RESEARCH ON EFFICIENT INDEXING OF LARGE-SCALE GEOSPATIAL DATA BASED ON MULTI-LEVEL GEOGRAPHIC GRID

YIN GAO [1], HAIRUI DUO [2] *, JIAN CHE [1], SHIQUAN ZHAO [1], BIANLI ZHAO [1]

[1] National Geomatics Center of China, 100830, Beijing, China
[2] Qinghai Normal University, 810016, Xining, China

KEY WORDS: Geographic Grid; Massive Geospatial Data; Spatial Index; Multi-level Framework; Spatial Scale; Dimension Reduction.

ABSTRACT:

With the implementation of unified natural resource management in China, national geographic conditions monitoring data have been identified as fundamental data for natural resource survey and monitoring. The efficiency of information extraction from massive spatio-temporal data to support natural resource management has emerged as a critical indicator for maximizing the value of geographic conditions monitoring data and enhancing data-driven decision management. Traditional spatial indices are computationally intensive, and when confronted with immense data volume or uneven data scale, issues such as extensive index computations and poor scale adaptability arise, impeding the efficient retrieval of complex geospatial data. In response to the need for efficient indexing of massive geospatial monitoring data at a scale of 100 million, a multi-level geographic spatial index framework based on geographic grids is proposed. Within the geographic conditions spatio-temporal database, a three-level spatial index of "zone-grid-space" is constructed, utilizing massive land cover data for analysis and testing. The results demonstrate that the multi-level spatial index method exhibits excellent scale adaptability, and grid coding dimensionality reduction and numerical operations effectively reduce the computational load of spatial retrievals of complex vector patches. This method significantly improves the retrieval efficiency of large-scale national geographic conditions data, providing an efficient technique for lightweight information extraction of large-scale monitoring geospatial data within spatial computing systems. The method holds reference value for on-demand retrieval, analysis, and decision-making of natural resource spatio-temporal big data.

1. INTRODUCTION

Following the successful completion of the first national geographic conditions survey in China, annual full-coverage, large-scale time series data of national geographic conditions have been generated since 2016, leading to the establishment of a national spatio-temporal information database (Gao et al. 2018; LI et al. 2016a; Wenzhong et al. 2017). Based on large-scale cloud storage and spatial database facilities, efficient storage and management of national geographic conditions monitoring data at a scale of 100 million has been realized. Distributed storage and online maps-based vector rendering are used to solve the problems of spatial data management, scheduling and dynamic browsing, providing strong support for national geographic conditions data management and online information services.

As the geographic conditions monitoring becomes integrated into the national natural resources investigation and monitoring system since 2018, the refined land cover data as the basic geographic conditions data is being directly applied in numerous natural resource supervision cases and the approval of basic farmland(Jiping et al. 2019). This highlights the fundamental value of geographic conditions monitoring data in natural resource management, analysis, and decision-making. However, the information service and decision computing capabilities of geographic conditions monitoring data are facing new demands within this new context of transitioning from indirect to direct service. The efficiency of information extraction to meet the requirements of natural resource management has become a key aspect to effectively utilize the geographic conditions monitoring data and improve the ability of information service and support.

The earth irregularity and spatial representation system based on Euclidean space and map projection will face the problems of discontinuous spatial representation and complex spatial computation in the case of large-scale spatial data(Béjar et al. 2023; Bondaruk et al. 2020; LI et al. 2010; Sahr et al. 2003). In order to support efficient geospatial data management, spatial indices, represented by B-tree, R-tree, and quadtree, have emerged as core technologies in various geospatial databases, playing an irreplaceable role in mainstream, mature geographic conditions database solutions(Li et al. 2016b; Yao and Li 2018).

Well-known open-source suites, such as PostGIS, implement spatial indices through R-tree or GIST tree indices (supporting geohash and projection), which transform PostgreSQL into a powerful spatial database with efficient spatial data management, quantity measurement, and topology analysis capabilities(Obe and Hsu 2021). MyISAM, the storage engine for the open-source MySQL database, has long supported spatial indexing based on R-tree indexing. The powerful InnoDB engine began supporting spatial indexing after version 5.7.4 labs, significantly improving spatial data management capabilities(Piórkowski 2011). Oracle, a prominent commercial database, has implemented R-tree, quadtree, and other spatial indices within the Oracle Spatial geospatial suite, accumulating years of expertise in spatial data storage and management. In addition, the development of Spatial and GeoRaster spatial functions and geometric operators has been fully advanced since the Oracle 12c version, which is committed to promoting comprehensive commercial geospatial database solutions(Bach 2014).

It is well known that index computation based on geospatial data has the characteristics of intensive computation(Park 2014; Wang et al. 2015; Yao et al. 2019). With the development of Earth observation technology, geospatial big data presents the characteristics of rapid expansion of data volume and diversification of spatial scale. A single type of spatial index has

more and more problems such as large amount of coordinate computation and poor scale adaptability, which gradually cannot meet the needs of rapid extraction and service of large-scale geospatial information(Bondaruk et al. 2020; Li et al. 2019). Geographic grids take the Earth as a whole into consideration, employing regular grids or cells to express Earth's space continuously and seamlessly, with each grid or cell representing the spatial range or position of the region it occupies. Grid coding algorithms transform spatial grids or cells into numerical codes, substituting complex spatial computations with simple coding operations. Consequently, on-demand location expression, spatial correlation, and statistical analysis can be realized in geographic space, ultimately achieving dimensionality reduction of spatial expression and simplifying spatial computation(Qian et al. 2019; Wang et al. 2020).

Internationally, geographic grids that have attracted widespread attention encompass the United States National Grid (USNG), the British National Grid Reference (BNGR), among others (Authority 2000; Cao et al. 2022; Yao et al. 2019). The US Military Grid Reference System (MGRS) and the Global Area Reference System (GARS) have been employed as global location coding standards for the US military to fulfill the requirements of cooperative operation of combat systems and command organizations(Sanjeewa 2016). In the civilian sector, innovations in the Internet economy, spearheaded by Google Earth, have adopted multi-tier tile grids to standardize global massive remote sensing image data and streamline online services, thereby popularizing geographic conditions services and becoming the Internet map service standard. To enhance grid segmentation expression accuracy, Google introduced the Google S2 spherical segmentation algorithm, which leverages the more universal Hilbert curve space-filling algorithm in multidimensional space expression(Kmoch et al. 2022). This algorithm has been extensively applied in Internet location services such as Google Maps search and Uber taxi proximity search, leading the advancement of location coding and Internet services development, and gaining adoption in an increasing number of fields.

In China, numerous expert teams from institutions, including the Chinese Academy of Sciences, Peking University, Wuhan University, and China University of Mining and Technology, have proposed various grid systems featuring unique characteristics for diverse fields and perspectives. Practical progress has been made in grid systems such as GeoSOT, SDOG, S3G, DOG, HQBS, and others, some of which have been validated and applied in professional fields like national public security, navigation and positioning, postal logistics, disaster reduction, and land management (Bondaruk et al. 2020; Jieqing et al. 2016; Qian et al. 2019). With the rapid development of new services such as earth observation, location service and sharing economy, grid information service has gradually transitioned from professional application to popular service. Internet services such as location search, shared bikes, etc., adopt the Z-curve hashing binary idea to carry out multilevel finite binary coding of geographical latitude and longitude coordinates, which solves the problem of near-real-time location expression and adjacent search of massive space points, and achieves good application results, indicating that geographic grid has good application potential in spatial big data retrieval(Bondaruk et al. 2020; Cao et al. 2022; Kmoch et al. 2022).

In response to this new demand in the era of spatio-temporal big data, international standards organizations and research institutions have formulated relevant standards for geographic grids. For example, the Open Geospatial Consortium (OGC) released the Discrete Global Grid System (DGGS) standard, which provides a framework for dividing the Earth's surface into layered tessellation of grid cells(Peterson 2016). National Geomatics Center of China and other research institutions have also jointly formulated national standards for China's geospatial grid, striving to create unified and standardized geospatial grid standards from the implementation level, realize multi-level standardized grid division of geospatial, and provide a unified standardized grid framework(Standardization 2009). This work has important practical value to promote the development of geographical grid from research to application in the new era.

In response to the demand for efficient indexing of massive geospatial monitoring data, this study presents a large-scale geospatial data indexing framework based on a multi-level geographic grid. The research investigates efficient index technologies for massive data, employing a typical grid algorithm, and concentrates on current practical challenges, such as multi-level index architecture, grid index efficiency, and multi-scale adaptability. National-scale land cover data is utilized for experimental analysis and performance verification, with the aim of providing technical support to enhance the indexing efficiency of large-scale geospatial data and facilitate natural resource management.

2. GENERAL FRAMEWORK

2.1 Efficient Indexing of Multi-scale Grid Encoding Model

Land cover data is one of the core achievements of geographic national conditions monitoring, characterized by comprehensive coverage, diverse types, fine patches, objective information, and strong timeliness (Pandey et al. 2021). It has unique advantages in supporting resource surveys, ecological protection, spatial planning, urban management, and macro decision-making. From the perspective of spatial computation and information extraction, land cover data indexing mainly faces three bottlenecks.

2.1.1 Wide range and large data volume: A single data layer of land cover seamlessly covers China's land territory, with a broad spatial distribution range, over 260 million patches in a single data layer, and data volume exceeding 300GB. Any spatial retrieval requires spatial computation of 260 million patches across the national land area, posing management, scheduling, and computational challenges that far exceed the supporting capacity of traditional centralized database systems.

2.1.2 Complex patches and large computational workload: Land cover patches were collected for the first time based on aerial photography and satellite imagery with resolutions better than 1 meter across the country, including seven provinces with resolutions better than 0.5 meters. The rich data sources resulted in accurate and complex boundaries of land cover patches, with complex patches such as woodland, grassland, and sandy land having as many as 100,000 nodes, and road and water systems having dozens of inner loops within a single patch. Conventional spatial computations and relationship discernments often require complex computations.

2.1.3 Varying scales and indexing challenges: Land cover data divides the entire surface into 86 detailed classes, with a single land class collection index refined to a minimum field of 200 square meters. Situations where small patches of 200 square meters, such as buildings and artificial facilities, coexist with large patches spanning multiple square kilometers, such as deserts and water surfaces, are common. The scale disparity between patches makes traditional single spatial indexing unsuitable for all granularity patches.

2.2 Efficient Indexing Framework

Traditional spatial indexing focuses on designing indices for individual spatial elements. When supporting the spatial indexing and computation of the aforementioned land cover data, it presents significant limitations, such as high scanning costs for large volumes, complex patch computations, and imbalanced computation due to scale differences. To address these issues, a discrete and lightweight multi-level spatial indexing framework has been proposed, incorporating "storage separation-logical coupling, lightweight matching-spatial filtering, spatial indexing-precise computation" to reduce data volume and simplify computational complexity, as shown in Table 1.

Spatial level	Indexing Target	Indexing Method
Large-scale	Divide and conquer to reduce data volume	Employ various separation strategies, such as distributed storage and partitioned storage, to physically separate massive nationwide data according to spatial range, enabling a divide-and-conquer approach for massive data. Spatial retrieval scope is narrowed from a global to a local scale, eliminating a large number of irrelevant unit data, reducing the target data volume from nationwide to spatially relevant data units, and significantly lowering the national data volume.
Medium-scale	Lightweight indexing for rapid data filtering	Use lightweight indexing methods, such as geographic grid encoding, to quickly perform spatial matching and data filtering within a certain error range, excluding a large number of irrelevant data records within the target unit. The target data is rapidly narrowed down to the relevant range within the target grid, greatly reducing the spatial computational cost of precise data filtering.
Small-scale	Precise computation to extract accurate results	Adopt spatial indexing and precise spatial data discernment methods, such as spatial computation, to eliminate a small amount of data within the grid computation error range in a limited target data scope. Obtain the final spatial retrieval results.

Table 1. Multi-level spatial indexing framework

2.3 Multi-level Indexing Model

Based on the multi-level indexing framework and the objective situation of national land cover data, a three-level indexing method, "partition-grid-space," is proposed to gradually narrow down the indexing scope during indexing, avoiding the inefficiency caused by single-level indexing of the entire data set and effectively improving the retrieval efficiency of spatial data. The structure of the multi-level grid indexing model is shown in Figure 1.

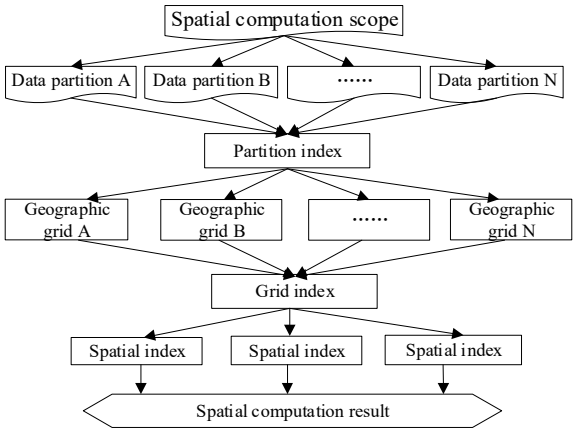


Figure 1. Multi-level grid spatial index model

2.3.1 Partition: Geographic national condition information extraction and data analysis have typical regional relevance, generally focusing on specific spatial units for information extraction. Therefore, data can be divided according to administrative regions, regular units, or social functional areas with uniform granularity and seamless continuity. Different unit data can be stored in the form of physical partitions, achieving physical separation and independence of national data volume, ensuring logical aggregation of data between units and physical separation of unit content data, and improving access efficiency and agility for each unit data. When accessing national data for information retrieval, indexing grids are formed based on the divided units, narrowing down the national data to several spatially relevant units within the target area, significantly reducing data scanning and access costs.

2.3.2 Grid: For each data unit, a grid index is established with elements as the objects, performing grid-based location matching and lightweight data filtering within each unit. A large number of geometric elements unrelated to grid units are excluded, narrowing down the target data to grid-related elements, significantly reducing the number of patch elements involved in actual spatial computations, and lowering the final spatial computation cost.

2.3.3 Space: At the limited spatial unit granularity level, the fast retrieval efficiency of spatial indexing can be utilized for precise spatial computations, forming the final spatial retrieval results. For each grid-related element, general spatial indexing and spatial judgment algorithms are employed for precise geometric computations and relationship discernments, extracting truly relevant elements within the scope of each target unit. The final data results are formed by merging the partitions unit by unit.

3. ALGORITHM

3.1 Index Construction Algorithm

Based on the general multilevel grid indexing model and considering the national geographical situation monitoring data, a multilevel grid indexing algorithm for geographical situation data is constructed to improve the indexing efficiency of massive geographical situation data. A county-level survey unit is used as the partition encoding algorithm, GeoSOT or Geohash as the grid encoding algorithm, and the common R-tree as the spatial indexing algorithm, forming a county-level administrative region-geographical grid-R-tree spatial indexing algorithm for geographical situation data. The multilevel indexing construction algorithm for national geographical situation data is shown in Figure 2.

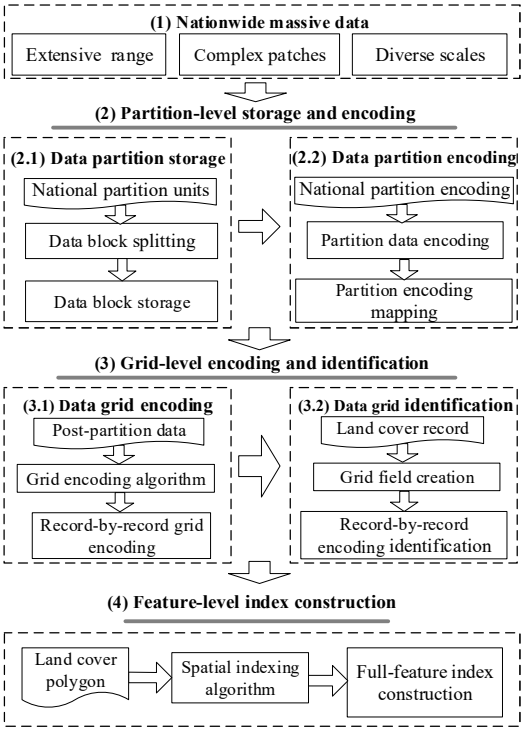


Figure 2. Index construction algorithm

3.1.1 County-level survey unit partition encoding algorithm: For the partition-level indexing method, county-level survey units are used as partition units to physically divide and store national data. The spatial range of the county-level survey unit is used as an irregular partition grid unit, and the data within each county-level survey unit is assigned a corresponding administrative region code, realizing the partition-level indexing encoding algorithm. The county-level survey unit storage grid encoding algorithm is: Code_partition = Code_Pac, where Pac is the corresponding administrative region code of the county-level survey unit.

3.1.2 Multigranularity grid encoding algorithm: For the grid-level indexing method, regular binary grids such as GeoSOT or Geohash are used as geographical grid partition algorithms. The minimum spatial bounding binary geographical grid of each vector feature is calculated, and the decimal geographical grid code is assigned to the corresponding vector feature, realizing the multigranularity grid encoding algorithm. Based on the multigranularity grid encoding algorithm, the geometric intersection computation of features is transformed into grid code character matching, greatly improving the efficiency of spatial intersection computation and realizing the fuzzy rapid filtering of spatial features.

3.1.3 R-tree spatial indexing algorithm: For the feature-level spatial indexing method, the general R-tree spatial indexing algorithm is used for spatial encoding.

3.2 Index Calculation Process

Under the multilevel grid indexing framework, a hierarchical indexing calculation method is adopted to achieve spatial retrieval calculation of massive data. The indexing calculation method is shown in Figure 3.

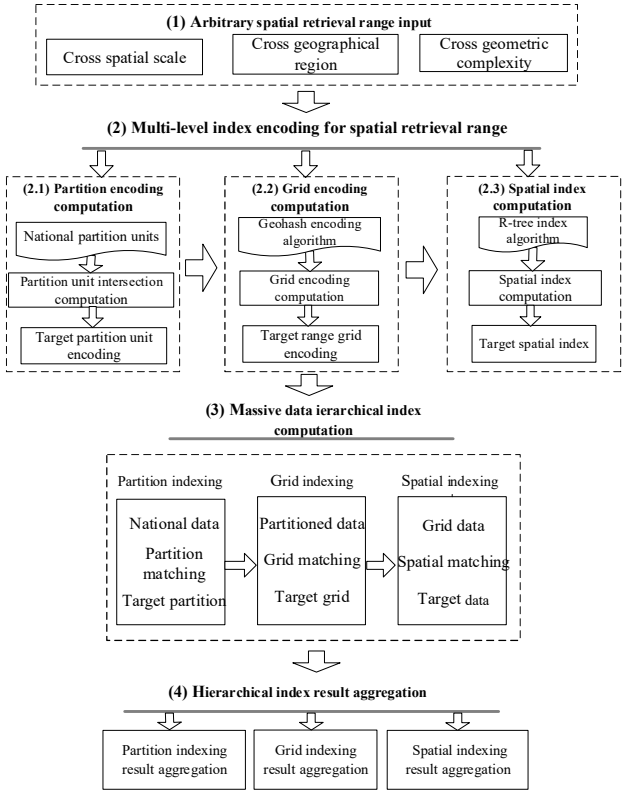


Figure 3. Index computation process

3.2.1 Arbitrary spatial retrieval range input: Efficient data retrieval is supported for any spatial range, with efficient support for complex spatial ranges spanning scales, regions, and multiple geometric complexity.

3.2.2 Multi-level index encoding: For the input spatial range, multilevel indexing encoding of the spatial retrieval range is performed, successively obtaining the partition unit encoding, grid encoding, and spatial indexing corresponding to the target range.

3.2.3 Hierarchical index computation: Based on the multi-level indexing computation results of the spatial retrieval unit, hierarchical index computations are carried out successively for the target data using the massive national data.

3.2.4 Hierarchical index result aggregation: For the outcome data of the hierarchical indexing calculation, the results are summarized by partition, forming the accurate indexing result data of the spatial retrieval range, and realizing the multilevel lightweight and fast spatial indexing of massive spatial data.

4. EXPERIMENT

Utilizing the national billion-scale geographical information data, this study employs the Oracle Spatial universal spatial database platform to perform cross-scale, cross-regional, and multi-form spatial range data index calculations. The computational efficiency of the multi-level indexing algorithm based on geographical grids and its applicability in different regions and scales are analyzed and tested. A horizontal efficiency comparison with the classic single-level R-tree spatial index is conducted, providing technical support for constructing an efficient online index and calculation of geographical information.

4.1 Experiment Data

4.1.1 Land cover data: Land cover data serves as the core vector output data for geographical information. The national surface is divided into 86 detailed categories, with more than 260 million patches. The minimum patch collection is refined to a field scale of 200 square meters. Complex patches, such as forests, grasslands, water surfaces, and roads, can have up to 100,000 single-element nodes. These data possess typical characteristics such as rich information, accurate representation, full coverage, fine-grained collection, and massive volume, constituting one of the performance bottlenecks for the computation and service applications of geographical information. The first national geographical information census land cover data is used as the test data (Figure 4) to conduct arbitrary range index calculations and performance comparison tests across the nation.

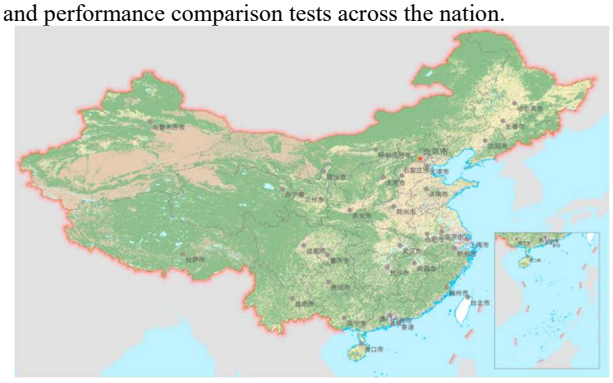


Figure 4. National land cover data distribution.

4.1.2 County-level survey unit: Following the organization of collected and updated data, the national data is divided into 2,777 county-level survey areas. County-level survey units are approximately equal in spatial granularity, and each unit objectively represents the continuous land cover within its respective area. County-level survey areas maintain stability across different years of monitoring data, making them ideal units for data partition organization in the database.

4.2 Experimental Methods

Cross-regional, multi-scale, and multi-form spatial retrieval ranges are used to carry out real-time spatial retrieval calculations for national data. Index calculation time consumption is recorded for different scenarios, and the computational efficiency of different data retrieval methods is comprehensively analyzed. Considering the differences in national regional characteristics and data granularity, rectangular and circular spatial units are employed. Five test areas are arbitrarily selected in the eastern, southern, western, northern, and central regions of the country, and ten progressively changing spatial retrieval units are constructed. A total of 100 diversified data index test units are created across the nation. For all index scenarios, both the multi-level index method proposed in this study and the traditional R-tree index method are independently calculated. The computational efficiency differences between the multi-level index method based on geographical grids and the traditional spatial index method in massive data indexing are compared and analyzed. The distribution of rectangular and circular test units, their spatial locations, and unit size settings are shown in Figures 5 and 6, respectively. The spatial area, patch quantity, and average patch area statistics of the 100 test units are illustrated in Figure 7.

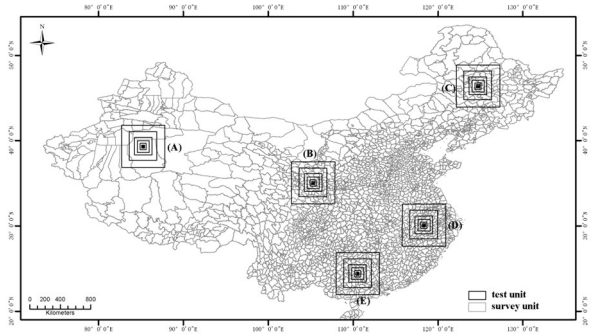
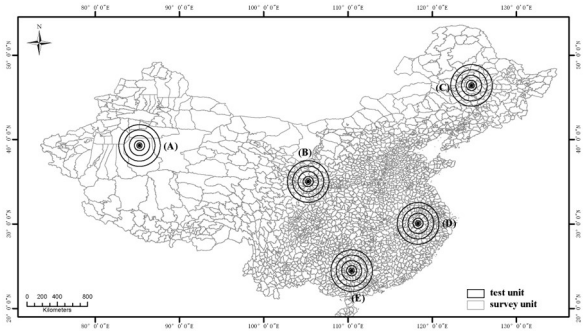


Figure 5. Rectangular spatial test units: Ten levels in five orientations: east, south, west, north, and central (A region denoted as rectangle west (abbreviated as RW), B region denoted as rectangle central (RC), C region denoted as rectangle north (RN), D region denoted as rectangle east (RE), and E region denoted as rectangle south (RS)). In each region, ten test units are set according to the spatial scale, with the largest square unit side length being 500 km. The unit side length is then successively halved, and the units are numbered from largest to smallest as 1-10.



. **Figure 6.** Circular spatial test units: Ten levels in five orientations: east, south, west, north, and central. A region denoted as circle west (abbreviated as CW), B region denoted as circle central (CC), C region denoted as circle north (CN), D region denoted as circle east (CE), and E region denoted as circle south (CS). The largest circular unit diameter is 500 km, and the unit diameter is then successively halved, with units numbered from largest to smallest as 1-10.

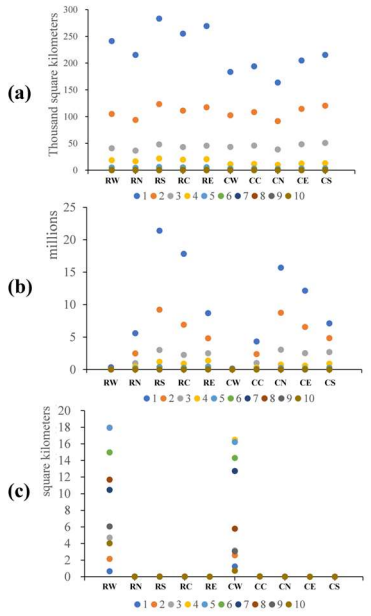


Figure 7. Spatial area, patch count, and average patch area of the 100 test units. (a) The geographical area of the 100 units, in thousands of square kilometers; (b) The number of surface cover patches contained within the 100 units, in millions; (c) The average area of surface cover patches within the 100 unit regions, in square kilometers.

4.3 Results and Analysis

Under the same computing platform, the calculation time consumption for the 100 different geographical locations, shapes, and granularities is recorded, and the computational efficiency comparison analysis is conducted. The comparison of multi-scale grid index calculation time consumption and conventional R-tree index time consumption for the 100 test units is shown in Figure 8. The multiplicative relationship statistics between the multi-scale grid index calculation time consumption and the conventional R-tree index time consumption for the 100 test units are displayed in Figure 9.

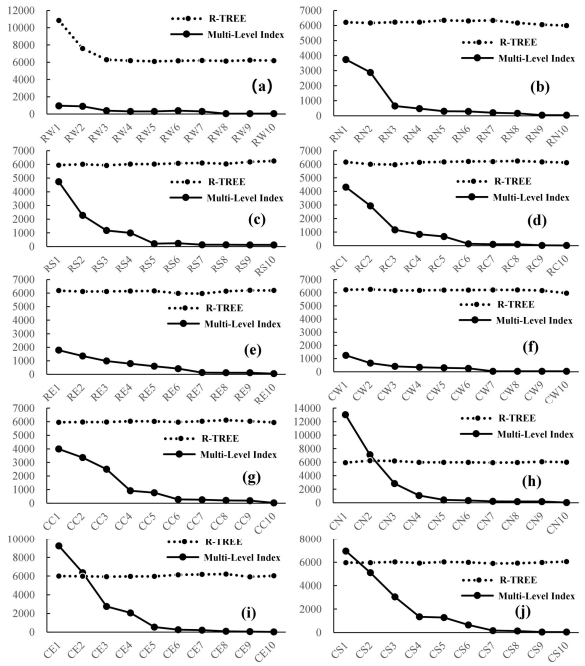


Figure 8. Comparison of computational time for multi-scale grid indexing and conventional R-TREE indexing for the 100 test units. The y-axis represents the computation time, measured in seconds.

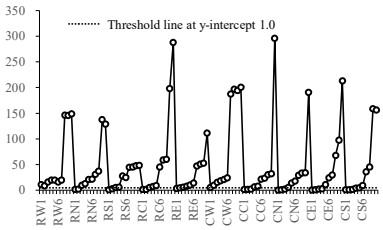


Figure 9. Ratio of computational time for multi-scale grid indexing and conventional R-TREE indexing for the 100 test units. The dashed line represents the threshold line at a y-intercept of 1.0, indicating the point where the approaches exhibit equal speed.

By conducting spatial index tests and comparisons across various geographical regions, spatial scales, and data geometric complexities, it can be deduced that the multi-scale geographic grid index employed in this study exhibits a significantly superior overall performance in comparison to the R-tree index. However, in the case of larger area retrieval units in a circular configuration, the R-Tree single index approach outperforms the multi-scale grid approach. The main analysis results are as follows:

4.3.1 Computation time consumption: The average time consumption based on the general parallel computation is 6,154.50 seconds. The relatively stable time consumption across different regions, shapes, and sizes indicates that traditional spatial computations suffer from imbalanced computational problems. In the parallel computation of national ultra-large-scale spatial data, the time consumption for different granularity computing units is relatively long.

4.3.2 Parallel computation balance: The balanced parallel computation of the spatially adaptive multi-scale grid index has an average time consumption of 1,209.51 seconds. As the spatial location, shape, and size of the computation units change, the computation time consumption dynamically varies. This suggests that the spatially adaptive multi-scale grid index parallel computation method can dynamically adapt to spatial granularity and optimize the utilization of computational resources, resulting in remarkable performance improvements in national ultra-large-scale parallel computations.

4.3.3 Performance enhancement: The time consumption of the spatially adaptive multi-scale grid index is 1/5 of the general parallel computation, with considerable variation across different spatial granularities. In the best-case scenario, it is 295.90 times faster than the general parallel computation, achieving geometric performance improvement. This indicates that the balanced parallel computation method of the multi-scale grid index has a significant improvement effect on ultra-large-scale computations. With the county-level survey area as the spatial constraint unit, when spatial granularity is adaptively matched according to the computation scenario, this method will achieve geometric performance improvement in spatial parallel computations.

5. CONCLUSIONS

In response to the challenges of efficient indexing and agile application for large-scale geographic spatial data, this study proposes a multi-level geographic grid indexing model and

algorithm. Building upon general geographic spatial data indexing, this model fully exploits the advantages of physical data partition indexing, administrative region business logic indexing, and spatial data indexing to search for hundreds of millions of complex land cover data points nationwide. An index algorithm combining macro, meso, and micro indices is proposed, and experimental verification is conducted on the basis of nationwide land cover data at the 100 million level under different data volumes and complexities. The results demonstrate that the multi-level geographic grid index proposed in this study achieves multi-level performance improvements compared to traditional single-level spatial indices. It displays significant potential in data management and computational analysis of results, such as surveying and monitoring. Further verification and optimization are required in order to address the inadequate performance of the multi-scale grid method, specifically in the scenario involving large-area retrieval units in a circular configuration. This particular case warrants additional investigation and improvement in future research endeavors.

The innovative value of the geographic grid lies in addressing spatial expression discontinuity and projection computation complexities within the Euclidean framework. Under the premise of certain accuracy requirements, the geographic grid meeting accuracy demands is adopted to achieve rapid spatial expression and computation, thereby realizing lightweight and agile spatial expression and computation services. In most cases, quickly returning results that meet accuracy requirements is more practically significant for high-performance information services of massive spatiotemporal data than waiting for an accurate result after a lengthy delay. The proposed method utilizes temporal monitoring data for multi-level index encoding, demonstrating significant potential in industries such as national natural resource management in the era of big data.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the National Geographic Conditions Monitoring Project for providing strong support for this study.

REFERENCES

Authority, T.V.,2000. United States National Grid. In: National Aeronautics and Space Administration.

Bach, M., 2014: *Expert Consolidation in Oracle Database 12c*. Apress.

Béjar, R., Lacasta, J., Lopez-Pellicer, F.J., & Nogueras-Iso, J., 2023. Discrete Global Grid Systems with quadrangular cells as reference frameworks for the current generation of Earth observation data cubes. *Environmental Modelling & Software* 162(1), 105656.

Bondaruk, B., Roberts, S.A., & Robertson, C., 2020. Assessing the state of the art in Discrete Global Grid Systems: OGC criteria and present functionality. *Geomatica* 74(1), 9-30.

Cao, J., Shi, S., Guo, Q., & Huang, G., 2022. Building Space Coding Based on Beidou Grid Position Code. *Mobile Networks and Applications* 27(4), 1700-1715.

Gao, Y., Liu, R., Liu, J., & Cheng, T., 2018. Research on the Construction of Remote Sensing Automatic Interpretation Symbol Big Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42(1), 381-384.

Jieqing, Y., Zhen, S., Lixin, W., Songwei, X., & Yongji, J., 2016. The Framework of Earth System Cellular Automata Based on SDOG-ESSG and Its Preliminary Experiment. *Acta Geodaetica et Cartographica Sinica* 45(S1), 40-52.

Jiping, L., Chun, D., Xiaochen, K., Shike, Q., Rong, Z., Bin, L., & Lijian, S., 2019. National Geographical Conditions Statistical Analysis in the Era of Big Data. *Geomatics and Information Science of Wuhan University* 44(1), 68-76.

Kmoch, A., Vasilyev, I., Virro, H., & Uuemaa, E., 2022. Area and shape distortions in open-source discrete global grid systems. *Big Earth Data* 6(3), 256-275.

LI, D., DING, L., & SHAO, Z., 2016a. Reflections on issues in national geographical conditions monitoring. *Geomatics and Information Science of Wuhan University* 41(2), 143-147.

LI, D., GONG, J., & SHAO, Z., 2010. From Digital Earth to Smart Earth. *Geomatics and Information Science of Wuhan University* 35(2), 127-132.

Li, S., Dragicevic, S., Castro, F.A., Sester, M., Winter, S., Coltekin, A., Pettit, C., Jiang, B., Haworth, J., & Stein, A., 2016b. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS journal of Photogrammetry and Remote Sensing* 115(1), 119-133.

Li, S., Pu, G., Cheng, C., & Chen, B., 2019. Method for managing and querying geo-spatial data using a grid-code-array spatial index. *Earth Science Informatics* 12(173-181.

Obe, R., & Hsu, L.S., 2021: *PostGIS in action*. Simon and Schuster.

Pandey, P.C., Koutsias, N., Petropoulos, G.P., Srivastava, P.K., & Ben Dor, E., 2021. Land use/land cover in view of earth observation: data sources, input dimensions, and classifiers—a review of the state of the art. *Geocarto International* 36(9), 957-988.

Park, K., 2014. Location-based grid-index for spatial query processing. *Expert Systems with Applications* 41(4), 1294-1300.

Peterson, P.R., 2016. Discrete global grid systems. *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology*1), 1-10.

Piórkowski, A., 2011. Mysql spatial and postgis–implementations of spatial data standards. *Electronic journal of Polish agricultural universities* 14(1), 1-8.

Qian, C., Yi, C., Cheng, C., Pu, G., Wei, X., & Zhang, H., 2019. Geosot-based spatiotemporal index of massive trajectory data. *ISPRS International Journal of Geo-Information* 8(284), 1-12.

Sahr, K., White, D., & Kimerling, A.J., 2003. Geodesic discrete global grid systems. *Cartography and Geographic Information Science* 30(2), 121-134.

Sanjeewa, G.,2016. Military Grid Reference System (MGRS) for the Future Joint Operations by the Sri Lankan Defense Forces. In: University of Sri Jayewardenepura, Nugegoda.

Standardization, N.T.C.f.,2009. Geographic grid. In. Beijing

Wang, S., Zhong, E., Lu, H., Guo, H., & Long, L.,2015. An effective algorithm for lines and polygons overlay analysis using uniform spatial grid indexing. In, *2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM)* (pp. 175-179): IEEE.

Wang, X., Wang, R., Zhan, W., Yang, B., Li, L., Chen, F., & Meng, L., 2020. A storage method for remote sensing images based on google s2. *IEEE Access* 8(1), 74943-74956.

Wenzhong, S., Pengfei, C., & Xiaokang, Z., 2017. Reliability Analysis in Geographical Conditions Monitoring. *Acta Geodaetica et Cartographica Sinica* 46(10), 1620-1626.

Yao, X., & Li, G., 2018. Big spatial vector data management: a review. *Big Earth Data* 2(1), 108-129.

Yao, X., Li, G., Xia, J., Ben, J., Cao, Q., Zhao, L., Ma, Y., Zhang, L., & Zhu, D., 2019. Enabling the big earth observation data via cloud computing and DGGS: Opportunities and challenges. *Remote Sensing* 12(1), 1-15.