDMS supported data-driven mechanisms in research of marine environmental changes.

The remainder of this paper is organized as follows. Section 2 introduces basic concepts about the marine spatiotemporal process, describes the process-oriented mining architecture and discusses its key technologies. Section 3 outlines the design of the modules and their logics, gives their technical workflows and integrates the modules in order to develop PoSDMS. Section 4 considers a monthly sea surface temperature (SST) dataset as a case study to evaluate the functions and performances of PoSDMS. Finally, discussions and conclusions are presented in Section 5.

2. Process-Oriented Mining System Architecture and Key Technologies

2.1. Basic Concepts

PoSDMS was aimed at developing a semi-/auto-analyzing system for exploring oceanic dynamics at an evolutionary scale with time series of raster-formatted datasets. As the evolution scale is generally different from the data observation scale, before designing the PoSDMS, some concepts about oceanic dynamics and evolution needed to be addressed from the perspective of system development.

Marine anomaly variation refers to the abnormal increase or decrease of marine environmental parameters relative to the mean status of long time series, which cover a specified spatial domain for a specified time range [8], e.g., monthly SSTA, seasonal sea surface salinity anomalies, etc.

Marine evolution process refers to the lifespan of a marine anomaly variation, which has a property of evolution from production through development to dissipation in space and time [13].

Marine process object refers to an object abstracted from a marine evolution process, which consists of marine snapshot objects and their evolutionary relationships among successive time snapshots.

Marine snapshot object refers to an object of marine anomaly variation at a specified time snapshot, generally a time of data observation, e.g., a snapshot object of SSTA at a snapshot time of satellite passing territory.

Marine evolution relationship refers to one of four relationships between successive time snapshots, i.e., a development, a merging, a splitting and a splitting–merging [14].

2.2. Spatiotemporal Dynamic Mining System Architecture with Raster-Formatted Datasets

Based on raster-formatted datasets, PoSDMS aimed to offer a platform for the discovery of dynamic evolutions and association patterns of marine environmental parameters at global and regional scales. Thus, PoSDMS was designed with four layers. From bottom to top, these were: a data layer, a technology layer, a function layer and an application layer, as shown in Figure 1. The data layer was responsible for data management from raster-formatted datasets through the middle datasets, and to the vector-formatted object datasets in a database, the data foundation of PoSDMS. The technology layer included raster processing technologies, GIS (Geographic Information System) spatial and temporal analysis technologies, process-oriented object extraction, representation, storage and data mining technologies and graph-based database technologies. These technologies supported the designs of the functions in the function layer, which included data pretreatment of raster-formatted datasets, extraction, representation, storage and management of process objects, clustering and association rule mining of process objects and visualization of process objects. In the application layer, PoSDMS explored the dynamic evolution patterns of marine environmental parameters, i.e., SST, sea surface salinity, sea surface precipitation and their association patterns with typical signs of global climate change, e.g., ENSO, PDO (Pacific decadal oscillation).

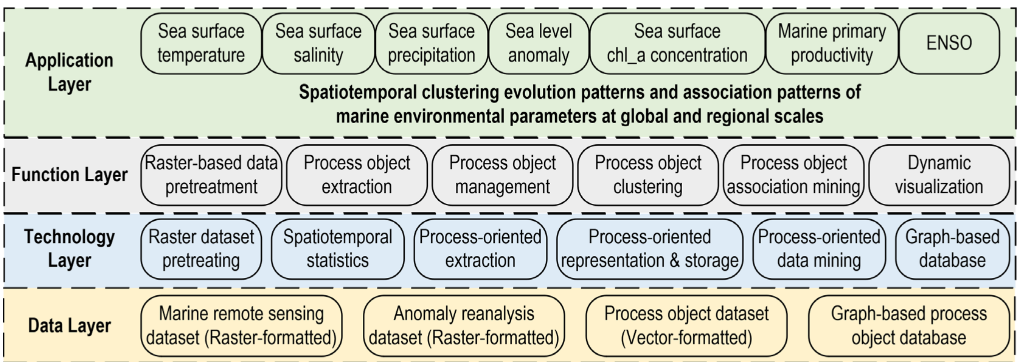

Figure 1. System architecture.

2.3. Key Technologies and Their Implementations

PoSDMS aimed to explore oceanic dynamics with time series of raster-formatted datasets. As oceanic dynamics require much more attention to be paid to evolutionary relationships rather than to static patterns, during the development of PoSDMS, two key technologies needed be addressed: first, how to represent and store marine snapshot objects and their evolutionary relationships, which determine the performance and efficiency of oceanic dynamics management; second, to design a process-oriented mining method to support the extraction, clustering, association rule mining and visualization of oceanic dynamics.

To deal with the first technology, PoSDMS integrated a process-oriented graph model to represent and store oceanic dynamics and their evolutionary relationships. The process-oriented graph model defined four types of nodes, i.e., a process node, a sequence node, a linked node and a state node, to represent and store marine objects, and two types of edges, i.e., an inclusionary and an evolutionary relationship, to represent and store a relationship between two objects [14]. During development, the Neo4j-based process-oriented graph database was built. Using an index-free adjacency to describe the relationships between objects [56], the process-oriented graph database performed one order of magnitude better in querying the spatial evolution of marine anomaly variations than Oracle, the object-relational database.

Regarding dynamic mining technology, the PoSDMS design involved a hierarchical mining strategy based on the process semantics of the marine process–evolution sequence-snapshot state [13]. The hierarchical mining strategy provided a foundation for the designs of the module used to obtain marine process objects and their evolutions [19], the module for integrating process-oriented similarity measuring functions and expanding clustering algorithms [26], and the module for constructing process-oriented mining transaction tables and expanding association rule mining algorithms [10], and the module for visualizing oceanic dynamics and their evolutionary relationships. Taking spatial information, thematic characteristics and their evolution in time into full consideration, the process-oriented mining technology improved the investigative capacity of PoSDMS to deal with oceanic dynamics.

3. Design and Implementation of PoSDMS

3.1. Modules and Their Logics

From raster-formatted datasets to marine knowledge, the principal capabilities of the PoSDMS were as follows:

1. It offered a set of tools to deal with large amounts and types of raster-formatted datasets with different spatial resolutions and different temporal resolutions. The data formats included, but were not limited to, the common GeoTiff, NetCDF (network Common Data Form), HDF4 (Hierarchical data format), HDF5 and HFA (Erdas imagine img).
2. It built a workflow through which to obtain marine anomaly variations in the form of process objects at global scale and efficiently managed them.
3. It explored dynamic evolution patterns and association patterns among marine environmental parameters, which included, but were not limited to, SST, sea surface salinity, sea surface precipitation, sea level anomaly, sea surface chl_*a* concentration and marine primary productivity.
4. It offered a flexible visualization component for the display of spatial and thematic characteristics of oceanic dynamics at a scale from process and sequence to snapshot in time, as well as their evolutionary relationships.

To achieve the aforementioned functions, PoSDMS developed six modules on the basis of marine spatiotemporal process semantics and graph database technologies: raster-formatted dataset pretreatment, process-oriented object extraction, data management, i.e., process-oriented graph database, process-oriented object clustering, process-oriented association rule mining and process-oriented visualization. These modules and their logics are shown in Figure 2.

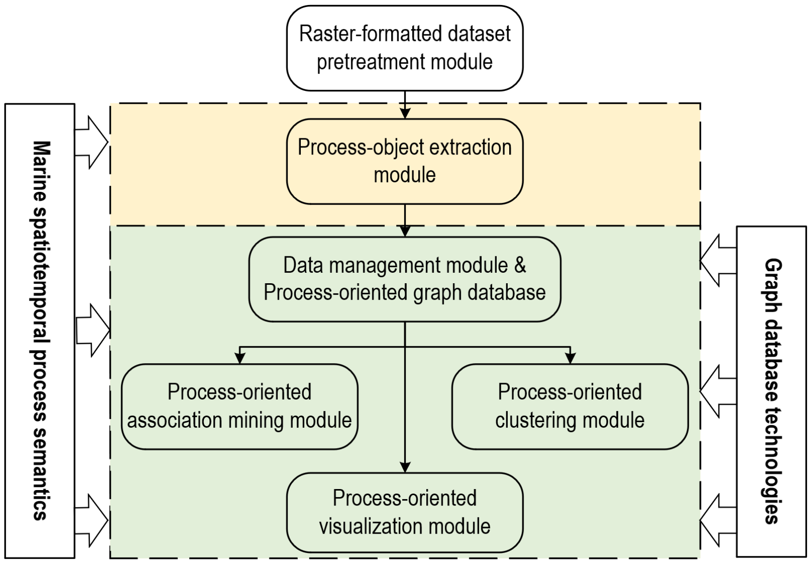

Figure 2. Modules and their logics.

3.2. Module Development and Integration

3.2.1. Raster-Formatted Dataset Pretreatment Module

The objective of this module transformed different types of datasets with different spatial resolutions and different temporal resolutions into time series of raster-formatted datasets within a uniform spatial and temporal resolution. The marine raster-formatted dataset mainly came from remote sensing images and historical climate records, which were stored in many formatted types, e.g., NetCDF, HDF4, HDF5, GeoTiff, etc. Additionally, the initial data were gathered at greatly different intervals, ranging from daily to annual; their spatial resolutions varied from meters to kilometers and even to global scales. To produce uniform datasets, PoSDMS developed a spatiotemporal transforming module by integrating a middle plug-in GDAL (Geospatial Data Abstraction Library) and defining a custom-defined data format based on HDF4 as a final file format. This module included four core algorithms, i.e., a spatiotemporal slicing algorithm, a spatiotemporal resampling algorithm, a spatiotemporal interpolating algorithm and a standard monthly averaged anomaly algorithm. Based on the GDAL plug-in, this module dealt with widely used raster formats including, but not limited to, NetCDF, HDF4, HDF5, GeoTiff, RST (Idrisi Raster format), HFA and GRASS Raster format. Figure 3 shows the workflow of the raster-formatted dataset pretreatment module, and Figure 4 gives an example showing a function and an interface of the spatial resampling algorithm.

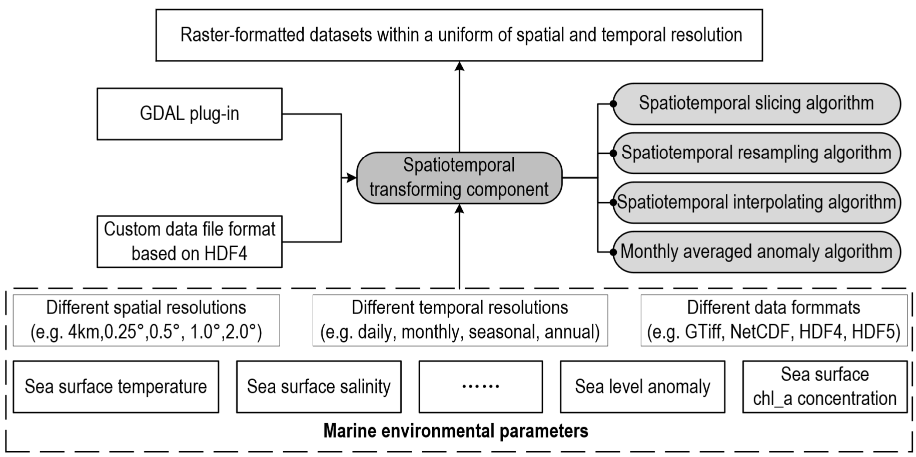

Figure 3. Workflow of raster-formatted dataset pretreatment module.

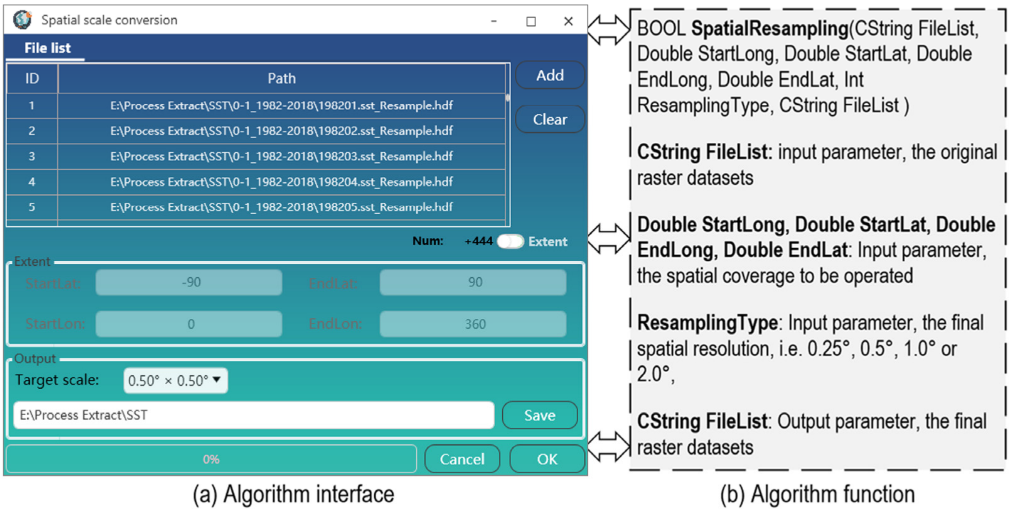

Figure 4. Example of spatial resampling algorithm function and interface.

3.2.2. Process-Oriented Extraction Module

From a time series of marine raster-formatted datasets within a uniform spatial and temporal resolution, this module utilized an integrated process-oriented extraction method to generate marine process objects and their evolutionary relationships. The idea of a process-oriented mining method was proposed by Xue et al. [13,14] and used to obtain rainstorm objects [31] and evolutionary objects of SSTA [19]. According to the marine process semantics of "marine process-evolution sequence-snapshot state", this module was divided into four algorithms, step by step. The four algorithms consisted of an extraction of snapshot objects from time series of raster-formatted datasets, a track of snapshot objects at successive time snapshots in an evolutionary sequence object, a reconstruction of process objects from sequence objects and an identification of evolutionary relationships from process objects. Figure 5 shows the workflow of the process-oriented extraction module.

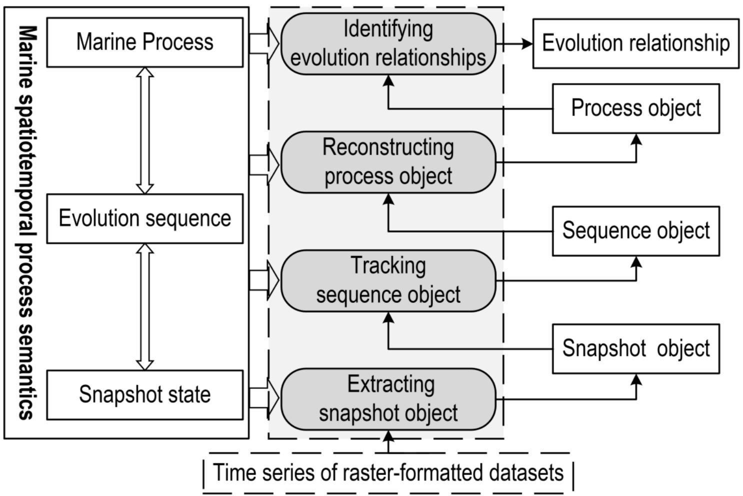

Figure 5. Workflow of process-oriented extraction module.

3.2.3. Process-Oriented Graph Database

According to the process-oriented node-edge storage structure [14], PoSDMS built the Neo4j-based graph database to store, manage and display marine process objects and their evolutionary relationships. Based on the process-oriented graph database, we developed an inputting and updating algorithm, an inquiring and searching algorithm, and the process-oriented visualizing strategy to manage and display marine process objects and their evolutions. Figure 6 shows their technical workflow.

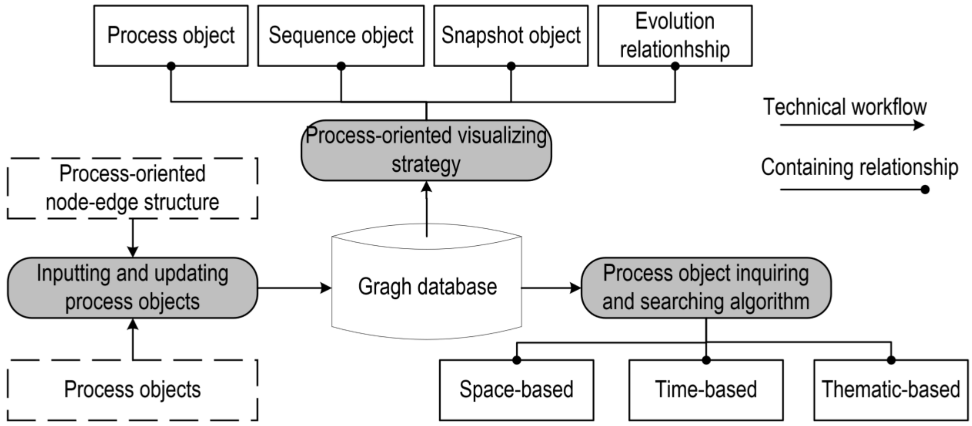

Figure 6. Workflow of building a process-oriented graph database, as well as its management and visualization.

The inputting and updating algorithm input process objects into the graph database with the node-edge structure, which included sub-algorithms for inputting process objects, labeling node types (i.e., a process node, a sequence node or a snapshot node) and building evolutionary relationships between nodes. The inquiring and searching algorithm designed the interfaces based on spatial structure, temporal evolution, thematic characteristics and their combinations to obtain user-interested process objects from the graph database. The visualization module designed a strategy for displaying process objects, sequence objects and snapshot objects using time series of spatial views, as well as a strategy for displaying evolutionary relationships using a node-edge view.

3.2.4. Process-Oriented Clustering Module

To discover clustering patterns of oceanic dynamics from marine process objects, this module expanded the clustering idea by defining new concepts about process neighborhoods and process similarities. The process-oriented clustering method included process-oriented similarity measurement functions and process similarity-based clustering algorithms. All these functions and algorithms were based on marine process semantics [13]. Figure 7 shows the workflow of the process-oriented clustering module.

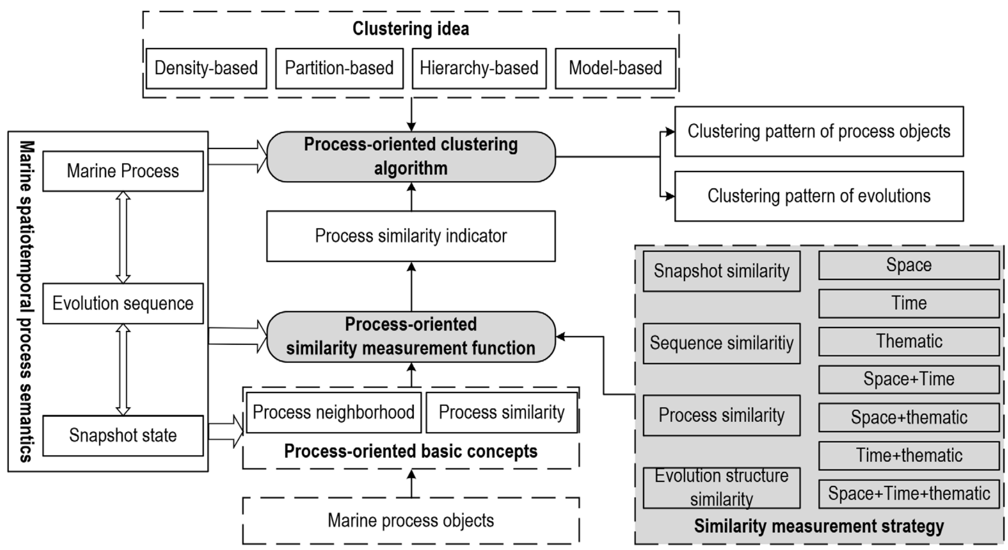

Figure 7. Workflow of process-oriented clustering module.

During the implementation of process-oriented similarity measurement functions, this module designed four types of similarity measurements according to marine process semantics, i.e., snapshot similarity, sequence similarity, process similarity and evolutionary structure similarity. Meanwhile, each type of similarity measurement could consider spatial, temporal and thematic characteristic or their different combinations. Thus, this module developed 28 similarity measurement functions in total. Using these process similarity indicators, the module expanded four process-oriented clustering algorithms based on K-mean, DBSCAN [57], SRNN [37] and DcSTCA (Dual-constraint SpatioTemporal Clustering Approach) [38].

3.2.5. Process-Oriented Association Rule Mining Module

This module was aimed at discovering the association rules among marine environmental parameters or between signals of global climate changes, e.g., ENSO and PDO. In the technical implementation of association rule mining, there were two key issues. One was the need to construct a mining transaction table, and the other was to design a mining algorithm. To achieve dynamic association patterns, this module fully considered marine process semantics to redefine concepts of support, confidence and lift. It also designed a process-oriented mining transaction table and process–sequence–state mining strategies. Figure 8 shows the workflow of the process-oriented association rule mining module.

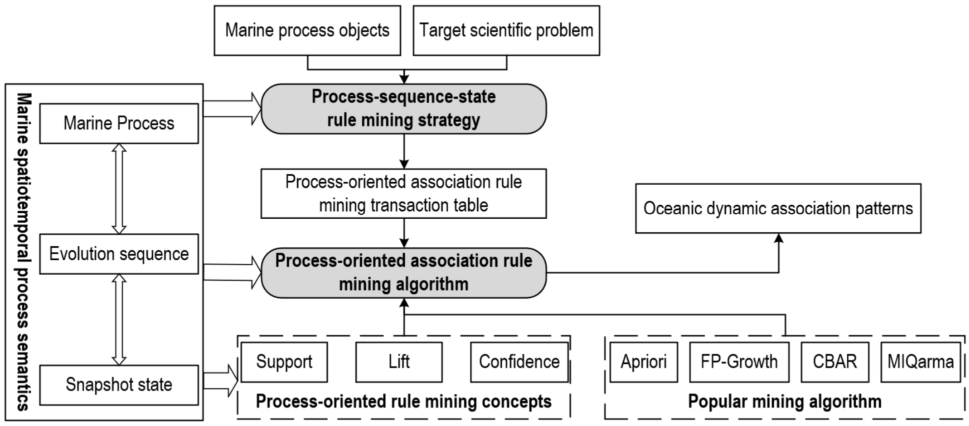

Figure 8. Workflow of process-oriented association rule mining module.

During the construction of the mining transaction table, in order to discover dynamic association patterns at different evolutionary phases (e.g., before, after, or within the lifespan), this module designed an interactive rule mining strategy to store a process object, a sequence object or a snapshot object as one record. This included spatial, temporal and thematic information. Regarding the mining algorithm, this module redefined the concepts of support, lift and confidence using the lifespan of an object, not a snapshot in time, and expanded four process-oriented association rule mining algorithms based on the idea of popular algorithms, i.e., Apriori [39], CBAR [40], FP-Tree [42] and MiQarma [41].

3.2.6. System Functions of PoSDMS

All the modules described in this paper were developed by Visual Studio 2018 and the third-party plug-ins GDAL version 2.0.2, Neo4j version 4.1.3 and ArcGIS runtime version 10.0. The tested hardware environment included an Intel core i7 CPU at 2.80 GHz, a 500 GB hard disk and 4.0 GB of memory. PoSDMS integrated all the modules to design six system functions for: graph database management and visualization, spatiotemporal scale transformation, spatiotemporal anomaly detection, marine process object extraction, marine process object clustering and marine process object association rule mining. The logics between these functions and their corresponding modules are shown in Figure 9. PoSDMS is a Windows-based standalone software which was developed by the authors and registered by the National Copyright Administration of P.R. China (No. 2022SR0406836).

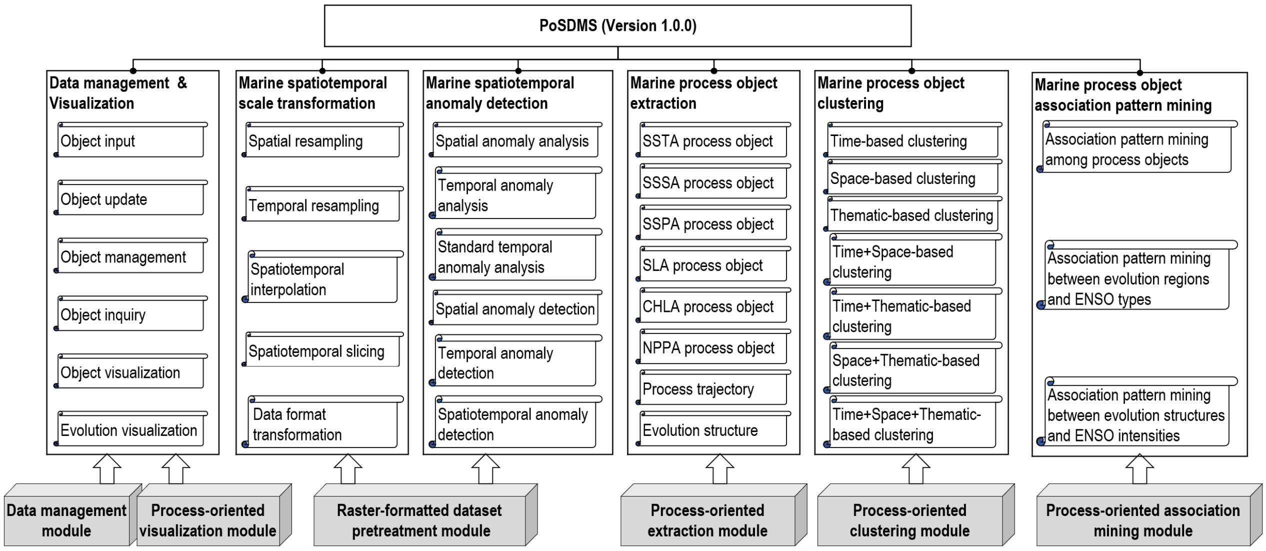

Figure 9. Logics between functions and modules of PoSDMS.

4. Case Study of Dynamic Analysis of SSTA in Global Ocean

PoSDMS buildt a process-oriented graph database based on Neo4j, named PoGDB, which stored 6 kinds of process objects of marine environmental parameters, i.e., SST, sea surface salinity, sea surface precipitation, sea level anomaly, sea surface chl_*a* concentration and marine primary productivity. PoSDMS was also capable of dealing with the six marine environments from their dynamic representation and storage through extraction, analysis, exploration and visualization. This paper took SST as a case study to evaluate the function and performance of PoSDMS.

4.1. Raster-Formatted SST Dataset and Its Pretreatment

The SST remote sensing dataset covered the period from January 1982 to December 2021 with a spatial resolution of 1° and a temporal resolution of 1 month; it was obtained from the NOAA Optimum Interpolation Sea Surface Temperature V2.0 provided by the NOAA/OAR/ESRL Physical Sciences Division, Boulder, Colorado, USA, and is available at http://www.esrl.noaa.gov/psd/ (accessed on 1 March 2022) [58]. The function ***Standard temporal anomaly analysis***, based on a standard monthly average anomaly algorithm, denoted as the *z*-score [59], was used to remove seasonal variations of SST that were mainly dominated by solar radiance. Thus, the monthly global SSTA dataset during the period of January 1982 to December 2021 was generated within a uniform of spatial and temporal resolution.

4.2. Process Objects of SSTA, Process-Oriented Graph Database and Visualization

Using the raster-formatted SSTA dataset aforementioned, the function ***SSTA process object***, in which the parameters were set similarly to PoAIES (Process-oriented Approach to Identify Evolution of SSTA) [19], was carried out to generated 417 process objects, 1108 sequence objects and 3687 snapshot objects, as well as 2738 development relationships, 275 merging relationships, 309 splitting relationships and 28 splitting-merging relationships. All the objects and relationships were stored into the PoGDB through the function ***Object Input***. Figure 10 shows the hierarchical view of process objects in PoGDB.

In Figure 10, there were six marine environmental parameters; each parameter was divided into two categories according to whether the anomaly variation was higher or lower than the mean value. Thus, there were 12 nodes of marine environmental parameters. The node WSST (Warmer SST, in which the SSTA value was higher than the mean value) consisted of 230 nodes of process objects of SSTA (partial shown in Figure 10). The process object SSTA with POID = 782 (Process Object ID) included 13 sequence objects. The sequence object with SOID = 782_10 included 8 snapshot objects. Among sequence objects and snapshot objects, their evolutionary relationships were clearly shown.

Based on the PoGDB, PoSDMS designed two visualization interfaces, the ***Object visualization*** and the ***Evolution visualization***. ***Object visualization*** displayed dynamic information of SSTA in a time series of snapshots, which focused more on changes of SSTA in space than evolution in time, while ***Evolution visualization*** took a node-edge to display the evolution of SSTA in time, with no consideration of the spatial structure of SSTA. Thus, the combination of two visualization interfaces integrated their respective advantages to displace the evolution of SSTA in space and time. Figures 11 and 12 display the dynamic evolutions of process objects of SSTA with POID = 782 using the object visualization interface and evolution visualization interface.

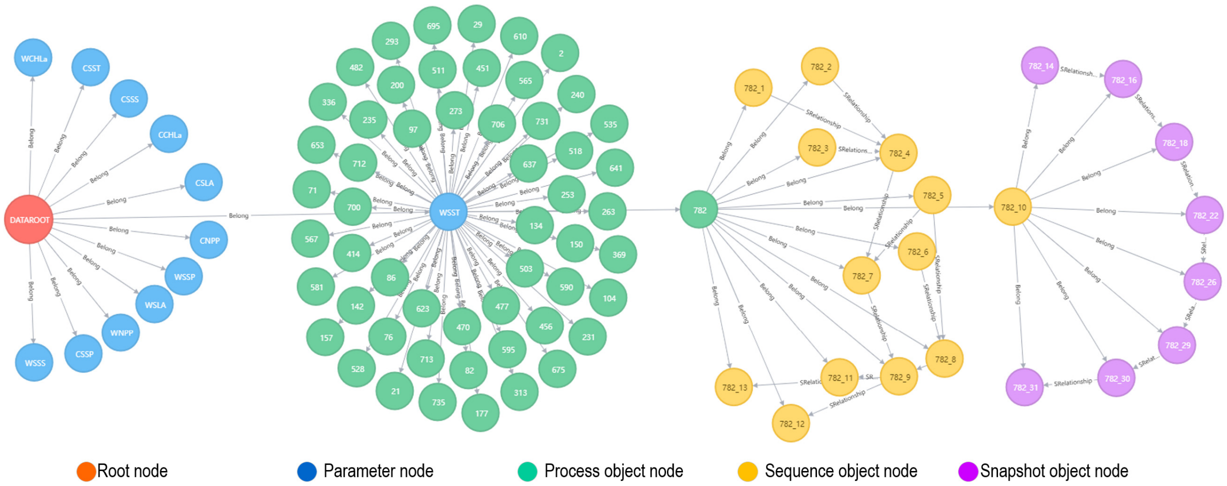

Figure 10. Hierarchical view structure of PoGDB.

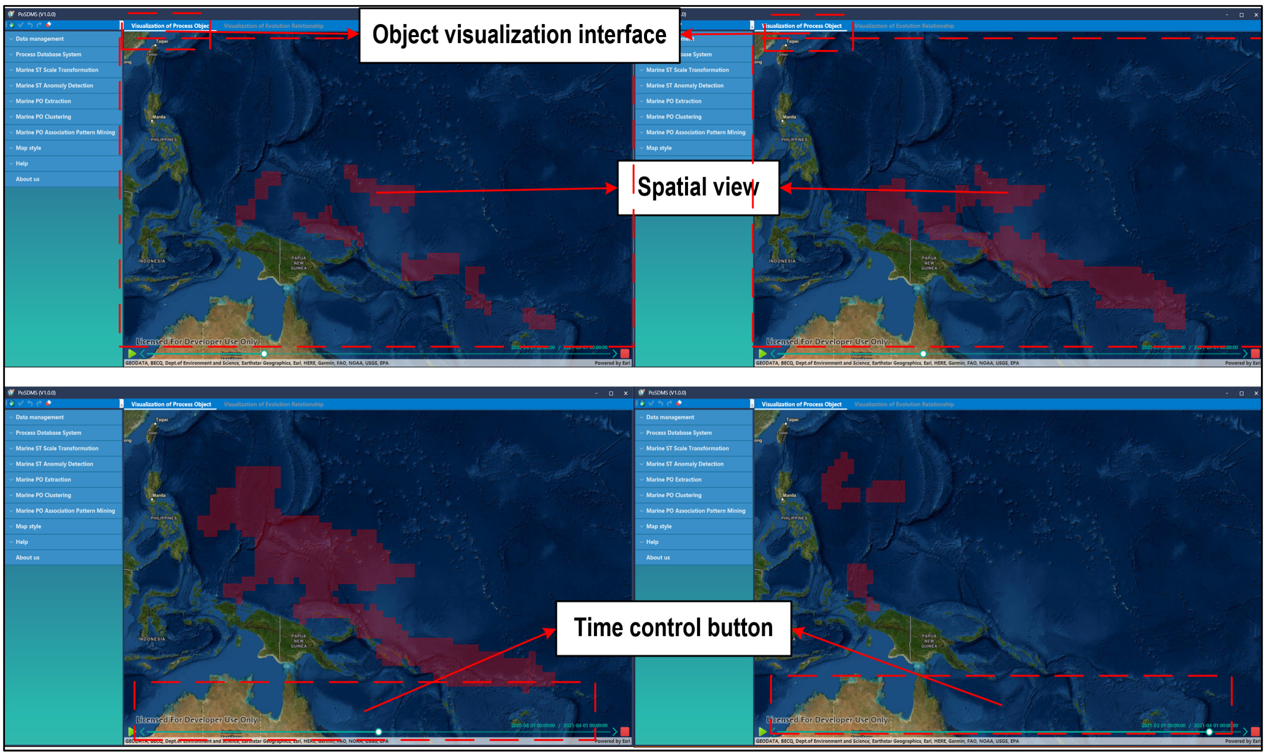

Figure 11. Spatial distribution of a specified process object of SSTA at different time snapshots.

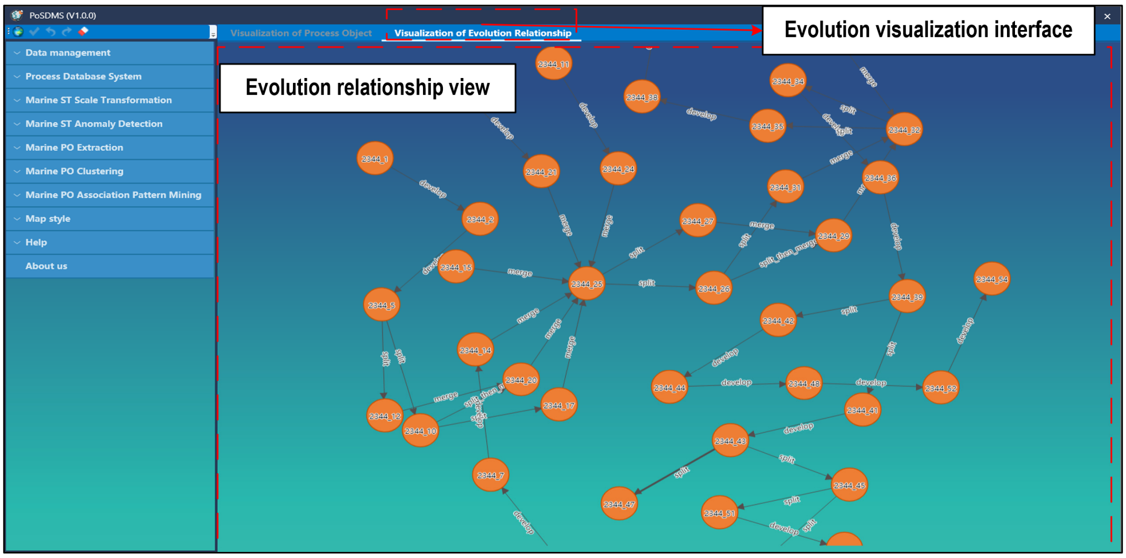

Figure 12. Evolutionary relationship of a specified process object of SSTA.

4.3. Clustering Pattern of SSTA Evolutions

To more thoroughly explore and analyze evolution patterns of SSTA, a process object of SSTA was generalized into a trajectory of SSTA in space and time. Thus, we developed a process-oriented spatiotemporal trajectory clustering method named PoSCM (Process-oriented Spatiotemporal Clustering Method). PoSCM calculated similarity measurements according to combinations of spatial, temporal and thematic characteristics of process objects of SSTA, and used a DBSCAN-based clustering algorithm to obtain evolution patterns of SSTA. Figure 13 shows the interface of the process-oriented trajectory clustering algorithm. Here, the clustering parameters were set the same as in Ref [26].

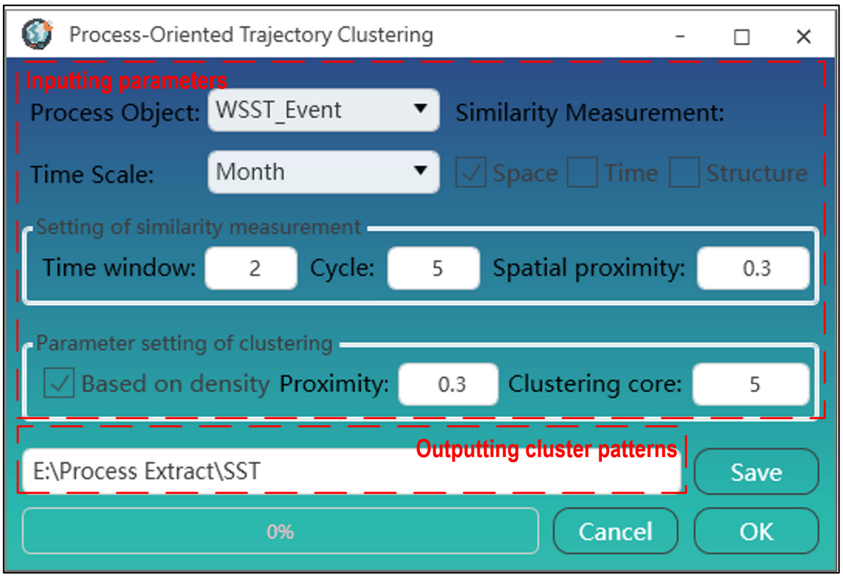

Figure 13. Interface of process-oriented clustering function.

Different combinations of similarity measurements based on spatial, temporal and thematic characteristics will meet with different scientific problems. Figure 14 shows clustering patterns in the Pacific Ocean while only considering thematic similarities of process objects of SSTA. In Figure 14, the origin and destination of the trajectory are the place of origin and dissipation of the process object of SSTA. The arrow indicates an evolutionary direction of SSTA, and the evolutionary structure shows the place of origin, place of dissipation, evolutionary relationships (developing, merging, splitting and splitting-merging relationships) and the places they occurred. To aid in finding more meaningful marine knowledge, each clustering pattern will require further exploration.

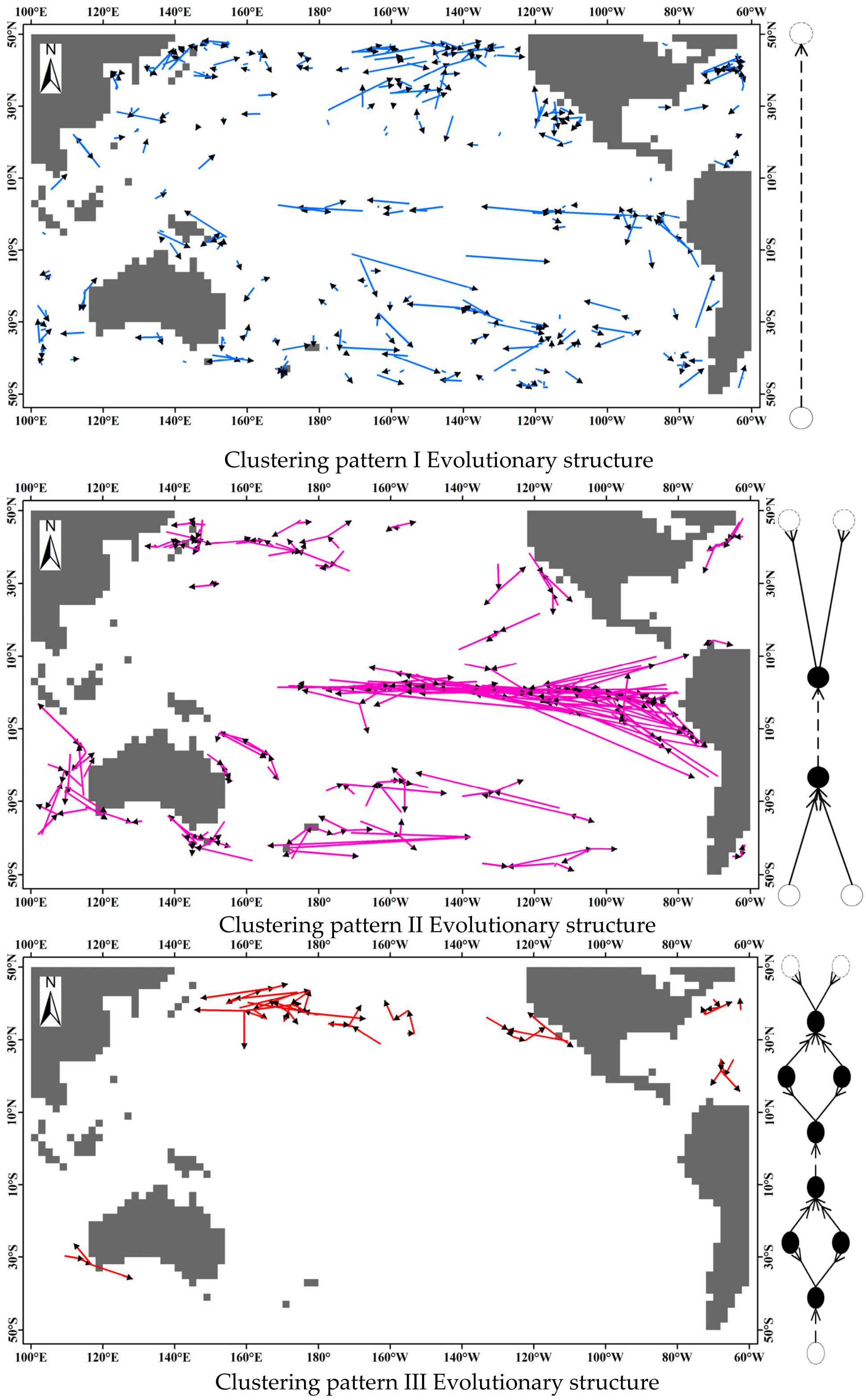

Figure 14. Clustering patterns in the Pacific Ocean, considering only thematic characteristics of process objects of SSTA.

5. Conclusions

Advanced Earth observation technologies could provide marine environmental parameters at large scales over long time periods, thus facilitating studies of their dynamic evolutions in space and time. To deal with such dynamic evolution discovery in marine environmental analyses, we developed a mining system at a scale of evolution, rather than of data observation, called PoSDMS. This system aimed to allow automatic/semi-automatic marine environmental analysis from raster-formatted datasets to provide new knowledge. This was achieved through six modules: a module of raster-formatted dataset pretreatment, a module of process-oriented extraction, a process-oriented graph database, a module of process-oriented visualization, a module of process-oriented clustering and a module of process-oriented association rule mining. These modules were developed on the basis of marine spatiotemporal process semantics [13,19] and graph database techniques [56] for processing raster-formatted datasets, extracting and storing oceanic dynamics, implementing mining processes and designing visualization interfaces. The detailed are listed as follows:

Raster-formatted datasets pretreatment module. By integrating the GDAL plug-in, the module dealt with large amounts of raster data formats, including, but not limited to, NetCDF, HDF4, HDF5, GeoTiff, RST, HFA, ASCII DEM and GRASS Raster format. The integrated algorithms solved the transformations with four spatial scales of 0.25°, 0.5°, 1.0° and 2.0° and three temporal scales of month, season and year, including geographical spatiotemporal statistics, spatiotemporal interpolation and standard monthly, seasonal and annual averaged anomalies [59].

Process-oriented graph database and Process-oriented oceanic dynamic management module, extraction module and visualization module. These modules integrated marine spatiotemporal process semantics [13], PoTGM (Process-oriented Two-tier graph Model) [14], PoAIR (Process-oriented Approach for Identifying Rainstorm) [31] and PoAIES [19] to obtain, represent, store and manage six types of anomalies of marine environmental parameters, i.e., SST, sea surface salinity, sea surface precipitation, sea level anomaly, sea surface chl_*a* concentration and marine primary productivity. Furthermore, these modules have an expanding capacity for handling other marine environmental parameters.

Process-oriented clustering module. This module designed interactive strategies for calculating similarity measurements based on spatial, temporal and thematic characteristics of process objects. The integrated algorithms included K-means, DBSCAN [57], SNN [37], DcSTCA [38] and PoSCM [26].

Process-oriented association rule mining module. By redefining the concepts about support, lift and confidence and constructing the mining transaction table at a scale of an evolution but not a time snapshot, this module expanded quantitative Apriori [39], CBAR [40], FP-Tree [42] and MiQarma [41].

Compared with independent popular techniques and algorithms, PoSDMS provided a platform to overcome the challenges of exploring oceanic dynamic information from time series of raster-formatted datasets. As all the key algorithms were developed at a scale of oceanic evolution, the preliminary results from a case study of SST were encouraging, and demonstrated that PoSDMS could be useful for obtaining oceanic dynamics thanks to its powerful processing capacity. While only SSTA in global oceans was taken as a case study to evaluate the functions and performance of PoSDMS, it would be equally capable of dealing with the other five marine environmental parameters, i.e., sea surface salinity, sea surface precipitation, sea level anomaly, sea surface chl_*a* concentration and marine primary productivity. Additionally, PoSDMS is expandable and could be used to deal with other marine environmental parameters, and it has been registered by the National Copyright Administration of P.R. China (No. 2022SR0406836).

The proposed PoSDMS system proved to be a promising analytical tool for dealing with oceanic dynamics using time series of raster-formatted datasets, but further development is still needed. Future studies will aim to expand the interfaces to integrate the latest mining methods and techniques, ensuring that PoSDMS keeps pace with the developments of data mining technologies in big data era. All of the key algorithms were developed in a serial and

Windows-based standalone version, which could limit the popularization and application of PoSDMS. Thus, a second study will aim to encapsulate the modules into independent components, design them in a parallel fashion and then migrate them onto cloud-based service platforms. In addition, PoSDMS discovered large amount of oceanic evolution patterns of marine environmental parameters at the sea surface; however, the dynamic information of the deep sea is equally or perhaps more important than that of the surface. Therefore, PoSDMS must expand its process-oriented analyzing module from three dimensions (Longitude, Latitude, Time) to four dimensions (Longitude, Latitude, Time, Depth).

Author Contributions: Conceptualization, C.X. and L.L.; methodology, L.L., Y.X. and C.W.; software, Y.X. and C.W.; data curation, C.N. and Y.X.; writing—original draft preparation, C.X. and Y.X.; writing—review and editing, C.X. and L.L.; upervision, C.X.; project administration, C.X.; funding acquisition, C.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA19060103), and by the National Natural Science Foundation of China (No. 41671401). And The APC was funded by XDA19060103.

Data Availability Statement: Not applicable.

Acknowledgments: We thank the NASA/JPL/PO.DAAC for providing the research data.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

CBAR	Cluster-Based Association Rule
CPU	Central Processing Unit
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DcSTCA	Dual-constraint SpatioTemporal Clustering Approach
DEM	Digital Elevation Model
ENSO	El Niño–Southern Oscillation
ESRL	Earth System Research Laboratories
GDAL	Geospatial Data Abstraction Library
GeoDMA	Geographic Data Mining Analyst
GIS	Geographic Information System
GRASS	Geographic Resources Analysis Support System
HDF	Hierarchical Data Format
HFA	Hierarchal File Format
MIQarma	Mutual-Information-based Quantitative Association Rule-Mining Algorithm
NetCDF	Network Common Data Form
NOAA	National Oceanic and Atmospheric Administration
OAR	Ocean Area Reconnaissance
PDO	Pacific Decadal Oscillation
PoAIES	Process-oriented Approach to Identify Evolution of SSTA
PoAIR	Process-oriented Approach for Identifying Rainstorm
PoGDB	Process-oriented Graph Database
PO	Process object
POID	Process Object ID
PoSCM	Process-oriented Spatiotemporal Clustering Method
PoSDMS	Process-oriented Spatiotemporal Dynamics Mining System
PoTGM	Process-oriented Two-tier Graph Model
P-V-M	Plateau-Valley-Mountain
RSMapMinig	Image-driven Remote-Sensing Mining System
SOID	Sequence Object ID
SRNN	Shared Reciprocal Nearest Neighborhood
SST	Sea Surface Temperature
SSTA	Sea Surface Temperature Anomalies
ST	SpatioTemporal
WSST	Warmer Sea Surface Temperature

References

1. Yuan, M.; Mark, D.M.; Egenhofer, M.J.; Peuquet, D.J. *Extensions to Geographic Representation: A Research Agenda for Geographic Information Science*; CRC Press: Boca Raton, FL, USA, 2004; pp. 129–156.
2. Shaw, S.-L.; Ye, X. Capturing Spatiotemporal Dynamics in Computational Modeling. In *The Geographic Information Science & Technology Body of Knowledge*, 1st Quarter 2019 ed.; Wilson, J.P., Ed.; UCGIS: Ithaca, NY, USA, 2019. [CrossRef]
3. Yi, J.W.; Du, Y.Y.; Liang, F.Y.; Zhou, C.H.; Wu, D.; Mo, Y. A representation framework for studying spatiotemporal changes and interactions of dynamic geographic phenomena. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1010–1027. [CrossRef]
4. Wang, J.; Duckham, M.; Worboys, M. A framework for models of movement in geographic space. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 970–992. [CrossRef]
5. Long, J.A.; Weibel, R.; Dodge, S.; Laube, P. Moving ahead with computational movement analysis. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 1275–1281. [CrossRef]
6. Dodge, S.; Song, G.; Tomko, M.; Weibel, R. Progress in computational movement analysis—Towards movement data science. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 2395–2400. [CrossRef]
7. Saulquin, B.; Fablet, R.; Mercier, G.; Demarcq, H.; Mangin, A.; Fantond' Andon, O.H. Multiscale Event-Based Mining in Geophysical Time Series: Characterization and Distribution of Significant Time-Scales in the Sea Surface Temperature Anomalies Relatively to ENSO Periods from 1985 to 2009. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 3543–3552. [CrossRef]
8. Xue, C.J.; Song, W.J.; Qin, L.J.; Dong, Q.; Wen, X.Y. A spatiotemporal mining framework for abnormal association patterns in marine environments with a time series of remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *38*, 105–114. [CrossRef]
9. Wang, H.; Du, Y.; Yi, J.; Wang, N.; Liang, F. Mining evolution patterns from complex trajectory structures—A case study of mesoscale eddies in the South China Sea. *Int. J. Geo-Inf.* **2020**, *9*, 441. [CrossRef]
10. Xue, C.J.; Su, F.Z.; He, Y.W. Process: A new view of geographical spatiotemporal dynamic analysis. *Adv. Earth Sci.* **2022**, *37*, 65–79. (In Chinese)
11. Yang, J.; Gong, P.; Fu, R.; Zhang, M.H.; Chen, J.M.; Liang, S.L.; Xu, B.; Shi, J.C.; Dickinson, R. The role of satellite remote sensing in climate change studies. *Nat. Clim. Chang.* **2013**, *3*, 875–883. [CrossRef]
12. Guttler, F.; Lenco, D.; Nin, J.; Teisseire, M.; Poncelet, P. A graph-based approach to detect spatiotemporal dynamics in satellite image time series. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 92–107. [CrossRef]
13. Xue, C.J.; Dong, Q.; Xie, J. Marine spatio-temporal process semantics and its application-taking the El Nino Southern Oscilation process and Chinese rainfall anomaly as an example. *Acta Oceanol. Sin.* **2012**, *31*, 16–24. [CrossRef]
14. Xue, C.J.; Wu, C.; Liu, J.; Su, F. A novel process-oriented graph Storage for dynamic geographic phenomena. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 100. [CrossRef]
15. Mcphaden, M.J.; Zebia, S.E.; Glantz, M.H. ENSO as an integrating concept in earth science. *Science* **2006**, *314*, 1740–1745. [CrossRef] [PubMed]
16. Yuan, M. Why are events important and how to compute them in geospatial research? *J. Spat. Inf. Sci.* **2020**, *21*, 47–61. [CrossRef]
17. Liu, W.B.; Li, X.G.; Rahn, D.A. Storm event representation and analysis based on a directed spatiotemporal graph model. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 948–969. [CrossRef]
18. Yu, M.; Yang, C.; Jin, B. A Framework for Natural Phenomena Movement Tracking–Using 4D Dust Simulation as an Example. *Comput. Geosci.* **2018**, *121*, 53–66. [CrossRef]
19. Li, L.; Xu, Y.; Xue, C.; Fu, Y.; Zhang, Y. A Process-Oriented Approach to Identify Evolutions of Sea Surface Temperature Anomalies with a Time-Series of a Raster Dataset. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 500. [CrossRef]
20. Lo, S.H.; Chen, C.T.; Russo, S.; Huang, W.R.; Shih, M.F. Tracking heatwave extremes from an event perspective. *Weather Clim. Extrem.* **2021**, *34*, 100371. [CrossRef]
21. Wu, B.; Yu, B.; Shu, S.; Wu, Q.; Zhao, Y.; Wu, J. A spatiotemporal structural graph for detecting land cover changes. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 397–425. [CrossRef]
22. Gowtham, A.; Anuj, K.; Vipin, K. Spatio-Temporal Data Mining: A Survey of Problems and Methods. *ACM Comput. Surv.* **2018**, *51*, 83.1–83.41. [CrossRef]
23. Khiali, L.; Ndiath, M.; Alleaume, S.; Lenco, D.; Ose, K.; Teisseire, M. Detection of spatio-temporal evolutions on multi-annual satellite image time series: A clustering-based approach. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *74*, 103–119. [CrossRef]
24. Yu, M.; Bambacus, M.; Cervone, G.; Clarke, K.; Duffy, D.; Huang, Q.; Li, J.; Li, W.; Li, Z.; Liu, Q.; et al. Spatiotemporal event detection: A review. *Int. J. Digit. Earth* **2020**, *13*, 1339–1365. [CrossRef]
25. Wang, H.; Du, Y.; Yi, J.; Sun, Y.; Liang, F. A new method for measuring topological structure similarity between complex trajectories. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 1836–1848. [CrossRef]
26. Liu, J.Y.; Xue, C.J.; Dong, Q.; Wu, C.B.; Xu, Y.F. A Process-Oriented Spatiotemporal Clustering Method for Complex Trajectories of Dynamic Geographic Phenomena. *IEEE Access* **2019**, *7*, 1237–1242. [CrossRef]
27. Yan, J.; Wang, L.; Song, W.; Chen, Y. A time-series classification approach based on change detection for rapid land cover mapping. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 249–262. [CrossRef]
28. Muñoz, C.; Wang, L.P.; Willems, P. Enhanced object-based tracking algorithm for convective rainstorms and cells. *Atmos. Res.* **2018**, *201*, 144–158. [CrossRef]
29. Debusscher, B.; Coillie, F.M.B.V. Object-Based Flood Analysis Using a Graph-Based Representation. *Remote Sens.* **2019**, *11*, 1883. [CrossRef]

30. Zhu, R.; Guilbert, E.; Wong, M.D. Object-oriented tracking of the dynamic behavior of urban heat islands. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 405–424. [CrossRef]
31. Xue, C.J.; Liu, J.Y.; Yang, G.H.; Wu, C.B. A process-oriented method for tracking rainstorms with a time-series of raster datasets. *Appl. Sci.* **2019**, *9*, 2468. [CrossRef]
32. Yi, J.; Du, Y.; Wang, X.; He, Z.; Zhou, C. A clustering analysis of eddies' spatial distribution in the South China Sea. *Ocean. Sci.* **2013**, *9*, 171–182. [CrossRef]
33. Caron, L.P.; Boudreault, M.; Camargo, S.J. On the Variability and Predictability of Eastern Pacific Tropical Cyclone Activity. *J. Clim.* **2015**, *28*, 9678–9696. [CrossRef]
34. Pampuch, L.A.; Drumon, A.; Gimeno, L.; Ambrizzi, T. Anomalous patterns of SST and moisture sources in the south atlantic ocean associated with dry events in southeastern brazil. *Int. J. Climatol.* **2016**, *36*, 4913–4928. [CrossRef]
35. Steinbach, M.; Tan, P.N.; Boriah, S.; Kumar, V.; Klooster, S.; Potter, C. The Application of Clustering to Earth Science Data: Progress and Challenges. In Proceedings of the 2nd NASA Data Mining Workshop, Pasadena, CA, USA, 23–24 May 2006.
36. Birant, D.; Kut, A. ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data Knowl. Eng.* **2007**, *60*, 208–221. [CrossRef]
37. Kawale, J.; Liess, S.; Kumar, A.; Steinbach, M.; Ganguly, A.; Samatova, N.F.; Semazzi, F.H.M.; Snyder, P.K.; Kumar, V. Data Guided Discovery of Dynamic Climate Dipoles. In Proceedings of the NASA Conference on Intelligent Data Understanding, Mountain View, CA, USA, 19–21 October 2011; pp. 30–44.
38. Liu, J.Y.; Xue, C.J.; He, Y.W.; Dong, Q.; Kong, F.P.; Hong, Y.L. Dual-constraint Spatiotemporal Clustering Approach for Exploring Marine Anomaly Patterns using Remote Sensing Products. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3963–3976. [CrossRef]
39. Srikant, R.; Agrawal, R. Mining quantitative association rules in large relational tables. In Proceedings of the ACM SIGMOD Conference on Management of Data, Montreal, QC, Canada, 4–6 June 1996; Volume 25, pp. 1–12. [CrossRef]
40. Tsay, Y.J.; Chiang, J.Y. CBAR: An efficient method for mining association rules. *Knowl.-Based Syst.* **2005**, *18*, 99–105. [CrossRef]
41. Xue, C.J.; Song, W.J.; Qin, L.J.; Dong, Q. A mutual-information-based mining method for marine abnormal association rules. *Comput. Geosci.* **2015**, *76*, 121–129. [CrossRef]
42. Wu, X.D.; Kumar, V.; Ross, Q.L.J.; Ghosh, J.; Yang, Q.; Motoda, H.; Mclachlan, G.J. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [CrossRef]
43. Lee, J.W.; Lee, Y.J. A knowledge discovery framework for spatiotemporal data mining. *Int. J. Inf. Process. Syst.* **2006**, *2*, 124–129. [CrossRef]
44. Compieta, P.; Martino, S.D.; Bertolotto, M.; Ferrucci, F.; Kechadi, T. Exploratory spatio-temporal data mining and visualization. *J. Vis. Lang. Comput.* **2007**, *18*, 255–279. [CrossRef]
45. Bertolotto, M.; Di Martino, S.; Ferrucci, F.; Kechadi, T. Towards a framework for mining and analysing spatio-temporal datasets. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 895–906. [CrossRef]
46. Yoo, J.S.; Bow, M. Mining spatial colocation patterns: A different framework. *Data Min. Knowl. Discov.* **2012**, *24*, 159–194. [CrossRef]
47. He, Z.; Deng, M.; Cai, J.; Xie, Z.; Guan, Q.; Yang, C. Mining spatiotemporal association patterns from complex geographic phenomena. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 1162–1187. [CrossRef]
48. Zhang, F.; Mao, R.; Du, Z.; Liu, R. Spatial and Temporal Processes Visualization for Marine Environmental Data Using Particle System. *Comput. Geosci.* **2019**, *127*, 53–64. [CrossRef]
49. Wang, Y.; Li, F.; Zhang, B.; Li, X. Development of a component-based interactive visualization system for the analysis of ocean data. *Big Earth Data* **2021**, 1–17. [CrossRef]
50. Korting, T.S.; Fonseca, L.M.G.; Camara, G. GeoDMA-Geographic Data Mining Analyst. *Comput. Geosci.* **2013**, *57*, 133–145. [CrossRef]
51. Pieri, G.; Cocco, M.; Salvetti, O. A marine information system for environmental monitoring: ARGO-MIS. *J. Mar. Sci. Eng.* **2018**, *6*, 15. [CrossRef]
52. Romani, L.A.S.; de Avila, A.M.H.; Chino, D.Y.T.; Zullo, J.; Chbeir, R.; Traina, C.; Traina, A.J.M. A New Time Series Mining Approach Applied to Multitemporal Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 140–150. [CrossRef]
53. Xue, C.J.; Dong, Q.; Li, X.; Fan, X.; Li, Y.; Wu, S. A remote-sensing-driven system for mining marine spatiotemporal association patterns. *Remote Sens.* **2015**, *7*, 9149–9165. [CrossRef]
54. Song, W.J.; Dong, Q.; XUE, C.J. A classified El Niño index using AVHRR remote-sensing SST data. *Int. J. Remote Sens.* **2016**, *37*, 403–417. [CrossRef]
55. Zhao, S.; Jin, F.-F.; Stuecker, M.F. Understanding lead times of warm water volumes to ENSO sea surface temperature anomalies. *Geophys. Res. Lett.* **2021**, *48*, e2021GL094366. [CrossRef]
56. Robinson, I.; Webber, J.; Eifrem, E. *Graph Database*, 2nd ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2015.
57. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the International Conference on Knowledge Discovery & Data Mining, Portland, OR, USA, 2–4 April 1996.

58. Reynolds, R.W.; Rayner, N.A.; Smith, T.M.; Stokes, D.C.; Wang, W. An Improved In Situ and Satellite SST Analysis for Climate. *J. Clim.* **2002**, *15*, 1609–1625. [CrossRef]
59. Zhang, P.; Steinbach, M.; Kumar, V.; Shekhar, S.; Tan, P.-N.; Klooster, S.; Potter, C. Discovery of patterns of Earth science data using data mining. In *Next Generation of Data Mining Applications*; Zurada, J., Kantardzic, M., Eds.; IEEE Press: Piscataway, NJ, USA, 2005; pp. 167–188.