Open Geospatial Engine: A Cloud-based Spatiotemporal Computing Platform

Xianyuan Zhang [1], Longgang Xiang [1], Peng Yue [2,3], Jianya Gong [2], Huayi Wu [1]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China - (zhangxianyuan, geoxlg, wuhuayi)@whu.edu.cn
[2] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China - gongjy@whu.edu.cn
[3] Hubei Province Engineering Center for Intelligent Geoprocessing, Wuhan University, Wuhan, China - pyue@whu.edu.cn

Keywords: Earth Spatiotemporal Big Data, Open Geospatial Engine, GeoCube, Cloud Computing, Multi-dimensional Joint Analysis.

Abstract

The quantity and diversity of Earth spatiotemporal big data have significantly increased in recent years, providing the potential to comprehensively analyse complex spatiotemporal problems from new perspectives. However, integrating, managing, and applying multi-source heterogeneous Earth spatiotemporal big data remains a challenge. To address this issue, this study proposes a cloud-based spatiotemporal computing platform, the Open Geospatial Engine (OGE), for the unified organization and joint analysis of Earth spatiotemporal big data in multiple dimensions and scales. The framework of this platform comprises three modules: data management, computing engine, and service interface. The data management module adopts the GeoCube model to effectively integrate and manage multi-source heterogeneous spatiotemporal data through multi-dimensional aligned tiles. The computing engine module seamlessly maps the GeoCube model to the cloud environment by extending Spark RDD, making the GeoCube model capable of high-performance distributed computing. Following OGC specifications, the service interface module integrates and shares data, operators, and spatiotemporal analysis models through extensible APIs in an interactive development environment. Combined with a series of high-performance optimization techniques, OGE simplifies data queries and the construction of complex analytical applications by integrating these three modules. The applicability of OGE is demonstrated by case studies involving multi-dimensional queries and joint analysis of long-time-series and heterogeneous spatiotemporal data.

1. Introduction

With the continuous advancement of Earth observation technologies, the Earth spatiotemporal big data are being continuously produced through orbital sensors, field measurements, and computer simulations (Li et al., 2017), leading to a significant increase in data volume, accompanied by an abundance in the diversity of available data types (Zhu et al., 2021). These data play an important role in various fields including resource utilization, economic development, national security, and social governance (Zhang, 2018). Despite diverse sources and structures of Earth spatiotemporal big data, they reflect the surface of the Earth from different aspects in various granularities, time phases, directions, and levels, collectively revealing the mechanisms underlying Earth's evolution, urban operational patterns, and human activity modes (Xu et al., 2016). In the era of Earth spatiotemporal big data, the integration of technologies such as spatiotemporal big data management, cloud-based distributed computing, artificial intelligence, and other related technologies holds unprecedented potential for understanding and discerning the Earth's dynamics. Consequently, the establishment of a complete processing framework of Earth spatiotemporal big data, encompassing data storage, management, computing, and application services, for comprehensively and efficiently mining and analysing spatiotemporal information, has become a highly focused research area globally (Liao, 2021).

Efficient data access is important for spatiotemporal analysis and application. Achieving technological breakthroughs in unified management of multi-source data combined with cloud computing technology is essential to meet real-time and comprehensive analysis requirements (Wang et al., 2019; Xu et al., 2021). Given the challenges and opportunities presented by the wide-range, multi-scale, multi-type, long-time-series, and high-dimensional characteristics of Earth spatiotemporal big data, a series of research programs have been launched internationally (Baumann et al., 2018; Mahecha et al., 2020). Analysis-ready data (ARD) and spatiotemporal data cubes for Earth Observation have gained widespread adoption (Gao et al., 2022; Xu et al., 2022a). In alignment with this trend, the Committee on Earth Observation Satellites (CEOS) proposed the Open Data Cube project (ODC, 2018), aiming to provide open-source solutions for the integrated management of Earth observation big data. Based on the ODC, a series of spatiotemporal data cube infrastructures have been developed, including the Australian Geoscience Data Cube (Lewis et al., 2017) and the Swiss Data Cube (Chatenoux et al., 2021; Giuliani et al., 2017), which are deployed on supercomputing platforms for unified management and access of remote sensing, meteorological, and ground station data. Additionally, The European Space Agency has established the Earth System Data Cube for joint analysis of multi-source data streams in the ocean, atmosphere, and other fields (Mahecha et al., 2020). Particularly noteworthy is the Google Earth Engine (GEE) initiated by Google in collaboration with Carnegie Mellon University and the United States Geological Survey (Gorelick et al., 2017), which has evolved into a cloud computing platform highly relied upon by geoscience-related scientists and engineers worldwide.

Although the platforms mentioned above have addressed multiple key issues in utilizing Earth spatiotemporal big data, there are still some shortcomings. On one hand, the data stored and managed primarily consist of corrected remote sensing images from various sources, with excessive emphasis placed on the multi-resolution pyramid structure during the process of standardization, while the unified expression and organization of types of Earth spatiotemporal data such as vector, raster and thematic data in the same area are ignored (Zhu et al., 2021), which leads to isolated data island and limits the descriptive capabilities of the Earth. On the other hand, the operational interfaces provided are still confined to simple overlay analysis and visualization based on scene-organized layers, lacking scalability (Xu et al, 2022b). The granularity of operations is relatively fixed, and there is a lack of systematic operations oriented to multi-level, heterogeneous, and multi-dimensional spatiotemporal features, making it challenging to fulfil the requirements of multi-dimensional joint analysis.

In this article, we propose a spatiotemporal computing platform, the Open Geospatial Engine (OGE), for unified organization and efficient joint analysis of Earth spatiotemporal big data. First, we provide an overview of the OGE framework, which integrates the application capabilities of Earth spatiotemporal big data. Subsequently, the multidimensional spatiotemporal data cube model adopted in OGE, GeoCube (Yue et al., 2020b), is described in detail, serving as a pivotal component for the multi-dimensional integration of heterogeneous spatiotemporal data. Based on the cube model, we introduce a mapping method for cube objects, facilitating efficient distributed spatiotemporal computing in the cloud environment. Finally, a prototype platform was built to verify the feasibility of the proposed framework. This study contributes to the literature in four main ways.

1) From the perspective of integrating Earth spatiotemporal big data, Geocube is adopted to solve the problem of organizing multi-type and multi-scale spatiotemporal data, providing capabilities of unified modelling and multi-dimensional operations for raster, field, vector, and thematic Earth spatiotemporal big data.

2) To give the GeoCube model the capability of high-performance distributed spatiotemporal computing, seamless mapping of various spatiotemporal data organized by GeoCube onto the distributed cloud environment is achieved.

3) A series of optimization schemes for spatiotemporal computing are integrated, including the multivariate hybrid organization and storage load balancing on the storage side, as well as the application of adaptive resource scheduling and the introduction of the LuoJiaNet and LuoJiaSet deep learning framework at the computing side. The efficiency of multi-dimensional reasoning and computing for complex spatiotemporal problems is enhanced.

4) A prototype platform is implemented to verify the feasibility of the proposed OGE framework by taking full advantage of open-source technologies and theories of big data and cloud computing. Feasible solutions are proposed for customized spatiotemporal analysis across multiple dimensions.

The remainder of this article is organized as follows. The overview of the OGE Framework is introduced in Section 2. The multi-dimensional data model of GeoCube is described in Section 3. The distributed spatiotemporal computation method designed for GeoCube is introduced in Section 4. The prototype platform and implementations are provided in Section 5. Finally, Section 6 concludes the article.

2. Open Geospatial Engine Framework

Massive spatiotemporal data and complex modelling algorithms play a significant role in spatiotemporal analysis (Deng et al., 2023), while service publishing enables their accessibility and operability. Thus, the data management, computing engine, and service interface are the three core modules within the OGE framework. As shown in Figure 1, we fully draw on innovative technologies emerging from fields such as the Internet of Things, big data, and artificial intelligence, combined with advanced models and methods in the field of Earth observation, propose OGE to provide a comprehensive platform for the processing and analysis of Earth spatiotemporal big data at various dimensions and granularities, with deep coupling and open sharing of computing power, data, operators, and spatiotemporal analysis models.

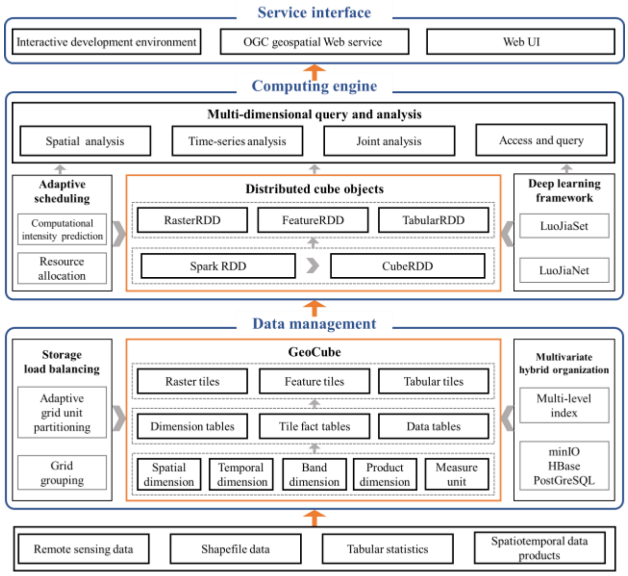

Figure 1. Overview of the OGE framework.

2.1 Data Management

The module of data management stores and manages a variety of Earth spatiotemporal big data from multiple sources, including remote sensing data, shapefile data, tabular statistics, and other spatiotemporal data products. Considering the advantages of the data cube model in data organization and analysis, the GeoCube model is adopted within OGE, aiming to establish a unified multi-dimensional framework capable of mapping heterogeneous, multi-source Earth spatiotemporal big data to standardized grid units, while aligning their attributes across dimensions through standard tile structures, which facilitates unified expression, organization and analysis of various types of spatiotemporal data. The detailed structure of GeoCube within the OGE framework is illustrated in Section 3.

Taking full advantage of distributed file and database storage, as well as relational and non-relational storage, a hybrid organization scheme based on the distributed file system minIO, the non-relational database HBase, and the relational-object database PostgreSQL is designed in OGE. In this scheme, the metadata of Geocube is stored in the relational database, the tile data is stored in the non-relational database, and the remote sensing images are stored in the distributed file system. A unified coding method for spatiotemporal cube units based on T-S-T (time-space-time) is proposed to link this multivariate hybrid storage system, which performs coarse-grained time coding, spatial coding, and fine-grained time coding at each coding level to improve the efficiency of spatiotemporal indexing. For query processing, a spatiotemporal range query optimization algorithm based on multi-level coarse filtering is

https://doi.org/10.5194/isprs-annals-X-4-2024-453-2024 |

proposed, which allows the coarse filtering step to be executed at the storage side to reduce network communication.

To address the issue of storage load imbalance arising from spatial clustering and temporal tides, a hierarchical computing method based on spatial density distribution derived from sampling results is proposed, executing split and merge operations in parallel at intermediate levels to adaptively partition spatial grids. The multi-level spatial grids are then represented by an undirected weighted graph structure and are grouped through graph partitioning algorithms. Finally, genetic algorithms are introduced to design an iterative refinement method for boundary units. By adjusting the grid ownership determined through graph partitioning, the imbalance of spatiotemporal data across different time granularity levels is further reduced.

2.2 Computing Engine

The computing engine module extends Apache Spark's resilient distributed dataset (RDD) based on the multi-dimensional partitioning structure of the GeoCube model in OGE (Yue et al., 2018), and the CubeRDD is designed for distributed spatiotemporal data cube objects. The CubeRDD can be further transformed into specific RDDs depending on data types, enabling the GeoCube model for multi-source data integration to be employed in high-performance distributed spatiotemporal computing environments. The CubeRDD method will be elaborated in detail in Section 4. Through the combination of CubeRDD and high-performance computing optimization methods, the computing engine module provides efficient multi-dimensional query and analysis capabilities.

The irregular distribution and pronounced heterogeneity of Earth spatiotemporal big data, particularly vector data, can lead to load imbalance issues during distributed computing. To address this challenge, machine learning methods are employed to characterize the spatiotemporal complexity features of data and algorithms (Yue et al., 2020a), enabling the automated prediction of the computational intensity of each computing node in processing workflows. During execution in OGE, resource allocation is adapted by simulated intensity results to improve the computational performance.

In the intelligent interpretation of multispectral remote sensing images based on multi-dimensional spatiotemporal information, OGE introduces the LuoJiaNet deep learning framework and the large-scale open-source remote sensing sample library, LuoJiaSet (Zhang et al., 2023), which provides basic remote sensing application models including remote sensing scene classification, target detection, ground object classification, change detection, and multi-view 3D reconstruction. Multi-dimensional applications compatible with the characteristics of remote sensing images have been designed, and the classification performance of the deep network combined with the GeoCube is improved under the guidance of the empirical knowledge model.

2.3 Service Interface

Based on the above key technologies, the service interface module of OGE provides multiple types of RESTFUL services of metadata, data, operators, and models through microservices. Following the OGC standard, it provides an interactive development environment with various spatiotemporal APIs.

On this basis, the Web UI of OGE is designed and implemented by mainstream Web frameworks to integrate and share data resources, operators, and spatiotemporal analysis capabilities. The interactive development environment supports users to form workflows by dragging graphical modules of operators or flexibly designing analytical applications based on Python and JavaScript in a low-code way. All these workflows are executed by the backend distributed engine through directed acyclic graphs (DAG) constructed by OGE. Additionally, with the assistance of service proxies, the method of GRASS and WPS is seamlessly imported, thereby expanding the spatiotemporal operator system of OGE.

3. Multi-dimensional Data Model

Earth spatiotemporal big data possess characteristics such as multi-type, heterogeneity, complex structure, and large volume. Due to different production modes and specifications, their resolutions and coordinate systems are usually inconsistent. Existing management frameworks for spatiotemporal big data are typically optimized for certain types of data, which leads to insufficient ability for joint analysis of multi-source heterogeneous spatiotemporal data. To address this issue, the GeoCube model (Figure 2) based on the fact constellation schema of online analytical processing (OLAP) is adopted in OGE. Through the designed multi-dimensional structure, as well as formal expression and organization method for its measure, GeoCube integrate various types of Earth spatiotemporal big data, including vector data, raster data, and tabular data, by mapping them to regular grid units. Thus, GeoCube provides unified representation, organization, and OLAP capabilities.

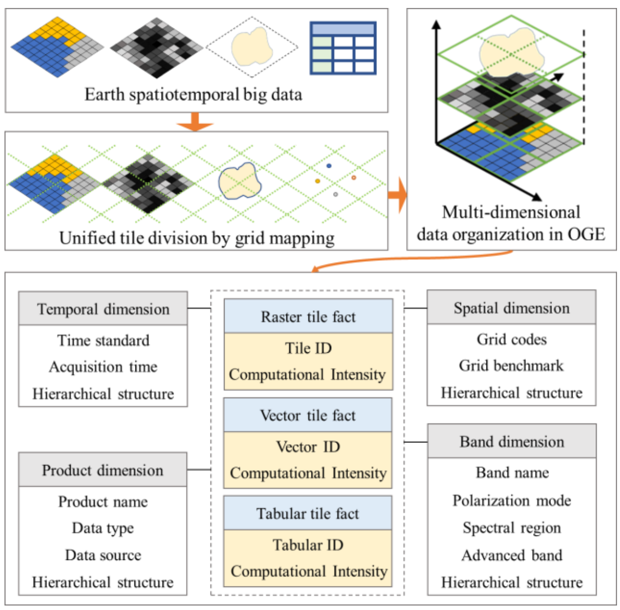

Figure 2. The GeoCube model in OGE.

3.1 Multiple Dimensions

To meet the needs of multi-dimensional spatiotemporal data analysis, four dimensions of GeoCube are involved: temporal dimension, spatial dimension, product dimension, and band dimension, which makes Earth spatiotemporal big data aligned in each dimension, so as to provide the ability of multi-dimensional real-time analysis.

The spatial and temporal dimensions are independent of the data imported, and provide a unified spatiotemporal reference as well as standard spatiotemporal partitioning. The spatial dimension describes the spatial locations of the cube, and its members are composed of a series of regular grids, which are uniquely identified by the grid codes and the spatial benchmark. Space-filling curves are used to code these discrete grids and reduce two-dimensional space to one dimension. This coding method assigns close codes to spatially adjacent data, thus making it beneficial for improving query efficiency and reducing communication frequency between computing nodes during distributed computing. The spatial benchmark describes the size, resolution, and reference of the spatial grid, defining the spatial scale of the cube. When the raster data and the vector data are uniformly mapped to a two-dimensional grid, the raster data are divided into physical blocks of the same size as the grid, while vector data is indexed into logical blocks, which helps to maintain the integrity of vector features. Additionally, the spatial dimension contains hierarchical information such as grid level, district level, and city level, which facilitates OLAP operation along the spatial dimension. The temporal dimension describes the time standard used, the duration of the spatiotemporal process under study, and the temporal distribution of spatiotemporal data. For long time-series data products, the acquisition time can either be the time obtained after the product is processed or the duration of the product's evolution process. There are intuitive hierarchical and adjacency relationships at the temporal dimension, such as the hierarchy from days to months and years, as well as the temporal similarity between adjacent time members, for example, the atmospheric temperature measured over consecutive days, which enables OLAP operation for sequential variation and spatiotemporal interpolation along the temporal dimension.

The band dimension describes the spectral band information of remote sensing images, including band names, polarization modes, spectral ranges, and advanced bands. The polarization mode is a unique attribute designed for synthetic aperture radar (SAR) images. The spectral range retains spectral range information of different data sources, allowing a band name to be mapped to different spectral range intervals. Considering that the results of raster calculation among bands can still be used for band analysis, the advanced band is proposed to describe hierarchical attribute information specifically designed for OLAP operation, referring to product-level bands generated based on a series of original bands, such as normalized difference water index (NDWI) bands and normalized difference vegetation index (NDVI) bands. The hierarchical relationship between advanced bands and original bands is also marked.

The product dimension is a thematic dimension for spatiotemporal data products, which contains information such as product name, data type, and data source. The product name is the most basic information, consisting of feature descriptions. The data type is used to distinguish between raster, vector, and tabular data products, and data source information describes production means including the sensors, satellite platforms and data publisher. It should be noted that vector and tabular data do not have satellite platform information.

3.2 Unified Tile Fact

The members of the temporal dimension, spatial dimension, product dimension, and band dimension jointly point to the measure information in the multi-dimensional structure of GeoCube. Raster data, vector data and tabular statistics with spatiotemporal information are uniformly mapped onto grids of the same benchmark as regular measure units with a uniform size. To manage measure units effectively, GeoCube further aggregates adjacent measure units in space and time into tiles, forming the tile fact structured as "dimension-measure". Three types of tile fact are supported in GeoCube: raster tile fact, vector tile fact, and tabular tile fact, and high-dimensional field data can be expressed by raster tile fact through dimension expansion. Each type of tile fact contains a spatial measure and a computational intensity measure. The spatial measure links to actual raster, vector, and tabular tile data, while the computational intensity measure records the computational intensity of the tile in different analytical functions. By identifying the computational intensity of different units in this multi-dimensional structure, the spatial-temporal heterogeneity is revealed, which supports efficient parallel computing.

4. Distributed Spatiotemporal Computation

Organizing data in a high-performance form is crucial for efficient distributed spatiotemporal computation of Earth spatiotemporal big data. The tile organization of the GeoCube model provides convenience for designing distributed cube objects, which enables efficient data processing by combining distributed objects with large-scale cloud computing.

As a mainstream distributed computing framework, Apache Spark is an efficient computing engine designed for large-scale data processing which employs RDD as the data container (Zaharia et al., 2010). The CubeRDD adopted in OGE extends Spark RDD and proposes a set of distributed cube objects. On the one hand, it inherits the distributed computing ability of Spark RDD, and on the other hand, it is compatible with multiple source data types by the tile fact of GeoCube, and achieves seamless mapping between the GeoCube model and cloud computing environment, as shown in Figure 3.

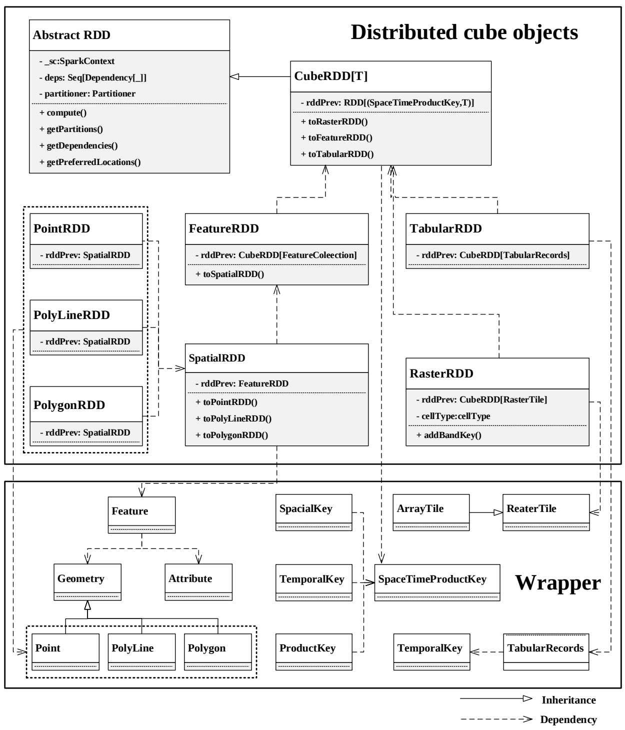

Figure 3. Distributed cub objects.

The value of CubeRDD is generic type T, allowing it to be converted into any data type to meet the mapping requirements of different data types. Furthermore, CubeRDD can be transformed into distributed cube raster objects (RasterRDD), distributed cube vector objects (FeatureRDD), and distributed cube tabular objects (TabularRDD), keeping consistency with the mapping between tiles in Geocube and cloud environments. The implemented distributed cube objects offer two advantages: 1) Consistency with the proposed GeoCube conceptual model, supporting analysis of multi-source, heterogeneous data within a unified multi-dimensional framework. For instance, RasterRDD can be grouped along the temporal dimension to support distributed time series analysis or algebraic analysis of band maps, while FeatureRDD can be grouped along the product dimension to support distributed nine-intersection model operations. Additionally, merging RasterRDD and FeatureRDD based on dimension keys enables raster tiles and vector tiles with identical spatiotemporal keys to fall within the same spatiotemporal cube unit, facilitating distributed joint analysis. And 2) In addition to raster, vector, and tabular data types, it can be extended to support other data types of cube objects by simply customizing dimension keys and measure types.

Pixels of remote sensing images are typically considered as the measure in spatiotemporal data cube platforms, which also serve as the smallest data unit for processing and analysis in GEE. However, the GeoCube model adopts raster tiles, vector tiles, or tabular tiles from multiple sources as the cube measure. To be consistent with the structure of the cube model, the proposed distributed spatiotemporal data cube objects are also organized by tiles. Therefore, OGE implements a large number of high-level distributed cube object operation interfaces, allowing users to perform flexible analysis without an in-depth understanding of the conceptual model. While the underlying conceptual model consists of cube units represented by raster or vector tiles, defining high-level interfaces enables users to directly engage in multi-dimensional analysis based on raster cells or vector features.

5. Prototype System and Implementation

Currently, the OGE platform has integrated a total of hundreds of millions of global vector data, terrain data, and image data, as well as remote sensing products, remote sensing samples, satellite virtual constellation, Internet of Things data and other Earth observation data, totalling approximately 20TB. OGE provides over a hundred distributed spatiotemporal operators covering raster, vector, and thematic data. Based on the above data and operators, the feasibility of OGE is verified by multi-dimensional query and analysis experiments.

5.1 Multi-dimensional Query

Through the construction and data importation of GeoCube, OGE aligns all kinds of spatiotemporal data across dimensions, enabling efficient and convenient multi-dimensional semantic query of raster and vector data. Figure 4 shows slicing the temporal dimension of the cube constructed based on Landsat 8 images through the drop-down list or inputting member information, without writing codes, the advanced member NDVI of a specific time on the band dimension in the study area can be obtained and displayed on the bottom layer.

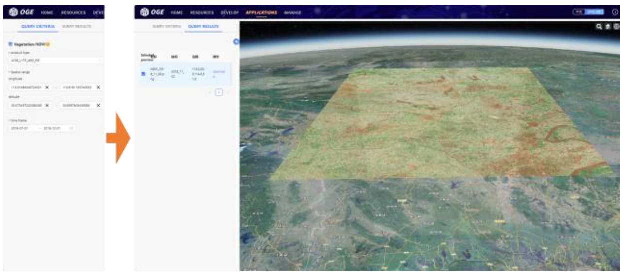

Figure 4. Multi-dimensional query.

5.2 Multi-dimensional Analysis

In terms of long-time-series spatiotemporal computations, take the analysis and computation of the water index as an example. The Landsat 8 L1-level data within the study area are imported to construct GeoCube. By utilizing the existing *Cube.NDWI* API, the advanced member NDWI is derived through roll-up along the band dimension, resulting in an updated cube named "NDWICube". Subsequently, the binarization process of NDWI is carried out and the product member "NDWI_Product" is added to the product dimension. Finally, the results are visualized. Figure 5 illustrates the codes required for this analysis, the key steps in the analysis process and execution time for different amounts of data.

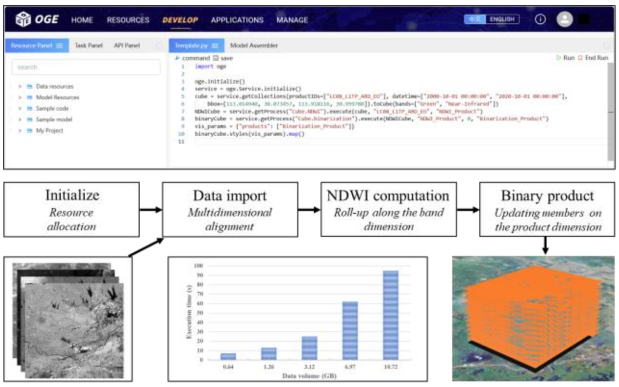

Figure 5. Long-time-series NDWI computation.

Figure 6 illustrates a multi-dimensional analysis of multi-source data, focusing on flood-affected areas and villages in a province in southern China in 2016. The data involved in this case included Landsat 7 Level-1 products before and after the disaster, village building vector data, and village-level demographic statistics. After initializing the cube and importing the above data, the member of NDWI on the band dimension is generated by the roll-up of high-resolution remote sensing images before and after the disaster. Subsequently, water body change detection is performed, resulting in flood-affected area products. Overlay analysis is then conducted between the flood-affected area products, village building vector data and tabular demographic statistics to extract villages and associated populations affected by the disaster.

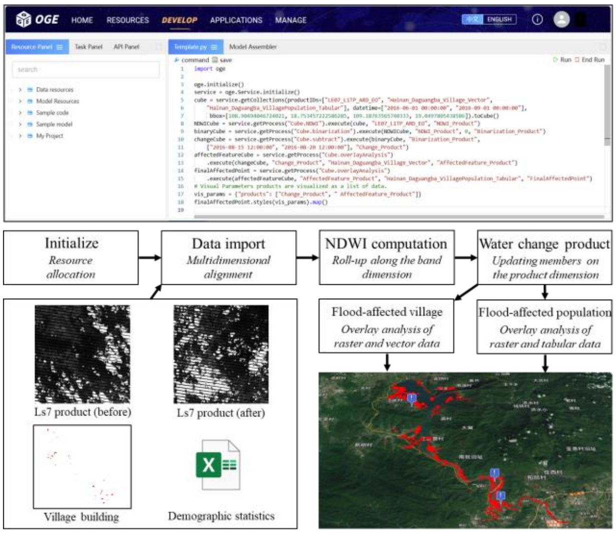

Figure 6. Multi-dimensional analysis of multi-source data.

The codes as well as the analysis process for above multi-dimensional joint analysis are shown in the upper part of figure 6, and the figure also illustrates the results provided by the web side of OGE, where the red area is the flood-affected area, and the points marked with symbols are affected villages.

Obviously, for multi-dimensional analysis such as long-time-series spatiotemporal computations and vector-raster joint analysis, users will fall into tedious code-writing tasks when using GEE and other platforms. However, with the development interface provided by OGE, users can achieve the same functions with only a few lines of code. The reason is that the GeoCube is an Earth observation data model for multi-source fusion and spatiotemporal alignment in long-time-series. The OGE not only provides standardized data representation and organizational structure with open access, but also encapsulates the operators, models and workflows required for spatiotemporal analysis, helping to shorten the development cycle of spatiotemporal applications.

6. Conclusion

In this study, we propose a spatiotemporal computing platform, the OGE, based on cloud computing and big data technologies, for unified organization and efficient joint analysis of multi-source and heterogeneous Earth spatiotemporal big data in multiple dimensions. Compared with previous works and similar platforms, the proposed OGE supports the multi-dimensional unified expression, organization, and analysis of multiple types of data such as vector, raster, and tabular data. The implementation and cases demonstrate the applicability of this framework

Future work includes the optimization of the distributed computing engine, the enrichment of spatiotemporal operators, and the expansion of research fields to improve the ability of data management and spatiotemporal analysis.

References

Baumann, P., Rossi, A. P., Bell, B., Clements, O., Evans, B., Hoenig, H., Hogan, P., Kakaletris, G., Koltsida, P., Mantovani, S., Marco Figuera, R., Merticariu, V., Misev, D., Pham, H. B., Siemen, S., Wagemann, J. 2018. Fostering Cross-Disciplinary Earth Science Through Datacube Analytics. *Earth Observation Open Science and Innovation*, 91-119. doi.org/10.1007/978-3-319-65633-5_5.

Chatenoux, B., Richard, J., Small, D., Roeoesli, C., Wingate, V., Poussin, C., Rodila, D., Peduzzi, P., Steinmeier, C., Ginzler, C., Psomas, A., Schaepman, M., Giuliani, G. 2021. The Swiss data cube, analysis ready data archive using earth observations of Switzerland. *Scientific Data*, 8(1), 295. doi.org/10.1038/s41597-021-01076-6.

Deng, H., Yue, P., Yu, D., Cao, Z., Liu, R., Wang, K. 2023. A Geospatial Data and Model Hub for Online Geospatial Analysis. 2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 1-6. doi.org/10.1109/Agro-Geoinformatics59224.2023.10233547.

Gao, F., Yue, P., Cao, Z., Zhao, S., Shangguan, B., Jiang, L., Hu, L., Fang, Z., Liang, Z. 2022. A multi-source spatio-temporal data cube for large-scale geospatial analysis. *International Journal of Geographical Information Science*, 36(9), 1853-1884. doi.org/10.1080/13658816.2022.2087222.

Giuliani, G., Chatenoux, B., De Bono, A., Rodila, D., Richard, J., Allenbach, K., Dao, H., Peduzzi, P. 2017. Building an Earth Observations Data Cube: lessons learned from the Swiss Data Cube (SDC) on generating Analysis Ready Data (ARD). *Big Earth Data*, 1(1-2), 100-117. doi.org/10.1080/20964471.2017.1398903.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27. doi.org/10.1016/j.rse.2017.06.031.

Lewis, A., Oliver, S., Lymburner, L., Evans, B., Wyborn, L., Mueller, N., Raevksi, G., Hooke, J., Woodcock, R., Sixsmith, J., Wu, W., Tan, P., Li, F., Killough, B., Minchin, S., Roberts, D., Ayers, D., Bala, B., Dwyer, J., Dekker, A., Wang, L. 2017. The Australian Geoscience Data Cube — Foundations and lessons learned. *Remote Sensing of Environment*, 202, 276-292. doi.org/10.1016/j.rse.2017.03.015.

Li, D., Wang, M., Dong, Z., Shen, X., Dong, Z. 2017. From Earth Observation Satellite to Earth Observation Brain. *Geomatics and Information Science of Wuhan University*, 42(2), 134-149. doi.org/10.13203/j.whugis20160526.

Liao, X. 2021. Scientific and technological progress and development prospect of the earth observation in China in the past 20 years. *National Remote Sensing Bulletin*, 25(1), 267-275. doi.org/10.11834/jrs.20211017.

Mahecha, M., Gans, F., Brandt, G., Christiansen, R., Cornell, S., Fomferra, N., Kraemer, G., Peters, J., Bodesheim, P., Camps-Valls, G., Donges, J., Dorigo, W., Estupinan-Suarez, L., Gutierrez-Velez, V., Gutwin, M., Jung, M., Londoño, M., Miralles, D., Papastefanou, P., Reichstein, M. 2020. Earth system data cubes unravel global multivariate dynamics. *Earth Syst. Dynam*, 11(1), 201-234. doi.org/10.5194/esd-11-201-2020.

ODC. 2018. CEOS OpenDataCube. https://www.opendatacube.org (11 March 2024).

Wang, S., Zhong, Y., Wang, E. 2019. An integrated GIS platform architecture for spatiotemporal big data. *Future Generation Computer Systems*, 94, 160-172. doi.org/10.1016/j.future.2018.10.034.

Xu, C., Du, X., Fan, X., Giuliani, G., Hu, Z., Wang, W., Liu, J., Wang, T., Yan, Z., Zhu, J., Jiang, T., Guo, H. 2022a. Cloud-based storage and computing for remote sensing big data: a technical review. *International Journal of Digital Earth*, 15(1), 1417-1445. doi.org/10.1080/17538947.2022.2115567.

Xu, C., Du, X., Fan, X., Yan, Z., Kang, X., Zhu, J., Hu, Z. 2021. A Modular Remote Sensing Big Data Framework. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-11. doi.org/10.1109/TGRS.2021.3100601.

Xu, C., Du, X., Jian, H., Dong, Y., Qin, W., Mu, H., Yan, Z., Zhu, J., Fan, X. 2022b. Analyzing large-scale Data Cubes with user-defined algorithms: A cloud-native approach. *International Journal of Applied Earth Observation and Geoinformation*, 109, 102784. doi.org/10.1016/j.jag.2022.102784.

Xu, G., Liu, Q., Chen, L., Liu, L. 2016. Remote sensing for China's sustainable development: Opportunities and challenges. *National Remote Sensing Bulletin*, 20(5), 679-688. doi.org/10.11834/jrs.20166308.

Yue, P., Gao, F., Shangguan, B., Yan, Z. 2020a. A machine learning approach for predicting computational intensity and domain decomposition in parallel geoprocessing. *International Journal of Geographical Information Science*, 34(11), 2243-2274. doi.org/10.1080/13658816.2020.1730850.

Yue, P., Shangguan, B., Zhang, M., Gao, F., Cao, Z., Jiang, L., Fang, Z. 2020b. Geocube: Towards the Multi-Source Geospatial Data Cube in Big Data Era. IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium. doi.org/10.1109/IGARSS39084.2020.9324632.

Yue, P., Wu, Z., Shangguan, B. 2018. Design and Implementation of a Distributed Geospatial Data Storage Structure Based on Spark. Geomatics and Information Science of Wuhan University, 43(12), 2295-2302. doi.org/10.13203/j.whugis20180034.

Zaharia. M., Chowdhury, M., Franklin, M.J, Shenke, S., Stoica, I. 2010. Spark: cluster computing with working sets// Proceedings of the 2nd USENIX conference on Hot topics in cloud computing, Boston, MA. doi.org/10.5555/1863103.1863113.

Zhang, B. 2018. Remotely Sensed Big Data Era and Intelligent Information Extraction. Geomatics and Information Science of Wuhan University, 43(12), 1861-1871. doi.org/10.13203/j.whugis20180172.

Zhang, Z., Zhang, M., Gong, J., Hu, X., Xiong, H., Zhou, H., Cao, Z. 2023. LuoJiaAI: A cloud-based artificial intelligence platform for remote sensing image interpretation. Geo-spatial Information Science, 26(2), 218-241. doi.org/10.1080/10095020.2022.2162980

Zhu, L., Su, X., Tai, X. 2021. A High-Dimensional Indexing Model for Multi-Source Remote Sensing Big Data. Remote Sensing,13(7), 1314. doi.org/10.3390/rs13071314.